

Genomics of Adaptation to Local Climate in Sitka spruce (*Picea sitchensis*)

by

Jason A. Holliday

B.Sc. University of Victoria, 2001

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate Studies

(Forestry)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2009

© Jason A. Holliday 2009

Abstract

Genecological studies in widely distributed tree species have revealed steep genetic clines along environmental gradients for traits related to adaptation to local climate. In the face of a changing climate, the ecological and economic importance of conifers necessitates an appraisal of how molecular genetic variation shapes quantitative trait variation. I have combined transcript profiling with association mapping to better understand the genomic architecture of adaptation to local climate in conifers, using Sitka spruce (*Picea sitchensis*) as a model. A microarray study during the fall hardening period revealed wholesale remodeling of the transcriptome within a population originating in the centre of the species range, and substantial variation in the autumn transcriptome was observed when populations from the northern and southern limits of the range were compared. Based on these data, a suite of candidate genes was selected and screened for single nucleotide polymorphisms (SNPs) in a panel of 24 individuals. A diverse array of biological processes were represented among the candidate genes, including stress response, carbohydrate, lipid and phenylpropanoid metabolism, light signal transduction, and transcriptional and posttranscriptional regulation. Nucleotide diversity in Sitka spruce was approximately average for a conifer ($\pi = 3.49 \times 10^{-3}$), and linkage disequilibrium decayed rapidly. Tests of selective neutrality suggest widespread purifying selection within these candidate genes, though evidence for positive selection was detected within a few. In addition, I observed evidence for diversifying selection in 8% of the studied genes, which exhibited high population differentiation relative to the genome-wide average F_{ST} of 0.12. To identify genetic determinants of phenotypic variation in locally adaptive traits, an Illumina GoldenGate assay was used to genotype 768 SNPs in a mapping population comprised of 410 individuals from 12 geographical populations collected from across the species range. After correcting for population structure and relative kinship, associations were detected in 28 of the candidate genes, which cumulatively explained 28% and 34% of the phenotypic variance in cold hardiness and budset, respectively. Most notable among these associations were five genes putatively involved in light signal transduction, the key pathway regulating autumn growth cessation in perennials. This study represents a significant step toward the goal of characterizing the genomic underpinnings of adaptation to local climate in conifers, and provides a substantial resource for breeding and conservation genetics in a changing climate.

Table of Contents

Abstract	ii
Table of Contents	iii
List of Tables	v
List of Figures	vi
Acknowledgements	vii
Co-Authorship Statement	viii
 Chapter 1: Literature Review and Research Objectives	
1.1 Introduction	1
1.2 Adaptation to local climate in temperate and boreal forest trees	1
1.3 Genomic dissection of local adaptation	3
1.4 Candidate gene selection	6
1.5 Nucleotide diversity and linkage disequilibrium in forest trees	11
1.6 Epigenetics and local adaptation	11
1.7 Sitka spruce as a model for studying the genomics of local adaptation	12
1.8 Research objectives	12
1.9 Literature cited.....	14
 Chapter 2: Global Monitoring of Autumn Gene Expression Within and Among Phenotypically Divergent Populations of Sitka spruce (<i>Picea sitchensis</i>)	
2.1 Introduction.....	20
2.2 Material and methods.....	22
2.3 Results.....	26
2.4 Discussion.....	30
2.5 Conclusions.....	37
2.6 Literature cited.....	48
 Chapter 3: Evidence for Positive and Diversifying Selection Among 174 Cold Hardiness-related Candidate Genes in Sitka spruce (<i>Picea sitchensis</i>)	
3.1 Introduction.....	52
3.2 Material and methods.....	54
3.3 Results.....	59
3.4 Discussion.....	63
3.5 Conclusions.....	69
3.6 Literature cited.....	81

Chapter 4: Association Mapping of Autumn Cold Hardiness and Budset Timing in Sitka spruce (*Picea sitchensis*)

4.1 Introduction	84
4.2 Material and methods.....	85
4.3 Results.....	89
4.4 Discussion.....	91
4.5 Conclusions.....	97
4.6 Literature cited.....	108

Chapter 5: Thesis Conclusion

5.1 Introduction.....	111
5.2 Wholesale remodeling of the transcriptome during the fall	111
5.3 Diversity and selection in local adaptation-related candidate genes	112
5.4 Significant associations between genotype and phenotype.....	114
5.5 Limitations of the current study	114
5.6 Future research.....	115
5.7 Literature cited.....	118

List of Tables

Table 2.1. Analysis of variance of cold injury data.....	38
Table 2.2. Functional categorization of within-population microarray results.....	39
Table 2.3. Functional categorization of among-population microarray results	42
Table 3.1. Summary of nucleotide diversity and tests of selective neutrality	70
Table 3.2. Candidate genes with significant tests of selective neutrality.....	71
Table 3.3. Model comparison for MLHKA test	72
Table 3.4. SNP outliers detected using BayesFst	73
Table 3.5. Pairwise F_{ST} estimates.....	74
Table 4.1. Origins of study populations for the association study and associated climatic variables.....	98
Table 4.2. Associations between SNPs and phenotypic traits.....	99

Lists of Figures

Figure 2.1. Origins of study populations for the microarray study, daily temperatures during sampling of the outdoor common garden, and phenotypic measurements of cold injury.....	44
Figure 2.2. Experimental design for the microarray study and summary of results	45
Figure 2.3. Gene ontology terms overrepresented among differentially expressed genes.....	46
Figure 2.4. Real-time PCR analysis of gene expression.....	47
Figure 3.1. Origins of study populations for the diversity study	75
Figure 3.2. Values for Tajima’s D and Fay and Wu’s H	76
Figure 3.3. Diversity vs. divergence.....	77
Figure 3.4. Results of BayesFst.....	78
Figure 3.5. Mantel test of isolation by distance.....	79
Figure 3.6. Distribution of pairwise LD estimates.....	80
Figure 4.1. Range map of Sitka spruce showing population sampled for Chapter 4.....	101
Figure 4.2. Regression of budset and cold hardiness on population location.....	102
Figure 4.3. Genetic clustering using the program ‘STRUCTURE’.....	103
Figure 4.4. Distribution of pairwise kinship estimates.....	104
Figure 4.5. Venn diagram of significant associations.....	105
Figure 4.6. Clinal variation in SNP frequencies.....	106
Figure 4.7. Genotypic effects for SNPs exhibiting clinal variation.....	107

Acknowledgements

I first acknowledge my friend and supervisor, Sally Aitken. This thesis would not have been possible without her patience, support, and extraordinary knowledge of forestry, genetics, and evolution. I would like to express my gratitude to my committee members Jörg Bohlmann, Kermit Ritland, and Carl Douglas, whose advice and support throughout my graduate career improved this thesis in many ways. I also thank Carol Ritland for her guidance in design and execution of many of the experiments described in this thesis, as well as Dylan Thomas and James Cummins for assistance with DNA extractions and generation of sequence data, and Ligia Mateiu, Nima Farzaneh, and Sam Yeaman for computing assistance.

I also thank current and former members of the Aitken lab, with whom I had many fruitful discussions throughout the years. In particular, I thank Christine Chourmouzis, Pia Smets and Joanne Tuytel for helping with field and greenhouse experiments, as well as cold hardiness phenotyping, and Makiko Mimura for her efforts in establishing the Sitka spruce common garden.

This work was supported by Genome Canada, Genome British Columbia, the Natural Sciences and Engineering Research Council of Canada (NSERC), a University of British Columbia Graduate Fellowship and an NSERC Postgraduate Scholarship.

Co-Authorship Statement

This thesis was written as a series of related manuscripts on the advice of my supervisory committee. Chapter 1 provides the context, literature review and research objectives. Chapter 2, 3 and 4 are manuscripts to be submitted to journals, and followed a typical scientific journal style including introduction, materials and methods, results, discussion, and conclusions. Chapter 5 summarizes results and discussion of this thesis and addresses further 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, this research would not have been possible without the contribution of my supervisor, Sally Aitken. She initially developed the concept for the association approach to studying local adaptation, and provided guidance throughout the execution of each of the chapters. To recognize her efforts, I have included her as a co-author on all chapters. Dr. Aitken also subsequently obtained funding through a collaborative proposal to Genome Canada with Jorg Bohlmann and Kermit Ritland. Dr. Bohlmann, and his postdoctoral fellow Steven Ralph helped me design and execute Chapter 2, and Rick White provided expertise in microarray data analysis. To recognize their efforts, they are included as co-authors on Chapter 2. Kermit Ritland provided useful input on the design and analysis of Chapter 3, and as such, he has been included as a co-author on this chapter.

1. Literature Review and Research Objectives

1.1 Introduction

Seasonal cold acclimation in coniferous and broadleaf trees of temperate and boreal regions is a remarkable physiological transition. The same dormant trees that in some cases survive submersion in liquid nitrogen (-196°C), once acclimated, can be killed during active growth by ambient temperatures only slightly below freezing (Sakai 1960; Weiser 1970). Susceptibility to frost damage of woody perennials in their native environment usually occurs during acclimation or after the onset of deacclimation (Sakai and Larcher 1987). Cold hardiness is therefore often thought of as a set of component traits, which include not only cold hardiness *per se*, but also timing of growth cessation and budset, initiation of cold acclimation, and timing of bud flush (Howe et al. 2003). Genecological studies in widely distributed tree species such as European aspen (*Populus tremula*), Sitka spruce (*Picea sitchensis*), and Scots pine (*Pinus sylvestris*) have revealed steep genetic clines along environmental gradients for these cold hardiness-related traits (Cannell and Sheppard 1982; Aitken and Adams 1996; Hurme et al. 1997; St Clair 2006; Hall et al. 2007; Mimura and Aitken 2007a). Whereas there is generally little difference in cold hardiness during the active growing season, significant variation both within and among populations reveals itself during the fall. In the southern portion of the ranges, moderate temperatures enable extended growing seasons, which may be accentuated by greater inter-specific competition for light resources. In contrast, the cold winter temperatures of the north necessitate a short growing season. Though generally well adapted to their local environment at present, the changing climate will force temperate and boreal tree species to either adapt, migrate, or be extirpated (Aitken 1999; Aitken et al. 2008). Conservation of existing variation in the genes involved in adaptation to local climate, particularly those involved in the growth and dormancy cycle and cold hardiness, is therefore crucial to maintain the adaptive capacity of our forests (Aitken et al 2008; Aitken 1999). This goal can be facilitated by a thorough appraisal of the genomic architecture of local adaptation to climate, including the number of genes involved, their respective effect sizes, linkages among them, and epistatic interactions.

1.2 Adaptation to local climate in temperate and boreal forest trees

1.2.1 Cold hardiness

The suite of component local adaptation-related traits outlined above are linked in that they stem from the necessity that perennial plants of the temperate and boreal regions cold acclimate each autumn, and deacclimate each spring. Because growth cessation is a prerequisite to cold acclimation,

the transition between growth and dormancy must be timed correctly to maximize annual growth without substantial risk of frost damage (Weiser 1970). Photoperiod and temperature regulate this progression, which for the fall is broken down into two phases (Weiser 1970). Growth cessation and bud set precede entry into phase I, the timing of which is largely determined by photoperiod. Cold hardiness increases slowly through phase I, but generally does not reach maximal levels until phase II, when freezing temperatures induce further hardening. Within 2-4 weeks of entry to phase II, maximum cold hardiness is typically achieved (Cannell et al. 1990). Freezing temperatures are not, however, required for all species to reach a high level of cold hardiness. For example, white and black spruce both acclimate to $< -40^{\circ}\text{C}$ under short photoperiod and at temperatures well above freezing (Colombo 1990; Silim and Lavender 1991).

Cold injury to non-acclimated plants is not the result of freezing alone, but also of accompanying cellular dehydration (Sakai and Larcher 1987). Cell walls and membranes effectively impede ice propagation, but once extracellular ice nucleation has occurred the nascent crystals grow rapidly through the apoplast (Ashworth 1992). Water is subsequently drawn out of cells as the extracellular concentration of solutes increases (Levitt 1980). If freezing occurs slowly enough, an equilibrium point may be reached where the osmolarity is the same on both sides of the membrane, and intracellular freezing is prevented by the high concentration of solutes (Palta et al. 1977). However, in many cases this equilibrium is never reached either because cooling is too rapid, diffusion of water across the membrane is too slow, or both (Sutinen et al. 2001). The result in some plants is a process known as supercooling, whereby intracellular water is cooled below the freezing point. Water requires a nucleation agent in order to crystallize and propagate. In supercooled tissue nucleation is suppressed, and in some cases liquid water can be cooled to -40°C (George and Burke 1984). At this point, ice crystals form *de novo*, and thus supercooling is not an effective cold tolerance strategy where minimum winter temperatures are below -40°C .

1.2.2 Cold hardiness in the context of climate change

The effects of climate change on cold injury in plants are not as straightforward as they may first appear. Although mean global temperatures are expected to increase, an increase in climatic variability is also anticipated (IPCC 2007). Such a development will likely have important consequences for perennial plant populations. Although cold acclimation is regulated largely by photoperiod, and hence will be unlikely to change substantially under climate change, unusually cold temperatures prior to acclimation could lead to injury. Similarly, the deacclimation period could see increased cold injury under climate change. Budburst is regulated largely by temperature, whereby first a chilling requirement must be met, followed by a heat sum. In spite of increased mean

temperatures, it is likely that the chilling requirement will continue to be met for most populations. However, with increased climatic variability, the heat-sum requirement may be met in mid-winter in some years. As plants are most susceptible to cold injury during active growth, if shoot elongation begins prematurely and is followed by sudden cold temperatures, cold damage could result in the destruction of the current years growth at a minimum, and it has the potential to kill the plant altogether. Seedlings are most likely to be injured or killed by such events, as their apex is closer to the ground where cold air tends to pool (Howe et al. 2003).

1.2.3 Quantitative trait variation within-species

There is a rich history of provenance trials in economically important conifers (Morgenstern 1996). These studies have been largely undertaken to identify seed sources that will be most productive in a particular environment, but have the side benefit of providing a resource for studies of genecology and local adaptation. Several species with wide latitudinal or elevational distributions have been studied in such common gardens, with *Pinus contorta*, the range of which spans 33° of latitude and 3900m of elevation, having perhaps the most substantial resource (Rehfeldt et al. 1999). In spite of very little population structure at neutral loci (i.e., low F_{ST}), wide variation in adaptive traits (i.e., high Q_{ST}) has been observed in this species (Xie and Ying 1995; Yang et al. 1996). This dichotomy is pervasive among studies of widely distributed tree species with continuous ranges (Savolainen et al. 2007), and is often invoked as evidence for local adaptation. Although modest deviations from equality of F_{ST} and Q_{ST} should be interpreted with care (Whitlock 2008), $Q_{ST} \gg F_{ST}$ provides reasonably strong evidence of variable selective pressures across a species range. When such a pattern is observed in species with high gene flow, the question then becomes, at what loci are the geographic populations that inhabit variable environments differentiated?

1.3 Genomic dissection of local adaptation

In spite of great interest in recent years, nearly all important questions related to the genomics of local adaptation are unanswered (Orr 2005). How many genes govern the relevant traits? What is the distribution of their effect sizes? What is their physical distribution within the genome? Until recently, the resources necessary to begin to answer these questions were only available for model species. However, non-model species provide more favorable biological frameworks for answering certain questions, and are generally of more ecological or economic importance. The decreasing cost of establishing genomics platforms (i.e., sequencing, expression profiling, high-throughput genotyping) has enabled the extension of these tools to forest trees, which has resulted in large expressed sequence tag (EST) databases, dense genetic maps, gene expression microarrays, and in the

case of *Populus trichocarpa*, a genome sequence (Ralph et al. 2006a; Ralph et al. 2006b; Ralph et al. 2008; Tuskan et al. 2006). Although economically important traits such as growth and wood quality have received substantial interest, the importance of forest health necessitates an equal emphasis on understanding how molecular genetic variation shapes quantitative phenotypic variation in locally adaptive traits in the wild.

1.3.1 Quantitative trait loci mapping

In conifers, locally adaptive traits such as cold hardiness and budset timing are continuously distributed within-population, and their genomic dissection became possible in the 1990's through the combination of genetic maps and breeding populations to map the associated quantitative trait loci (QTL). To do this, controlled crosses segregating for the trait of interest are genotyped at numerous marker loci, which are typically microsatellites or single nucleotide polymorphisms (SNPs). Each individual is then assayed for the trait of interest, and statistical associations identified between marker alleles and phenotypic variants. This approach relies on the few recombination events that have occurred within the pedigree under study, which results in extensive linkage disequilibrium (LD) between the anonymous marker loci and the loci influencing the trait (the QTL) (Lander and Botstein 1989). A series of QTL studies in coastal Douglas-fir has been undertaken to better understand the genomic basis for variation in bud phenology and cold hardiness. For fall cold hardiness, 11 QTL were detected, with percent variance explained (PVE) per QTL ranging from 2.0-6.8, whereas nine were detected for spring cold hardiness, with PVE of 2.0-7.5 (Jermstad et al. 2001b; Wheeler et al. 2005). For timing of budflush, no fewer than 33 unique QTL were identified (Jermstad et al. 2001a). These results indicate that many genes underlie locally adaptive traits in conifers, and that modest effect sizes are the norm. Although a powerful approach to genetic dissection of complex traits, in non-model systems QTL mapping has significant limitations. First, associations are specific to the mapping population and cannot be transferred among populations or species. Second, in the absence of a genome sequence positional cloning of QTL is exceedingly difficult, and the identity of QTL is generally not known. In addition, although it is often assumed that a single gene underlies each QTL, in some cases QTL may reflect the cumulative effects of several or many genes. These limitations combined with the expansion of genomic resources has lead to a shift toward interest in a related technique, namely association mapping, that has important advantages over QTL studies (Neale and Savolainen 2004).

1.3.2 Association mapping

Originating in human genetics, association mapping is similar in many respects to QTL mapping but with one important distinction. Both of these methods track the co-segregation of marker alleles and phenotypic variants. In contrast to QTL mapping, however, association mapping employs large unstructured populations with minimal LD and highly abundant markers (usually single nucleotide polymorphisms (SNPs)) to identify genetic variants that associate with phenotypic variation (Lander and Schork 1994; Neale and Savolainen 2004). Whereas in QTL mapping markers associate with phenotypes due to linkage with functional variants, in association mapping the much more rapid decay of LD in the mapping population means that a marker may associate with the phenotype because it is the actual functional variant. In model systems with relatively high LD in natural populations, such as *Arabidopsis thaliana* and *Homo sapiens*, genome-wide association studies are possible (Wang et al. 2005), wherein tens to hundred of thousands of SNPs across the entire genome are tested for association, and positive results are followed up in independent mapping populations. Such studies are essentially finer scale QTL mapping, and as in QTL mapping, associations generally indicate linkage with the causative locus (though the physical distance between marker and QTL is expected to be much smaller in association studies). In contrast, non-model species and model species with very low LD (e.g., maize, *Zea mays*) require the *a priori* identification of candidate genes with a plausible functional role in the trait of interest. Annotations of homologs in model species and gene expression studies form the principle lines of evidence for candidate gene selection, and where a QTL map exists, co-localization of candidates with QTLs can provide additional support. Clearly the likelihood of success by this approach depends crucially on the selection of appropriate candidate genes. In spite of the potential complexities in identifying candidate genes, conifers are well suited to candidate-gene based association studies (Neale and Savolainen 2004; Gonzalez-Martinez et al. 2006c). One of the major challenges in designing an effective association study is population structure. In crop plants in particular, breeding history in domesticated lines and limited gene flow in wild populations has created complex stratification that could lead to spurious association (Flint-Garcia et al. 2003). Similar issues have led to false association in other systems, most notably (or infamously) in humans (Knowler et al. 1988). Fortunately, most temperate and boreal tree species do not suffer from these problems to a large extent. Population structure at selectively neutral sites is generally low, and gene flow via pollen is efficient (Hamrick and Godt 1996). In species where disjunct peripheral populations exist, population structure can be elevated (Mimura 2006), but this can be accounted for by excluding these populations, or by explicitly modeling population structure or kinship in the analysis (Yu et al. 2006).

1.4 Candidate gene selection

Spruce species have among the most substantial genomic resources of any conifer. Three North American spruce species have been targeted for extensive EST sequencing in recent years: Sitka spruce, white spruce (*Picea glauca*), and interior spruce (*Picea engelmannii* x *Picea glauca*). To date, more than 200,000 EST reads representing approximately 46,000 putative unique transcripts have been generated (Ralph et al. 2008; Pavy et al. 2005), with sampling from Sitka and white spruce contributing most to these resources. Approximately 75% of the EST reads assembled into contiguous sequences, but a high level of singleton unigenes (~26,000) suggests that many low-copy transcripts have been captured. Extensive normalization was carried out for many of the libraries, which lead to increased rates of gene discovery by filtering out highly abundant transcripts (Ralph et al. 2008). cDNA libraries from which the sequence reads were generated spanned many tissues, including bark, xylem, phloem, shoots, and lateral and apical buds. Various developmental stages were represented among these, and tissue was also harvested that had been subjected to both simulated and actual insect attack (Ralph et al. 2008). Although the focus was not on late season or dormant tissue, one of the libraries described in Pavy et al. (2005) comprised a developmental sampling sequence of vegetative buds that included the dormant state. Deeper sequencing of late season tissue, including both vegetative buds and needle tissue, has revealed a relatively high level of putative unique transcripts (K. Reid and J. Holliday, unpublished data), particularly considering the extensive sequencing that has already been carried out. In addition to EST sequencing, several full-length cDNA libraries were generated from which approximately 6500 full-length cDNAs were sequenced. In the absence of the genome sequence for spruce, these full-length cDNAs represent a substantial resource, although the libraries from which they were constructed were mostly from tissue stressed by insect feeding (Ralph et al. 2008). Further full-length sequencing of libraries from other tissues would enhance studies of local adaptation. Indeed, one of the recently-developed dormancy libraries is full-length (though the effort and expense involved in generating full-length reads from such a library is substantial, and it is therefore not possible to employ a full-length cDNA sequencing strategy for this library at this time).

On the basis of the sequence resources described above, three generations of spruce cDNA microarrays have been developed. The most recent of these comprises 21,840 elements (Ralph et al, 2006b; Friedmann et al. 2006). It is necessary to employ the latter in conjunction with these sequence resources to select candidate genes with a reasonable probability of being involved in the trait of interest. Although sequence similarity to cold hardiness and local adaptation-related genes in *Arabidopsis* may be sufficient to choose candidate genes in angiosperm trees (e.g. *Populus trichocarpa*), in the absence of corroborating evidence for function (e.g., gene expression), this is

unlikely to be a successful strategy in conifers. Not only did they diverge from angiosperms close to 400 mya (Kenrick and Crane 1997), but the evolution of conifer genomes has followed a very different trajectory than that of angiosperms. Whereas the difference in genome size and gene family size between poplar and *Arabidopsis*, for example, can be largely explained by whole genome duplication (Tuskan et al. 2006), the genomes of *Pinus* and *Picea* species appear to have expanded to their current massive sizes in large part through tandem gene duplications and expansion of multigene families (Kinlaw and Neale 1997; Ahuja and Neale 2005). As such, BLASTX searches between *Arabidopsis* and gymnosperms typically return many possible gymnosperm orthologs with no way to differentiate them. This problem is confounded by the redundancy present in EST databases, which hold by far the most sequence information for conifers.

With a relatively small number of mechanistic studies available, selection of candidate genes for population genomic dissection of local adaptation in forest trees relies on data from gene expression profiling. Gene expression microarrays for aspen and Scots pine have provided the first near global pictures of the autumn transcriptome in trees (Schrader et al. 2004; Joosen et al. 2006; Druart et al. 2007a; Ruttink et al. 2007). Far from a period of cellular ‘quiescence’, cold acclimation involves extensive remodeling of the transcriptome. Some of the genes that are upregulated appear to be involved in cold hardiness, though it is difficult to disentangle gene expression related to the concomitant process of dormancy induction. In the first such study reported, gene expression during active growth and dormancy in cambial meristem tissue from aspen trees was compared (Schrader et al. 2004). Of 13,490 cDNA clones on the array, 1773 were upregulated in the dormant tissue, whereas 1620 downregulated. In another study, a 1.5K microarray was recently used to study the transcriptional response of apical buds in a Scots pine (*Pinus sylvestris*) provenance grown in three separate field sites along a latitudinal transect, and identified a number of candidate pine cold hardiness-related genes (Joosen et al. 2006). One of the caveats of employing gene expression profiling to select candidate genes is that it is generally not possible to take all genes that are upregulated under a particular treatment (e.g., short days) forward to an association panel, and genes that are not strongly upregulated may nevertheless be important to the phenotype. Pairing down of that list requires a comprehensive survey of the literature to identify those genes most likely to be involved in the trait of interest. The following is an overview of the various gene families and functional categories of genes that are likely to be involved in either local adaptation in general, or cold hardiness in particular.

1.4.1 Dehydrins

Dehydrins (also known as late embryogenesis abundant, LEA) were among the first cold hardiness-related candidate genes identified in trees. Interest in these genes stems from their importance in cold hardiness in annual plants, and from a practical perspective, their abundance in cold hardy tissue (and hence ease of isolation). LEA-like transcripts have been identified in several tree species, including peach (*Prunus persica*) (Wisniewski et al. 1996; Wisniewski et al. 1999), Scots pine (Joosen et al. 2006), and white spruce (Richard et al. 2000; Liu et al. 2004). Although they are clearly important to the cold acclimation process, the protective mechanism is unclear. In angiosperms, dehydrins have been found localized in the cytoplasm, nucleus, and adjacent to cellular membranes (Close 1997; Koag et al. 2003). It has been suggested that they may function as solubilizing agents for cellular macromolecules (Close 1997), though this is not proven. Interestingly, a study of a nematode LEA-like protein suggests structural alterations upon dehydration leading to a coiled-coil-like structure similar to that of intermediate filaments (Goyal et al. 2003; Wise and Tunnacliffe 2004). This suggests a possible additional role for dehydrins in increasing mechanical strength of cells that have become dehydrated during cold acclimation (Wise and Tunnacliffe 2004).

1.4.2 Antifreeze proteins

Antifreeze proteins (AFPs) are a class of secreted proteins with homology to pathogenesis-related (PR) proteins chitinases, β -1,3-glucanases and thaumatin-like proteins (Yu and Griffith 1997). AFPs prevent the propagation of ice by binding irreversibly to the ice surface through multiple hydrophilic domains (Kuiper et al. 2001). Structural analysis of an insect AFP also suggest an important role for hydrophobic interactions in ice binding (Leinala et al. 2002). Extensive studies have been carried out in winter rye (*Secale cereale*), and although ice-binding domains in some species have been identified (Kuiper et al. 2001) there is no consensus domain that would enable identification of novel AFPs from a database search (Jia and Davies 2002; Griffith and Yaish 2004). Functional or gene expression studies are therefore needed to establish candidate antifreeze genes. The presence of gene products with antifreeze activity has been indirectly demonstrated in several tree species by a thermal hysteresis assay that measures differences in freezing and melting points of water in cold acclimated tissue. These include the woody angiosperms *Populus deltoides* (eastern cottonwood) and *Quercus alba* (white oak), as well as the gymnosperm *Ginkgo biloba* (Duman and Olsen 1993). In *Picea abies*, a candidate antifreeze protein, *af70*, is expressed in response to cold treatment (Sabala et al. 1997), and an apoplastic Douglas-fir chitinase with antifreeze activity has also been isolated (Zamani et al. 2003).

1.4.3 Light signal transduction

Phytochromes and their interactions with circadian oscillators form the basis for photoperiodism in plants (Thomas and Vince-Prue 1997), and are thought to be the primary regulators of nightlength-mediated bud set in perennials (Howe et al. 1996; Horvath et al. 2003). Indeed, overexpression of an oat *PHYTOCHROME A (PHYA)* gene in hybrid aspen (*Populus tremula* x *P. tremuloides*) blocked growth cessation and cold acclimation under short days (Olsen et al. 1997). Conifer phytochromes similar to *PHYA* and *PHYB* have been isolated from Norway spruce and Scots pine, respectively (Clapham et al. 1999). In *Arabidopsis* *PHYA* controls expression of the flowering time gene *FLOWERING LOCUS T (FT)* through regulation of *CONSTANS (CO)* (Yanovsky and Kay 2002), and a crucial role for *FT* in growth cessation was recently demonstrated in hybrid aspen (Bohlenius et al. 2006). Whether functional orthologs to *FT* exist in conifers is currently unknown, but *FT*-like transcripts have been identified in spruce EST databases. One of these transcripts was shown to be upregulated strongly following transfer of Norway spruce seedlings to short days (Gyllenstrand et al. 2007). This result is surprising, as it is the opposite pattern to that found in angiosperm trees. Given the similarity of angiosperm *FT* gene family members, it is possible that this gene has been misidentified as *FT* when it is in fact an ortholog to another member of the gene family (e.g., *TERMINAL FLOWER 1*). It seems likely that if they do exist and function in a similar fashion, conifer *FT* orthologs would exhibit similar expression patterns to those seen in aspen (i.e., downregulation following transition to short days).

1.4.4 Transcription factors

Whereas the C-repeat binding factors (CBFs) and their upstream regulator, *INDUCER OF CBF EXPRESSION 1 (ICE1)*, are crucial for transient cold tolerance in annuals, their role in seasonal cold acclimation in perennials is unclear (Stockinger et al. 1997; Chinnusamy et al. 2003; Van Buskirk and Thomashow 2006). Gene expression microarray studies in both angiosperm trees and conifers across the growth cessation, cold acclimation, and dormancy induction period show conflicting results in this area (Schrader et al. 2004; Joosen et al. 2006; Druart et al. 2007a; Ruttink et al. 2007). Whereas, Schrader et al. found a *CBF1*-like gene in *Populus tremula* upregulated 6.7-fold in a contrast between active and dormant cambium, Ruttink et al. (2007) found no differential expression of either *ICE1* or CBFs in *P. tremula* x *P. alba* buds up to 6-weeks following the transition to short days. In the latter, many canonical downstream targets of the CBF regulon were induced. It is possible that the coarse nature of the sampling has contributed to variation in results for CBF expression, as CBFs appear to be functional in angiosperm trees. *P. trichocarpa* transformed with an *Arabidopsis CBF1* gene increased constitutive freezing tolerance (Benedict et al. 2006), and

overexpression of a sweet cherry (*Prunus avium*) CBF in *Arabidopsis* conferred increased freezing tolerance (Kitashiba et al. 2004). The situation in conifers is less clear. Although phylogenetic analysis of AP2 transcription factors in spruce suggests the existence of a CBF/DREB clade (S. Ralph, 2008, unpublished data), microarray studies during cold acclimation in Scots pine did not find conifer CBF-like genes to be induced (Joosen et al. 2006).

1.4.5 Metabolic remodeling

Cellular membranes are the primary sites of freezing injury. Plasma membranes from non-acclimated plants suffer expansion-induced lysis and lipid phase transitions; however, membranes from cold-acclimated plants resist this damage (Steponkus 1984; Webb et al. 1993; Webb and Steponkus 1993). The hydrophilic nature of dehydrins may function in part to stabilize membranes (Close 1997), and changes in the composition of membranes themselves likely also plays a role in their ability to resist rupture. In *Arabidopsis*, higher membrane phospholipid content as well as an increase in unsaturated phospholipids occurs during cold acclimation at low, above freezing temperatures (Uemura et al. 1995). Increased unsaturation of phosphatidylglycerol through constitutive expression of glycerol-3-phosphate acyltransferase also increased chilling resistance in tobacco (*Nicotiana tabacum*) (Murata 1992). Similar results have been observed in conifers. For example, needles of Scots pine exhibit an increase in phospholipid content upon cold acclimation (Ketchie et al. 1987), and acclimated white spruce has an increased level of unsaturated phospholipids (Senser and Beck 1982). Unfortunately, the difficulty in analyzing the lipid fraction in metabolite screens has left the global picture of lipid remodeling during cold acclimation obscure. Inferences from gene expression studies, however, suggests some of the important players in this process. For example, in *Arabidopsis* overexpression of the lipid transfer protein *EARLII* leads to reduced electrolyte leakage during freezing damage (Bubier and Schlappi 2004). In cold acclimated European birch (*Betula pendula*) leaves, an increase in transcription of several fatty acid desaturases was accompanied by an increase in the level of unsaturated phospholipids (Martz et al. 2006).

Although membrane remodeling likely plays a role in preventing rupture upon extracellular freezing, soluble carbohydrates have a dual role in protection of membranes in that they directly stabilize membranes (Bryant et al. 2001) and are thought to reduce osmotic gradients across those membranes. A positive correlation between sugar content and hardiness has been observed in Scots pine, lodgepole pine, Norway spruce, and red spruce (*Picea rubens*) (Aronson et al. 1976; Dehayes 1992; Ogren et al. 1997). In particular, di- and trisaccharides such as sucrose and raffinose increase during cold acclimation (Wang and Zwiazek 1999; Druart et al. 2007a; Ruttink et al. 2007). Conversely, starch concentration has been shown to be at a minimum in winter in red pine (*Pinus*

resinosa), Scots pine and red spruce (Alscher et al. 1989; Hansen et al. 1996), and in aspen, starch breakdown appears to be regulated by short days (Druart et al. 2007). Sugar metabolism genes upregulated during the fall in hybrid poplar include galactinol synthase, raffinose synthase, and sucrose synthase (Ruttink et al. 2007).

1.5 Nucleotide diversity and linkage disequilibrium in conifers

Several surveys of nucleotide diversity have been undertaken in temperate and boreal forest trees (Brown et al. 2004; Krutovsky and Neale 2005; Gonzalez-Martinez et al. 2006a). To date, these have generally taken the form of re-sequencing of a handful of genes in a few tens of individuals from several geographic populations. The surprising result that has emerged from these studies is that forest trees do not have a level of diversity that would be expected given their life history traits and previously characterized allozyme variation. In any event, there is ample variation in forest trees to exploit for the purposes of association genetics, though this is quite variable among loci.

One of the features of conifer genomes that conforms to predictions based on life history traits is the decay of LD. In the studies conducted so far, intragenic LD was shown to decay to non-significant levels within a few hundred base pairs, and intergenic LD was virtually non-existent (Brown et al. 2004; Krutovsky and Neale 2005; Gonzalez-Martinez et al. 2006a; Heuertz et al. 2006). These results bode well for the potential of identifying ‘quantitative trait nucleotides’ (QTN) in forest tree association studies, but suggest a problem that is yet to be fully appreciated. Namely, the decay of LD may be so rapid that the use of so-called haplotype-tagging SNPs may be inefficient at detecting selection on adjacent loci. This is the approach that has been applied with good success so far in human association studies, and reduces the number of SNPs that must be genotyped. However, with the exceedingly low LD that appears to be pervasive in forest trees, it will likely be necessary to use saturated genotyping strategies (i.e., genotype all SNPs above a given frequency threshold of, e.g., 10%), although the haplotype-tagging approach has successfully reduced genotyping effort in loblolly pine, which appears to have somewhat higher LD than other species so far studied (Brown et al. 2004; Gonzalez-Martinez et al. 2006b). More generally, the extent to which selectively non-neutral candidate genes have higher than average LD will affect the likelihood of success of the haplotype-tagging approach.

1.6 Epigenetics and local adaptation

Although mutation and selection undoubtedly shape the genetics of local adaptation, it is becoming increasingly apparent that epigenetics may play an important role in this process. Recent studies in Norway spruce show that the temperature during both zygotic and somatic embryo

development can dramatically affect cold hardiness and bud phenology in the offspring (Johnsen et al. 2005a; Johnsen et al. 2005b; Kvaalsen and Johnsen 2007). In some cases, the offspring's phenotype varied the equivalent of six degrees of latitude from what was expected given the geographic origin of the parents. A crucial question that remains outstanding in this area is the extent to which these traits are persistent, both within an individual's lifetime and in its offspring and subsequent generations.

1.7 Sitka spruce as a model for studying the genomics of local adaptation

Spruce species have become one of two model systems for conifer genomics in North America (the other being loblolly pine). Of these two models, spruce now enjoys the largest single-pass and full-length EST resources, a 21.8k cDNA microarray, and dense genetic maps (Pavy et al. 2008). These features make spruce generally a good choice for studies of the genomics of local adaptation. Among the possible spruce species to be used for this purpose, Sitka spruce stands out due to its enormous latitudinal range and linearly heterogeneous environment along that range. It was not unexpected as a result of this climatic gradient that when populations from across the range were grown in a common garden, substantial population differentiation for traits related to local adaptation would be observed. What was unexpected was the strength of the clines, with Q_{ST} in some cases approaching 90%. Although association genetics is ultimately best employed in a single panmictic population, an initial pass will benefit by the ability to take advantage of this wealth of among-population phenotypic variation. Another benefit of this population variation is the ability to use F_{ST} outlier detection approaches for identifying genes involved in local adaptation.

1.8 Research objectives

In the following chapters I investigate the genomics of local adaptation in a model conifer, Sitka spruce by (i) assessing gene expression both within and among-population during the cold acclimation period, (ii) studying genetic variation in a suite of local adaptation-related candidate genes, and (iii) testing those candidate genes for associations with adaptive variation in cold hardiness. In Chapter 2, I conduct a study of global gene expression across five timepoints in a wild population of Sitka spruce from the centre of the species range, and compare gene expression in north, central, and southern populations at early and late autumn timepoints. In Chapter 3, I select a suite of 180 candidate genes to determine the level of nucleotide diversity and linkage disequilibrium within the species, and identify a number of those genes which may have been targets of recent selection using a variety of statistical tests against the neutrality of mutations. In Chapter 4 I use a rangewide panel of Sitka spruce genotypes to identify associations between a subset of the single

nucleotide polymorphisms identified in chapter 3 and autumn cold hardiness. Finally, in Chapter 5 I summarize my most significant findings and suggest future research direction.

1.9 Literature cited

- Ahuja, M. R., and D. B. Neale. 2005. Evolution of genome size in conifers. *Silvae Genetica* **54**:126-137.
- Aitken, S. N. 1999. Conserving adaptive variation in forest ecosystems. *Journal of Sustainable Forestry* **8**:1-10.
- Aitken, S. N., and W. T. Adams. 1996. Genetics of fall and winter cold hardiness of coastal Douglas-fir in Oregon. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere* **26**:1828-1837.
- Aitken, S. N., S. Yeaman, J. A. Holliday, T. Wang, and S. Curtis-McLane. 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications* **1**:95-111.
- Alscher, R. G., R. G. Amundson, J. R. Cumming, S. Fellows, J. Fincher, G. Rubin, P. Vanleuken, and L. H. Weinstein. 1989. Seasonal changes in the pigments, carbohydrates and growth of red spruce as affected by ozone. *New Phytologist* **113**:211-223.
- Aronson, A., T. Ingstad, and L. G. Loof. 1976. Carbohydrate metabolism and frost hardiness in pine and spruce seedlings grown at different photoperiods and thermoperiods. *Physiologia Plantarum* **36**:127-132.
- Ashworth, E. N. 1992. Formation and spread of ice in plant tissues. *Horticultural Review* **13**:215-255.
- Benedict, C., J. S. Skinner, R. Meng, Y. J. Chang, R. Bhalerao, N. P. A. Huner, C. E. Finn, T. H. H. Chen, and V. Hurry. 2006. The CBF1-dependent low temperature signalling pathway, regulon and increase in freeze tolerance are conserved in *Populus* spp. *Plant Cell and Environment* **29**:1259-1272.
- Bohlenius, H., T. Huang, L. Charbonnel-Campaa, A. M. Brunner, S. Jansson, S. H. Strauss, and O. Nilsson. 2006. CO/FT regulatory module controls timing of flowering and seasonal growth cessation in trees. *Science* **312**:1040-1043.
- Brown, G. R., G. P. Gill, R. J. Kuntz, C. H. Langley, and D. B. Neale. 2004. Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proceedings of the National Academy of Sciences of the United States of America* **101**:15255-15260.
- Bryant, G., K. L. Koster, and J. Wolfe. 2001. Membrane behaviour in seeds and other systems at low water content: the various effects of solutes. *Seed Science Research* **11**:17-25.
- Bubier, J., and M. Schlappi. 2004. Cold induction of EARLI1, a putative Arabidopsis lipid transfer protein, is light and calcium dependent. *Plant Cell and Environment* **27**:929-936.
- Cannell, M. G. R., and L. J. Sheppard. 1982. Seasonal changes in the frost hardiness of provenances of *Picea sitchensis* in Scotland. *Forestry* **55**:137-153.
- Cannell, M. G. R., P. M. Tabbush, J. D. Deans, M. K. Hollingsworth, L. J. Sheppard, J. J. Philipson, and M. B. Murray. 1990. Sitka spruce and Douglas-Fir seedlings in the nursery and in cold storage - Root growth potential, carbohydrate content, dormancy, frost hardiness and mitotic index. *Forestry* **63**:9-27.
- Chinnusamy, V., M. Ohta, S. Kanrar, B. H. Lee, X. H. Hong, M. Agarwal, and J. K. Zhu. 2003. ICE1: a regulator of cold-induced transcriptome and freezing tolerance in Arabidopsis. *Genes & Development* **17**:1043-1054.
- Clapham, D. H., H. U. Kolukisaoglu, C. T. Larsson, M. Qamaruddin, I. Ekberg, C. Wiegmann-Eirund, H. A. Schneider-Poetsch, and S. von Arnold. 1999. Phytochrome types in *Picea* and *Pinus*. Expression patterns of *PHYA*-Related types. *Plant Molecular Biology* **40**:669-678.
- Close, T. J. 1997. Dehydrins: A commonality in the response of plants to dehydration and low temperature. *Physiologia Plantarum* **100**:291-296.

- Colombo, S. J. 1990. Bud dormancy status, frost hardiness, shoot moisture-content, and readiness of black spruce container seedlings for frozen storage. *Journal of the American Society for Horticultural Science* **115**:302-307.
- DeHayes, A. L. 1992. Winter injury and developmental cold tolerance in red spruce. Pp. 295-337 in C. Eagar, and M. B. Adams, eds. *Ecology and decline of red spruce in the eastern United States*. Springer-Verlag, Berlin.
- Druart, N., A. Johansson, K. Baba et al. 2007. Environmental and hormonal regulation of the activity-dormancy cycle in the cambial meristem involves stage-specific modulation of transcriptional and metabolic networks. *Plant Journal* **50**:557-573.
- Duman, J. G., and T. M. Olsen. 1993. Thermal hysteresis protein activity in bacteria, fungi, and phylogenetically diverse plants. *Cryobiology* **30**:322-328.
- Flint-Garcia, S. A., J. M. Thornsberry, and E. S. Buckler. 2003. Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology* **54**:357-374.
- Friedmann, M., S. G. Ralph, D. Aeschliman, J. Zhuang, K. Ritland, B. E. Ellis, J. Bohlmann, and C. J. Douglas. 2007. Microarray gene expression profiling of developmental transitions in Sitka spruce (*Picea sitchensis*) apical shoots. *Journal of Experimental Botany* **58**:593-614.
- George, M. F., and M. J. Burke. 1984. Supercooling of tissue water to extreme low-temperature in overwintering plants. *Trends in Biochemical Sciences* **9**:211-214.
- Gonzalez-Martinez, S. C., E. Ersoz, G. R. Brown, N. C. Wheeler, and D. B. Neale. 2006a. DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in *Pinus taeda* L. *Genetics* **172**:1915-1926.
- Gonzalez-Martinez, S. C., E. Ersoz, G. R. Brown, N. C. Wheeler, and D. B. Neale. 2006b. DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in *Pinus taeda* L. *Genetics* **172**:1915-1926.
- Gonzalez-Martinez, S. C., K. V. Krutovsky, and D. B. Neale. 2006c. Forest tree population genomics and adaptive evolution. *New Phytologist* **170**:227-238.
- Goyal, K., L. Tisi, A. Basran, J. Browne, A. Burnell, J. Zurdo, and A. Tunnacliffe. 2003. Transition from natively unfolded to folded state induced by desiccation in an anhydrobiotic nematode protein. *Journal of Biological Chemistry* **278**:12977-12984.
- Griffith, M., and M. W. Yaish. 2004. Antifreeze proteins in overwintering plants: a tale of two activities. *Trends in Plant Science* **9**:399-405.
- Gyllenstrand, N., D. Clapham, T. Kallman, and U. Lagercrantz. 2007. A Norway spruce FLOWERING LOCUS T homolog is implicated in control of growth rhythm in conifers. *Plant Physiology* **144**:248-257.
- Hall, D., V. Luquez, V. M. Garcia, K. R. St Onge, S. Jansson, and P. K. Ingvarsson. 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.
- Hamrick, J. L., and M. J. W. Godt. 1996. Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **351**:1291-1298.
- Hansen, J., G. Vogg, and E. Beck. 1996. Assimilation, allocation and utilization of carbon by 3-year-old Scots pine (*Pinus sylvestris*, L.) trees during winter and early spring. *Trees-Structure and Function* **11**:83-90.
- Heuertz, M., E. De Paoli, T. Kallman, H. Larsson, I. Jurman, M. Morgante, M. Lascoux, and N. Gyllenstrand. 2006. Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [*Picea abies* (L.) Karst]. *Genetics* **174**:2095-2105.
- Horvath, D. P., J. V. Anderson, W. S. Chao, and M. E. Foley. 2003. Knowing when to grow: signals regulating bud dormancy. *Trends in Plant Science* **8**:534-540.

- Howe, G. T., S. N. Aitken, D. B. Neale, K. D. Jermstad, N. C. Wheeler, and T. H. H. Chen. 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.
- Howe, G. T., G. Gardner, W. P. Hackett, and G. R. Furnier. 1996. Phytochrome control of short-day-induced bud set in black cottonwood. *Physiologia Plantarum* **97**:95-103.
- Hurme, P., T. Repo, O. Savolainen, and T. Paakkonen. 1997. Climatic adaptation of bud set and frost hardiness in Scots pine (*Pinus sylvestris*). *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere* **27**:716-723.
- IPCC. 2007. Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Jermstad, K. D., D. L. Bassoni, K. S. Jech, N. C. Wheeler, and D. B. Neale. 2001a. Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. I. Timing of vegetative bud flush. *Theoretical and Applied Genetics* **102**:1142-1151.
- Jermstad, K. D., D. L. Bassoni, N. C. Wheeler, T. S. Anekonda, S. N. Aitken, W. T. Adams, and D. B. Neale. 2001b. Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. II. Spring and fall cold-hardiness. *Theoretical and Applied Genetics* **102**:1152-1158.
- Jia, Z. C., and P. L. Davies. 2002. Antifreeze proteins: an unusual receptor-ligand interaction. *Trends in Biochemical Sciences* **27**:101-106.
- Johnsen, O., O. G. Daehlen, G. Ostreng, and T. Skroppa. 2005a. Daylength and temperature during seed production interactively affect adaptive performance of *Picea abies* progenies. *New Phytologist* **168**:589-596.
- Johnsen, O., C. G. Fossdal, N. Nagy, J. Molmann, O. G. Daehlen, and T. Skroppa. 2005b. Climatic adaptation in *Picea abies* progenies is affected by the temperature during zygotic embryogenesis and seed maturation. *Plant Cell and Environment* **28**:1090-1102.
- Joosen, R. V. L., M. Lammers, P. A. Balk, P. Bronnum, M. Konings, M. Perks, E. Stattin, M. F. Van Wordragen, and A. H. M. van der Geest. 2006. Correlating gene expression to physiological parameters and environmental conditions during cold acclimation of *Pinus sylvestris*, identification of molecular markers using cDNA microarrays. *Tree Physiology* **26**:1297-1313.
- Kenrick, P., and P. R. Crane. 1997. The origin and early evolution of plants on land. *Nature* **389**:33-39.
- Ketchie, D. O., J. Bervaes, and P. J. C. Kuiper. 1987. Lipid-Composition of Pine Needle Chloroplasts and Apple Bark Tissue as Affected by Growth Temperature and Daylength Changes .1. Phospholipids. *Physiologia Plantarum* **71**:419-424.
- Kinlaw, C. S., and D. B. Neale. 1997. Complex gene families in pine genomes. *Trends in Plant Science* **2**:356-359.
- Kitashiba, H., T. Ishizaka, K. Isuzugawa, K. Nishimura, and T. Suzuki. 2004. Expression of a sweet cherry DREB1/CBF ortholog in *Arabidopsis* confers salt and freezing tolerance. *Journal of Plant Physiology* **161**:1171-1176.
- Knowler, W. C., R. C. Williams, D. J. Pettitt, and A. G. Steinberg. 1988. Gm3-5,13,14 and Type-2 diabetes-mellitus - an asociation in American-Indians with genetic admixture. *American Journal of Human Genetics* **43**:520-526.
- Koag, M. C., R. D. Fenton, S. Wilkens, and T. J. Close. 2003. The binding of maize DHN1 to lipid vesicles. Gain of structure and lipid specificity. *Plant Physiology* **131**:309-316.
- Krutovsky, K. V., and D. B. Neale. 2005. Nucleotide diversity and linkage disequilibrium in cold-hardiness- and wood quality-related candidate genes in Douglas fir. *Genetics* **171**:2029-2041.
- Kuiper, M. J., P. L. Davies, and V. K. Walker. 2001. A theoretical model of a plant antifreeze protein from *Lolium perenne*. *Biophysical Journal* **81**:3560-3565.
- Kvaalsen, H., and O. Johnsen. 2008. Timing of bud set in *Picea abies* is regulated by a memory of temperature during zygotic and somatic embryogenesis. *New Phytologist* **177**(1): 49-59.

- Lander, E. S., and D. Botstein. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**:185-199.
- Lander, E. S., and N. J. Schork. 1994. Genetic dissection of complex traits. *Science* **265**:2037-2048.
- Leinala, E. K., P. L. Davies, and Z. C. Jia. 2002. Crystal structure of beta-helical antifreeze protein points to a general ice binding model. *Structure* **10**:619-627.
- Levitt, J. 1980. Responses of plants to environmental stress. Academic Press, Inc., New York.
- Liu, J. J., A. K. M. Ekramoddoullah, D. Taylor, N. Piggott, S. Lane, and B. Hawkins. 2004. Characterization of Picg5 novel proteins associated with seasonal cold acclimation of white spruce (*Picea glauca*). *Trees-Structure and Function* **18**:649-657.
- Martz, F., S. Kiviniemi, T. E. Palva, and M. L. Sutinen. 2006. Contribution of omega-3 fatty acid desaturase and 3-ketoacyl-ACP synthase II (KASII) genes in the modulation of glycerolipid fatty acid composition during cold acclimation in birch leaves. *Journal of Experimental Botany* **57**:897-909.
- Mimura, M. 2006. Adaptation and Gene Flow in Sitka Spruce. Forest Sciences. University of British Columbia, Vancouver.
- Mimura, M., and S. N. Aitken. 2007. Adaptive gradients and isolation-by-distance with postglacial migration in *Picea sitchensis*. *Heredity* **99**:224-232.
- Morgenstern, E. K. 1996. Geographic variation in forest trees: genetic basis and application of knowledge in silviculture. University of British Columbia Press, Vancouver.
- Murata, N., O. Ishizaki-Nishizawa, S. Higashi, H. Hayashi, Y. Tasaka & I. Nishida. 1992. Genetically engineered alteration in the chilling sensitivity of plants. *Nature* **356**:710 - 713.
- Neale, D. B., and O. Savolainen. 2004. Association genetics of complex traits in conifers. *Trends in Plant Science* **9**:325-330.
- Ogren, E., T. Nilsson, and L. G. Sundblad. 1997. Relationship between respiratory depletion of sugars and loss of cold hardiness in coniferous seedlings over-wintering at raised temperatures: Indications of different sensitivities of spruce and pine. *Plant Cell and Environment* **20**:247-253.
- Olsen, J. E., O. Junttila, J. Nilsen, M. E. Eriksson, I. Martinussen, O. Olsson, G. Sandberg, and T. Moritz. 1997. Ectopic expression of oat phytochrome A in hybrid aspen changes critical daylength for growth and prevents cold acclimatization. *Plant Journal* **12**:1339-1350.
- Orr, H. A. 2005. The genetic theory of adaptation: A brief history. *Nature Reviews Genetics* **6**:119-127.
- Palta, J. P., J. Levitt, and E. J. B. Stadelmann, M.J. 1977. Dehydration of onion cells: a comparison of freezing vs. desiccation and living vs. dead cells. *Physiologia Plantarum* **41**:273-279.
- Pavy, N., C. Paule, L. Parsons et al. 2005. Generation, annotation, analysis and database integration of 16,500 white spruce EST clusters. *BMC Genomics* **6**.
- Pavy, N., B. Pelgas, S. Beauseigle, S. Blais, F. Gagnon, I. Gosselin, M. Lamothe, N. Isabel, and J. Bousquet. 2008. Enhancing genetic mapping of complex genomes through the design of highly-multiplexed SNP arrays: application to the large and unsequenced genomes of white spruce and black spruce. *BMC Genomics* **9**:21.
- Ralph, S., C. Oddy, D. Cooper et al. 2006a. Genomics of hybrid poplar (*Populus trichocarpa* x *deltoides*) interacting with forest tent caterpillars (*Malacosoma disstria*): normalized and full-length cDNA libraries, expressed sequence tags, and a cDNA microarray for the study of insect-induced defences in poplar. *Molecular Ecology* **15**:1275-1297.
- Ralph, S. G., H. Yueh, M. Friedmann et al. 2006b. Conifer defence against insects: microarray gene expression profiling of Sitka spruce (*Picea sitchensis*) induced by mechanical wounding or feeding by spruce budworms (*Choristoneura occidentalis*) or white pine weevils (*Pissodes strobi*) reveals large-scale changes of the host transcriptome. *Plant Cell and Environment* **29**:1545-1570.

- Ralph, S. G., H. J. E. Chun, N. Kolosova et al. 2008. A conifer genomics resource of 200,000 spruce (*Picea* spp.) ESTs and 6,464 high-quality, sequence-finished full-length cDNAs for Sitka spruce (*Picea sitchensis*). *BMC Genomics* **9**.
- Rehfeldt, G. E., C. C. Ying, D. L. Spittlehouse, and D. A. Hamilton. 1999. Genetic responses to climate in *Pinus contorta*: Niche breadth, climate change, and reforestation. *Ecological Monographs* **69**:375-407.
- Richard, S., M. J. Morency, C. Drevet, L. Jouanin, and A. Seguin. 2000. Isolation and characterization of a dehydrin gene from white spruce induced upon wounding, drought and cold stresses. *Plant Molecular Biology* **43**:1-10.
- Ruttink, T., M. Arend, K. Morreel, V. Storme, S. Rombauts, J. Fromm, R. P. Bhalerao, W. Boerjan, and A. Rohde. 2007. A Molecular timetable for apical bud formation and dormancy induction in poplar. *Plant Cell* **19**:2370-2390.
- Sabala, I., H. Franzen, and S. vonArnold. 1997. A spruce gene, af70, constitutively expressed in somatic embryos and induced by ABA and low temperature in seedlings. *Physiologia Plantarum* **99**:316-322.
- Sakai, A. 1960. Survival of the twig of woody plants at -196 C. *Nature* **185**:393-394.
- Sakai, A., and W. Larcher. 1987. Frost survival in plants. Responses and adaptation to freezing stress. Springer-Verlag, Berlin.
- Savolainen, O., T. Pyhajarvi, and T. Knurr. 2007. Gene flow and local adaptation in tees. *Annual Review of Ecology Evolution and Systematics* **38**:595-619.
- Schrader, J., R. Moyle, R. Bhalerao, M. Hertzberg, J. Lundeberg, P. Nilsson, and R. P. Bhalerao. 2004. Cambial meristem dormancy in trees involves extensive remodelling of the transcriptome. *Plant J* **40**:173-187.
- Senser, M., and E. Beck. 1982. Frost resistance in spruce (*Picea abies*): The lipid composition of frost resistant and frost sensitive spruce chloroplasts. *Zeitschrift Fur Pflanzenphysiologie* **105**:241-253.
- Silim, S. N., and D. P. Lavender. 1991. Relationship between cold hardiness, stress resistance and bud dormancy in white spruce (*Picea glauca* (Moench) Voss.). seedlings. Pp. 9-14. 11th Annual Meeting of the B.C. Forest Nursery Association, Prince George.
- St Clair, J. B. 2006. Genetic variation in fall cold hardiness in coastal Douglas-fir in western Oregon and Washington. *Canadian Journal of Botany-Revue Canadienne De Botanique* **84**:1110-1121.
- Steponkus, P. L. 1984. Role of the plasma membrane in freezing injury and cold acclimation. *Annual Review of Plant Physiology and Plant Molecular Biology* **35**:543-584.
- Stockinger, E. J., S. J. Gilmour, and M. F. Thomashow. 1997. *Arabidopsis thaliana* CBF1 encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. *Proceedings of the National Academy of Sciences of the United States of America* **94**:1035-1040.
- Sutinen, M., R. Arora, M. Wisniewski, E. Ashworth, R. Strimbeck, and J. Palta. 2001. Mechanisms of Frost Survival and Freeze-Damage in Nature. Pp. 187-219 in F. J. Bigras, and S. J. Columbo., eds. *Conifer Cold Hardiness*. Kluwer Academic Publishers, Dordrecht.
- Thomas, B., and D. Vince-Prue. 1997. Photoperiodism in plants. Academic Press Limited, London.
- Tuskan, G. A.S. DiFazioS. Jansson et al. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**:1596-1604.
- Uemura, M., R. A. Joseph, and P. L. Steponkus. 1995. Cold acclimation of *Arabidopsis thaliana* - Effect on plasma membrane lipid composition and freeze induced lesions. *Plant Physiology* **109**:15-30.
- Van Buskirk, H. A., and M. F. Thomashow. 2006. *Arabidopsis* transcription factors regulating cold acclimation. *Physiologia Plantarum* **126**:72-80.

- Wang, W. Y., B. J. Barratt, D. G. Clayton, and J. A. Todd. 2005. Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet* **6**:109-118.
- Wang, Y. F., and J. J. Zwiazek. 1999. Spring changes in water relations, gas exchange, and carbohydrates of white spruce (*Picea glauca*) seedlings. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere* **29**:332-338.
- Webb, M. S., S. W. Hui, and P. L. Steponkus. 1993. Dehydration-induced lamellar-to-hexagonal-II phase transitions in DOPE-DOPC mixtures. *Biochimica et Biophysica Acta* **1145**:93-104.
- Webb, M. S., and P. L. Steponkus. 1993. Freeze-induced membrane ultrastructural alterations in rye (*Secale cereale*) leaves. *Plant Physiology* **101**:955-963.
- Weiser, C. J. 1970. Cold resistance and injury in woody plants. *Science* **169**:1269-1278.
- Wheeler, N. C., K. D. Jermstad, K. Krutovsky, S. N. Aitken, G. T. Howe, J. Krakowski, and D. B. Neale. 2005. Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. IV. Cold-hardiness QTL verification and candidate gene mapping. *Molecular Breeding* **15**:145-156.
- Whitlock, M. C. 2008. Evolutionary inference from Q(ST). *Molecular Ecology* **17**:1885-1896.
- Wise, M. J., and A. Tunnacliffe. 2004. POPP the question: what do LEA proteins do? *Trends in Plant Science* **9**:13-17.
- Wisniewski, M., T. J. Close, T. Artlip, and R. Arora. 1996. Seasonal patterns of dehydrins and 70-kDa heat-shock proteins in bark tissues of eight species of woody plants. *Physiologia Plantarum* **96**:496-505.
- Wisniewski, M., R. Webb, R. Balsamo, T. J. Close, X. M. Yu, and M. Griffith. 1999. Purification, immunolocalization, cryoprotective, and antifreeze activity of PCA60: A dehydrin from peach (*Prunus persica*). *Physiologia Plantarum* **105**:600-608.
- Xie, C. Y., and C. C. Ying. 1995. Genetic architecture and adaptive landscape of interior lodgepole pine (*Pinus contorta* ssp *latifolia*) in Canada. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere* **25**:2010-2021.
- Yang, R. C., F. C. Yeh, and A. D. Yanchuk. 1996. A comparison of isozyme and quantitative genetic variation in *Pinus contorta* ssp *latifolia* by F-ST. *Genetics* **142**:1045-1052.
- Yanovsky, M. J., and S. A. Kay. 2002. Molecular basis of seasonal time measurement in *Arabidopsis*. *Nature* **419**:308-312.
- Yu, J. M., G. Pressoir, W. H. Briggs et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* **38**:203-208.
- Yu, X. M., and M. Griffith. 1997. Characterization of winter rye antifreeze proteins in their native forms. *Plant Physiology* **114**:602-602.
- Zamani, A., R. Sturrock, A. K. M. Ekramoddoullah, S. B. Wiseman, and M. Griffith. 2003. Endochitinase activity in the apoplastic fluid of *Phellinus weirii*-infected Douglas-fir and its association with over wintering and antifreeze activity. *Forest Pathology* **33**:299-316.

2. Global Monitoring of Autumn Gene Expression Within and Among Phenotypically Divergent Populations of Sitka spruce (*Picea sitchensis*)¹

2.1 Introduction

Temperate and boreal trees alternate between periods of active growth in summer, and dormancy in winter. Beginning in late summer, and in response to short days, cessation of shoot elongation leads to initiation of cold acclimation. This period is characterized by suspension of mitotic activity that can be reversed given favorable conditions, and is followed late in the fall by bud endodormancy, which is maintained throughout the winter period until appropriate chilling and heat sum requirements are met (Rohde and Bhalerao 2007). Substantial cold hardiness is incompatible with growth, and tradeoffs exist in this annual cycle between competition for light resources and the need to acquire and maintain cold hardiness (Howe et al. 2003). As a result, timing of entry into and exit from dormancy is locally adaptive, and genetic clines observed in common gardens usually correspond to variation in climate of population origin along latitudinal or elevational gradients (Morgenstern 1996). In woody plants the period of autumn cold acclimation is typically divided into two phases (Weiser 1970). Phase I, which overlaps with growth cessation and dormancy induction, is triggered by a critical nightlength that is reached weeks or months before the first frost, and varies among families, populations, and species (Weiser 1970; Cannell and Sheppard 1982; Cannell et al. 1990; Aitken and Adams 1996). Cold hardiness steadily increases until the first sub-freezing temperatures occur. At this point, entry into phase II leads to further hardening, which is often rapid, and maximum cold hardiness is ultimately achieved (Weiser 1970). As maximum cold hardiness often far exceeds minimum recorded winter temperatures, cold injury is most commonly observed in spring and fall, before the onset of acclimation or after de-acclimation (Weiser 1970; Aitken and Adams 1996).

Although the phenotypic responses of conifers to long nights and low temperature are well studied, little is known about how these environmental cues are integrated at the molecular level. Much of what we do know about the molecular basis for plant cold tolerance is based on studies of herbaceous annuals such as *Arabidopsis thaliana* and winter rye (*Secale cereale*). These species

¹ A version of this chapter has been published. Holliday, J. A., S. G. Ralph, R. White, J. Bohlmann, and S. N. Aitken. 2008. Global monitoring of autumn gene expression within and among phenotypically divergent populations of Sitka spruce (*Picea sitchensis*). *New Phytologist* **178**:103-122.

provide favorable starting points for the study of cold hardiness because they are experimentally tractable. However, in contrast to perennials, annuals typically acclimate fully in response to a low temperature cue, which makes the regulation of cold hardiness in these two systems fundamentally different. In spite of this difference, few studies have focused on the suite of genes that contribute to cold hardiness in perennials. The recent completion of the *Populus trichocarpa* genome (Tuskan et al. 2006), as well as successful efforts to sequence large numbers of spruce (*Picea* spp.) and pine (*Pinus* spp.) expressed sequence tags (ESTs) (Ralph et al. 2006b) affords new opportunities in the study of cold hardiness in forest trees. Indeed, the first microarray study comparing the actively growing and dormant cambial meristem in *Populus tremula* reveals wide variation in gene expression between these two developmental states, some of which appears to be related to cold hardiness (Schrader et al. 2004). In addition, a 1.5K microarray was recently used to study the transcriptional response of apical buds in a Scots pine (*Pinus sylvestris*) provenance grown in three separate field sites along a latitudinal transect, and identified a number of candidate pine cold hardiness genes (Joosen et al. 2006). Although these studies provide the first insights into the genomics of cold hardiness and dormancy, a global picture of the temporal regulation of autumn gene expression in a conifer has yet to be reported. Similarly, the degree to which adaptation to local climate is manifested at the level of gene expression is unknown.

Conifers are among the most economically and ecologically important terrestrial plant species, and are also evolutionarily distinct from model angiosperms, having diverged ~380 million years ago (Kenrick and Crane 1997). This evolutionary distance makes selection of candidate genes in conifers based on sequence comparisons alone very challenging, a problem that is confounded by the evolutionary expansion of multigene families in conifers (Kinlaw and Neale 1997; Ahuja and Neale 2005). In the face of a changing climate, existing variation in the genes involved in adaptation to local climate provides the foundation for short-term survival and long-term adaptation (Aitken 1999). Conservation of these genetic resources is therefore crucial. In addition, rapid climate change may make marker-aided selection for traits related to local adaptation increasingly important in managed forests. Information on the genomic architecture of local adaptation to climate, including the number of genes involved, linkages among them, and pleiotropic effects on phenotypes will also shed greater light on the ability of populations to adapt rapidly to climate change than assuming a purely biometric approach based on quantitative genetic variation (Aitken et al., unpublished). In order to advance gene conservation and selective breeding efforts in an expedient, economic and rigorous way, we must first have an understanding of the genes involved. The goals of this study were threefold: (i) to analyze the reorganization of the transcriptome in Sitka spruce during autumn cold

acclimation; (ii) to identify transcripts that are differentially expressed among phenotypically divergent populations, and; (iii) to combine these transcriptional data with functional data from model species to identify a list of candidate genes for Sitka spruce cold acclimation that will form the basis for a future candidate gene-based association study. We used the Treenomix *Picea* spp. cDNA microarray comprising 21,840 unique elements to assay temporal and among-population variation in gene expression of Sitka spruce seedlings. Sampling of foliage for RNA extraction and cold hardiness phenotyping was carried out on five time points between August and December (2004) within a population originating at the approximate geographic centre of the species range (Prince Rupert, British Columbia). In addition, northern (Valdez, Alaska) and southern (Redwood, California) peripheral populations, which have markedly different cold acclimation phenotypes, were sampled at early and late autumn time points. Microarray expression data were validated for selected genes using real-time PCR.

2.2 Materials and methods

2.2.1 Plant material and tissue sampling

Foliage for RNA extraction and cold hardiness phenotyping was obtained from four-year-old Sitka spruce seedlings, which were grown from seed collected from natural populations spanning the species range in a raised-bed outdoor common garden in Vancouver, BC, Canada (49° N) (Mimura and Aitken 2007c). Needle tissues from current year upper lateral shoots were collected at five time points between late summer and early winter, 2004. Three of seventeen available populations were chosen for sampling based on geographic location and previously characterized genetic clines: Valdez, Alaska, USA (61° N) (AK), Prince Rupert, British Columbia, Canada (54° N) (BC) and Redwood, California, USA (41° N) (CA) (Figure 2.1a). Needle samples (~1g) were taken from eight individuals in the BC population on each of the five dates (Aug. 30, Oct. 18, Nov. 24, Dec. 1, and Dec. 13). To compare among-population gene expression, eight individuals from the AK and CA populations as well as eight additional individuals from the BC population were also sampled on the second and fourth dates. The common garden contained eight experimental blocks, and seedlings from each population were sampled from most blocks on each sampling date. Tissues were flash frozen in an N₂ vapor tank immediately upon collection, and subsequently stored at -80°C until processing.

In order to measure changes in gene expression triggered by both lengthening nights (phase I of cold acclimation) and the first sub-freezing temperatures (phase II of cold acclimation), we chose sampling dates that spanned both of these events. Cold acclimation begins in earnest when primary

growth has ceased and terminal bud formation has commenced ('bud set') (Weiser 1970). For the AK, BC and CA populations, the average dates of bud set in 2003 at the common garden site were July 2nd, August 28th and October 5th, respectively. As a compromise between these divergent dates, we conducted our first sample collection on August 30th. The second time point was approximately six weeks later, on October 18th, well before the arrival of the first freezing temperatures. We subsequently sampled just prior to the first frost event (which occurred on November 27th; sampling on November 22nd), and four days after that frost (December 1st). In order to capture any late responses, we conducted a final sampling 16 days after the first frost (December 13th).

2.2.2 Cold hardiness phenotyping

Cold hardiness for each individual sampled for RNA at each time point was measured using electrolytic leakage as a proxy for cell death (Hannerz et al. 1999). Briefly, 4-5 needles from the current year's growth of upper lateral shoots were cut into 0.5cm segments and frozen in 0.2 mL of water with a small amount of an ice nucleator (AgI). Samples were kept at 4°C overnight, and then the temperature was reduced by 4°C per hour and held for one hour at the selected test temperature. Samples were then thawed overnight at 4°C. A control was kept at 4°C throughout this process. After freezing, the electrolytic conductivity of the solution was measured. Frozen and control samples were then heat-killed at 95°C, and measured again. The following ratio expresses the result as an index of injury (I_t):

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

where $R_t = L_t/L_k$, $R_o = L_o/L_d$, L_t is the conductance of leachate from the sample frozen at temperature t , L_k is the conductance of the leachate from the sample frozen at temperature t and then heat-killed, L_o is the conductance of the leachate from the unfrozen sample, and L_d is the conductance of the leachate from the corresponding heat-killed, unfrozen sample.

Freeze testing temperatures were selected *a priori* for each sampling date based on cold hardiness in previous years (Mimura and Aitken 2007). We measured cold hardiness in all three populations (CA, BC, and AK) on all five dates, and tested at two temperatures on each date. These temperatures were: -8 and -11°C on August 30th, -10 and -14°C on October 18th, -10 and -20°C on November 22nd, -15 and -25°C on December 1st, and -15 and -25°C on December 13th. To provide a graphical representation of cold hardiness phenotypes across the acclimation period, mean I_t for 20 seedlings at or interpolated/extrapolated to -10°C are illustrated in Figure 2.1c. Actual values of I_t at each of two test temperatures for only those eight seedlings used for RNA extraction were used for

statistical analyses of phenotypic variation. On dates where three test temperatures were used, the two temperatures providing the greatest phenotypic variation were analyzed. Phenotypic data were subject to analyses of variance using the General Linear Model procedure of SAS (SAS 1989). Degrees of freedom were inadequate to test both experimental block and biological seedling pool simultaneously, and preliminary analyses of variance indicated that variation among experimental blocks was not significant. Sources of variation tested in analyses of variation among populations included population (CA, BC and AK), sampling date (Oct. 18th and Dec. 1st), test temperature nested within date, biological pool nested within date and population (two pools of four seedlings per population and sampling date), and population by date interaction. All effects were treated as fixed. In the analysis of temporal variance for the BC population, the model included sampling date, temperature within date (two temperatures), and biological pool nested within date (two pools of four seedlings per date). Reduced analyses of variance models were also run on each sampling date separately with terms involving date deleted from the above models. Least square phenotypic means were calculated for each population at each test temperature on each sampling date, and tests of differences between pairs of populations were conducted using a Bonferonni adjustment (SAS 1989).

2.2.3 RNA extraction, experimental design and microarray hybridization

To decrease the effects of biological variance among individual seedlings within populations, equal amounts of foliar tissue were pooled from four individuals prior to RNA extraction. Two pools were collected at each time point for each population, and total RNA was extracted following a previously published protocol (Kolosova et al. 2004). RNA quality was assessed by measuring spectral absorbance between 200 and 350 nm and by visual assessment on a 1% agarose gel. Because contaminants undetectable by these two methods can interfere with enzymatic manipulation, 5 µg of total RNA was reverse transcribed incorporating P-32 labeled dGTP. Resulting cDNA was run on a 1% alkaline buffer gel and visualized using a Storm phosphorimager (Amersham Biosciences, Piscataway, USA).

A factorial hybridization design with dye balance was chosen to assess gene expression among each of the five time points for the BC population (i.e., Aug 30 (BC1), Oct 18 (BC2), Nov 22 (BC3), Dec 1 (BC4), Dec 13 (BC5); Figure 2.2a). Among-population hybridizations were also conducted in a factorial fashion at the second and fourth time points (Oct 18 (CA2, BC2, AK2) and Dec 1 (CA4, BC4, AK4) (Figure 2.2b). Two biological replicates of this design were made for a total of 32 hybridizations. Hybridizations were performed using the Genisphere Array350 kit (Genisphere, Hatfield, USA). Hybridization conditions were the same as in Ralph *et al.* (2006), except that 40 µg of total RNA was used for each channel and hybridizations were incubated at 60°C. Complete details

of cDNA microarray fabrication and quality control will be described elsewhere (Ralph et al., unpublished). All microarray experiments were designed to comply with MIAME guidelines (Brazma et al. 2001). Raw and normalized data as well as TIFF images have been uploaded to the Gene Expression Omnibus under series accession number GSE8370. Sequences for array clones can be found by searching NCBI using the clone IDs given in Tables 2.2 and 2.3.

2.2.4 Microarray analysis

Slides were scanned and spot intensity quantified using ImaGene software (BioDiscovery, Inc., El Segundo, CA). To correct for background intensity, the lowest 10% of median foreground intensities was subtracted from the median foreground intensities. Data were then normalized by variance stabilizing normalization (VSN) to compensate for nonlinearity of intensity distributions (Huber et al. 2002). To identify significant changes in gene expression, a linear mixed-effects model was fit to the normalized intensities in the Cy3 and Cy5 channels of the 32 microarray slides. The model contained an adjustment for dye bias, an array effect indicating which Cy5/Cy3 pair was on each array, a treatment effect indicating sample population and time point, and a random effect to adjust for repeated measures on the same biological sample (Kerr et al. 2000). P-values were computed for each gene-by-treatment effect and Q-values were calculated to adjust for the false discovery rate (FDR) (Storey and Tibshirani 2003). All of the above statistical analyses of gene expression data were carried out using the R statistical package (www.r-project.org).

To identify themes in the time series expression data, we used the Cytoscape plug-in BiNGO to test for statistical overrepresentation of Gene Ontology (GO) terms within the GOSlim Plants ontology, among genes up- and downregulated two-fold between the first and fifth time points in the BC population (Shannon et al. 2003; Maere et al. 2005). BiNGO uses a Fisher's exact test to compute the probability that the number of differentially expressed genes in each GO category could have occurred by chance, given the total number of genes on the microarray in that GO category. In our case, this involved comparing the nearest *Arabidopsis* homologs for all upregulated genes to all *Arabidopsis* homologs on the microarray. The FDR was used to correct for multiple testing, with a cutoff of 0.05.

2.2.5 Real-time PCR

To validate microarray expression data, real-time PCR was conducted for eight genes, chosen to represent a variety of biological functions and expression levels on the microarray. Genes chosen for validation of the array results include homologs to *EARLII*, *GIGANTEA*, *CAXI*, *PHYA*, *CBL2*, *LEA*, and *MPK6*. Because of current interest in the role of *FLOWERING LOCUS T (FT)* in

daylength-mediated growth cessation responses in other species (see discussion section), we also used real-time PCR to determine the expression pattern of a spruce *FT/TFLI* homolog in our EST collection, though this transcript was not represented on the array. Gene expression was assayed across the five time points in the BC population. The same RNA pools (i.e., four individuals per pool per time point) were used as in the microarray component of the study. Prior to reverse transcription, 15µg total RNA for each of the five BC time points was treated with DNaseI (Invitrogen) according to manufacturer's instructions to remove genomic DNA. RNA was then divided into three aliquots of 5µg and cDNA synthesis completed for each aliquot independently using Superscript II reverse transcriptase (Invitrogen) with an oligo dT₁₂₋₁₈ primer. cDNA synthesis was assessed visually by gel electrophoresis prior to pooling of the three reactions.

Gene-specific primers were designed by aligning all BLAST matches in our EST collection with expect values less than E-50 to the target sequence. Primers were then placed in non-conserved regions of each target sequence. Primer specificity was assessed by visual inspection on a 2% agarose gel (single product of expected length) and melting curve analysis. Real-time PCR amplification conditions were identical to those described in Ralph *et al.* (2006), except that transcript abundance was normalized to translation initiation factor 5A (TIF5A, IS0013_F24, GenBank: DR448953) as its expression was invariant (data not shown).

2.3 Results

2.3.1 Cold hardiness

We observed highly significant ($p < 0.0001$) differences in cold hardiness across sampling dates within the central (BC) population (Table 2.1). Sub-freezing temperatures were not observed at our study site until November 27th (Figure 2.1b). However, both the central and northern populations (BC and AK) began developing cold hardiness well in advance of this date (Figure 2.1c). This result agrees with the well-established role of nightlength in regulating phase I of cold acclimation. Cold hardiness steadily increased through our sampling period in northern populations, and wide among-population variation was observed. Population differences were highly significant ($p < 0.0001$) on both Oct. 18th and Dec. 1st (Table 2.1). On Oct. 18th, the AK population had significantly higher hardiness (least square mean I_t = 20.3% across both test temperatures) than both BC and CA (I_t = 57.4 and 67.8% respectively), while BC and CA did not differ significantly. By Dec. 1st, all populations differed significantly, with least square mean I_t across both test temperatures of 10.5% for AK, 36.8% for BC, and 78.4% for CA.

2.3.2 Temporal variation in gene expression within-population

Differentially expressed genes were selected using two criteria: fold-change for at least one contrast in our time-series of $> 2.0x$ and $Q\text{-value} < 0.05$. To estimate the false discovery rate (FDR), we calculated $Q\text{-values}$ (Storey and Tibshirani 2003), and found, for example, the FDR for the BC5/BC1 contrast to be 5.6, 1.6, and 0.27% at $P = 0.05$, $P=0.01$, and $P=0.001$, respectively. Technical variance in our model generally exceeded biological variance, and although further biological replication may have been preferable, it would be unlikely to substantially enhance the accuracy of our gene expression estimates (data not shown).

Similar numbers of genes were up- and downregulated two-fold between the first (August 30th) and fifth (December 13th) time points in the central (BC) population. Out of 21,840 array elements, 1257 were upregulated and 967 downregulated (greater than two-fold changes; $Q < 0.05$) (Figure 2.2a). Many of these genes had no homology to *Arabidopsis*, including 549 of those upregulated, and 387 of those downregulated. Expression patterns across the five BC time points shows that transcripts for many genes began to increase by the second time point (Figure 2.2a). Similarly, genes that were downregulated generally began to decrease between the first and second time points. Among 1257 genes two-fold upregulated between the first and fifth time point, 309 were induced by the second time point (BC2/BC1), and 949 by the third (BC3/BC1). Few genes that were not activated early in the fall were subsequently induced after the third or fourth time point (i.e., following sub-freezing temperatures). In addition, most of the genes that were induced early plateaued in abundance by the fourth time point.

Ten ‘GOSlim Plants’ categories were overrepresented among genes upregulated in the BC population. These included, for example, ‘response to stress’ (45 genes), ‘response to abiotic stimulus’ (35 genes), membrane (148 genes), and ‘transport’ (51 genes) (Figure 2.3a). Only three GO terms were overrepresented among genes two-fold downregulated. These were ‘plastid’ (116 genes), ‘thylakoid’ (20 genes) and ‘secondary metabolism’ (19 genes) (Figure 2.3b).

Many of the induced genes were homologous to known or putative cold response genes in herbaceous annuals. We categorized genes that were upregulated in our study into nine broad categories that reflect the most prevalent themes in the cold acclimation literature. These are ‘stress responsive’, ‘carbohydrate metabolism’, ‘lipid metabolism and cell wall-related’, ‘light signaling’, ‘calcium signaling’, ‘phytohormone signaling’, ‘transcriptional regulation’, ‘posttranscriptional regulation and protein turnover’, and a ‘miscellaneous’ category for genes that did not easily fit into one of the above groups (Table 2.2). Stress-response genes were among the most strongly upregulated. These included many dehydrins and pathogenesis-related genes (e.g., chitinases, beta-1,3-glucanases, thaumatin family); genes related to oxidative stress such as peroxidases and

glutathione S-transferase; genes in the flavonoid/anthocyanin pathway (e.g., chalcone synthase, isoflavone reductase); and disease-responsive genes such as members of the dirigent, leucine-rich repeat, and hevein families (Table 2.2).

A number of carbohydrate metabolism genes were upregulated, including several involved in di- and trisaccharide synthesis (galactinol synthase, raffinose synthase, sucrose synthase, galactosyltransferase) and starch breakdown (alpha amylase) (Table 2.2). Upregulated lipid metabolism genes included a sphingolipid desaturase, squalene synthase and lipid transfer protein. The latter two were very strongly induced, with fold-increases of 16.9 and 34.4, respectively (BC5/BC1).

Signal transduction was one of the dominant themes among the genes upregulated in our study. It was unknown before we conducted this study whether signaling genes upregulated in spruce during cold acclimation would parallel those identified so far in *Arabidopsis*. We observed generally small (1.5 to 5- fold for BC5/BC1) but usually highly statistically significant increases in many signaling genes, some of which were homologous to genes with well-defined roles in the cold stress response of *Arabidopsis*. Others had less well understood functions. A small but significant increase in *PHYTOCHROME A* occurred between the first and second time point, whereas *PHYTOCHROME B* was downregulated across all sampling dates (Table 2.2). *GIGANTEA*, a gene known to be downstream of phytochromes under certain environmental stimuli (e.g., photoperiodic flowering) (Huq et al. 2000; Mizoguchi et al. 2005), was upregulated 5-fold. Our data also suggest a pivotal role for calcium signaling in spruce cold tolerance. Several genes involved in mediating the response to calcium were upregulated, including homologs to calmodulin 8, components of the calcineurin B-like (CBL) signaling module (CBL, CIPK) and the $\text{Ca}^{2+}/\text{H}^{+}$ antiporter *CAX1*. In addition, two annexin homologs and three C2-domain containing genes were induced (Table 2.2).

A large number of transcription factors (TFs) were upregulated over the course of our study, including those belonging to bZIP, bHLH, leucine zipper, myb, AP2, CCAAT-box, and NAC families (Table 2.2). Although fold-change values were generally modest, associated P-values were typically very low. Targeted breakdown of cellular proteins and transcripts appears to play a role in spruce cold acclimation as well. Several genes involved in ubiquitin-mediated protein degradation were induced, including members of the SCF complex (Kelch-repeat containing F-box, SKP1 interacting partner, COP10, Ring finger protein). A homolog to the RNA slicer *ARGONAUTE 1* (*AGO1*) was also moderately upregulated.

2.3.3. Among-population variation in gene expression

Among-population microarray hybridizations revealed substantial differential gene expression. Gene expression patterns corresponded to the genetic clines in cold hardiness-related traits among our study populations, with many genes upregulated in the northern (AK) and central (BC) compared to the southern (CA) population (Table 2.3 and Figure 2.2b). In the AK population, 326 genes were more than two-fold more highly expressed than in the CA population at the fourth time point. These genes also tended to be more highly expressed in BC compared to CA, but generally did not differ significantly between AK and BC. Interestingly, the expression level for most of these 326 genes did not vary significantly in the CA population between the second and fourth time point.

Many of the genes that were more highly expressed in the central and northern populations (BC and AK) relative to the southern CA population had annotations suggestive of involvement in cold hardiness. We grouped these genes in a similar fashion to the temporal component of our study, but collapsed some of the categories (e.g. ‘carbohydrate metabolism’ and ‘lipid metabolism’ together became ‘metabolism’) (Table 2.3). Examples of genes in the ‘stress responsive’ category that varied among populations include an ERD, a dehydrin, and a catalase. Several carbohydrate and lipid metabolism genes were also differentially expressed among populations, including a galactosyltransferase, a UDP-glucosyl transferase and a lipase. Among the genes involved in signal transduction and transcriptional regulation that were differentially expressed among populations were a CBL-like protein, calmodulin-binding protein, and a phospholipase.

2.3.4 Real-time PCR validation of microarray results

The genes chosen for microarray expression validation using real-time PCR (*CAX1*, *EARLII*, *CBL2*, *PHYA*, *GI*, *MPK6*, and *LEA*) generally had patterns of expression consistent with the microarray results (Figure 2.4). All of these genes were upregulated between the first and subsequent time points, and whereas some peaked in expression early (e.g., *MPK6*), others reached maximum expression late (e.g., *EARLII*). One particularly interesting gene was *CAX1*, which according to the microarray results peaked in expression at the third and fifth time points, with a decrease in relative expression at the fourth time point (Table 2.2). Although the decrease from time point three to four was not significant, our real-time data revealed the same pattern, and thus this cyclical expression does not appear to be an artifact. Real-time PCR measurements also suggest that a gene in our EST collection similar to *FLOWERING LOCUS T/TERMINAL FLOWER 1 (FT/TFLI)* was downregulated strongly between BC1 and subsequent time points.

2.4 Discussion

2.4.1 Cold hardiness

Phenotypic measurements of cold hardiness throughout our sampling period reinforce the importance of nightlength in determining the onset of cold acclimation in both the BC and AK populations, but suggest the CA population does not respond to this cue, in spite of experiencing the shortest daylength of its native environment on November 20th in Vancouver. Conversely, the AK population may not have a low-temperature requirement for entry to deep (phase II) cold acclimation, as this population achieved maximum cold hardiness well before the first frost. Measurements of cold injury taken at -25°C in the AK population showed similar and minimal levels of injury on November 22nd, December 1st and December 13th (data not shown).

2.4.2 Temporal within-population gene expression during cold acclimation

The winter period is one of bud dormancy and suspension of vegetative growth. Cold acclimation is inextricably linked to the concomitant processes of growth cessation and bud dormancy induction, and indeed the development of cold hardiness depends crucially on being preceded by budset. Although this makes it difficult or impossible to separate autumn gene expression changes related to these three processes in this study, the candidate genes we have identified, whether they are directly involved in cold hardiness, or indirectly through the establishment of the dormant state, are nevertheless relevant cold hardiness-related candidate genes.

Whereas repression of a range of cellular processes accompanies the period of cold acclimation and dormancy induction, low temperature stress necessitates induction of a complex suite of traits. These cold adaptive processes require a large investment of resources for gene expression and metabolic remodeling during a period when energy production via photosynthesis is decreasing (Clapham et al. 2001; Oquist et al. 2001). As a result, not only would it be inefficient to continue expressing genes for processes such as growth and reproduction, the limited energy available means that downregulation of these genes may be crucial for survival. Results of our microarray study through the period of cold acclimation demonstrate this redirection of resources. Gene Ontology categories statistically overrepresented among differentially expressed genes in our time series reveal that metabolic remodeling and stress response are dominant themes among upregulated genes (Figure 2.3a). Interestingly, both ‘transport’ and ‘transporter activity’ were overrepresented, suggesting that subcellular protein and metabolite targeting are important components of cold acclimation. In contrast, the downregulated autumn transcriptome was enriched in genes localized to the chloroplast, suggesting a redirection of resources away from photosynthesis. (Figure 2.3b). Although most

upregulated genes were induced early in our timecourse, many were not significantly upregulated until the third time point (BC3) (Figure 2.2a and Table 2.2). Because the seedlings experienced chilling (greater than 0°C but less than 5°C) temperatures after the second time point but prior to the third (Figure 2.1b), a role for low, above-freezing temperatures in enhancing nightlength-mediated transcription is possible. This pattern would also be congruent with positive feedback on expression of these genes following the initial signal (i.e., long nights). Separation of these environmental cues will require further study in controlled environments, and this work is in progress. In contrast, we did not observe substantial changes in gene expression following the first subfreezing temperatures, which occurred between time points three and four, suggesting that the subfreezing temperature cue for phase II of cold acclimation is not manifested at the level of large scale gene expression, and may trigger a more subtle response.

2.4.3 Stress response genes induced during cold acclimation

Plant stress response pathways often share both signaling components and response genes (Xiong et al. 2002). In some cases, apparently disparate stressors involve common mechanisms of damage. For example, ice propagation in the apoplast draws water out of the cell, and in this way mimics dehydration stress (Zwiazek et al. 2001). Dehydrins are one class of genes that respond to both of these stresses. It is thought that dehydrins function as chaperones to stabilize cellular macromolecules and prevent their coagulation as water exits the cell (Close 1997). In addition, *in vitro* studies suggest a role for dehydrins in stabilizing cellular membranes (Koag et al. 2003). Putative dehydrins were strongly upregulated (up to 30-fold) over the course of this study (Table 2.2). Several genes with less understood roles in the dehydration response were also induced, including an early response to dehydration (ERD) gene and members of the osmotin family.

Whereas dehydrins are thought to function as intracellular molecular chaperones as water is drawn from the cell, apoplastic antifreeze proteins limit extracellular ice propagation by adhering to the surface of nascent ice crystals and ice nucleators. Plant antifreeze proteins are homologous to three classes of pathogenesis-related proteins (i.e., chitinases, beta-1,3-glucanases and the thaumatin family) (Hon et al. 1995), and our expression results revealed candidate antifreeze genes in each of these categories (Table 2.2). Many of these genes were induced strongly by the second time point in the BC populations, indicating that they are likely regulated by photoperiod. In order to verify antifreeze activity, *in vitro* assays must be performed, as it is not known what sequence motifs, if any, differentiate antifreeze genes from pathogenesis-related genes (Hiilovaara-Teijo et al. 1999; Griffith and Yaish 2004)

Generation of radical oxygen species is another stress brought on by low temperatures, and damage to non-acclimated plants can be profound (e.g., degradation of lipid membranes) (Kendall and Mckersie 1989; Mckersie 1991). We observed substantial allocation of resources to transcription of genes that protect against oxidative damage. Upregulated genes include peroxidases, catalases, glutathione peroxidases and glutathione S-transferases (Table 2.2). In addition, several genes involved in flavonoid/anthocyanin biosynthesis were upregulated (e.g. chalcone synthase, isoflavone reductase). Induction of the flavonoid biosynthetic pathway could result in production of flavonoids that scavenge radical oxygen species, or of anthocyanins that block their generation.

2.4.4 Carbohydrate and lipid metabolism genes induced during cold acclimation

Many studies have addressed changes in carbohydrate composition in conifers during the fall, and point to an increase in non-reducing di- and trisaccharides, such as sucrose and raffinose (Pomeroy et al. 1970), and breakdown of polysaccharides, particularly starch (Alscher et al. 1989). Several explanations for these changes have been proposed. First, release of energy from starch could compensate for decreased photosynthetic activity; second, an increase in disaccharides provides free hydroxyl groups to replace hydrogen bonds with cellular macromolecules lost upon cellular dehydration; increasing intracellular solutes compensates for the osmotic potential generated across the plasma membrane upon extracellular ice formation; and finally, increased osmolarity in the cytoplasm depresses the intracellular freezing point (Zwiazek et al. 2001). We observed a 15-fold increase in the transcript for galactinol synthase, which catalyzes the first step in raffinose synthesis, and a 2.7-fold increase in a raffinose synthase-like transcript (Table 2.2). Several clones with high homology to both sucrose synthase and alpha amylase were also upregulated.

Cellular membranes are the primary site of freezing injury (Levitt 1980; Ziegler and Kandler 1980), and dehydration-related damage is in part a function of lipid composition (Uemura and Steponkus 1994). The ratio of saturated to unsaturated membrane lipids decreases during acclimation (Lynch and Steponkus 1987; Uemura et al. 1995). Genes that may be involved in membrane remodeling were upregulated during our timecourse, including a sphingolipid desaturase and several fatty acid synthases (Table 2.2). A lipid transfer protein similar to the vernalization-responsive *Arabidopsis* gene *EARLII* showed a 34- fold increase (Wilkosz and Schlappi 2000). In addition, we observed a 17-fold increase in squalene synthase 1. This result is interesting in light of a study of winter rye and spring oat (*Avena sativa*), which showed the former to have nearly four-fold more free sterols in the plasma membrane than the latter (Uemura and Steponkus 1994).

2.4.5 Signal transduction genes induced during cold acclimation

Phytochromes are thought to be the primary regulators of nightlength-mediated bud set and initiation of autumn cold acclimation in perennials (Howe et al. 1996; Horvath et al. 2003). Interestingly, phytochromes have also been recently implicated in upstream regulation of the cold stress machinery in *Arabidopsis*, suggesting a link between annual and perennial cold acclimation (Benedict et al., 2006). This role has been directly demonstrated in hybrid aspen trees (*Populus tremula* x *tremuloides*), wherein overexpression of *PHYA* blocked growth cessation and cold acclimation under short days (Olsen et al. 1997). Three features on our microarray are phytochrome homologs, two belonging to the *PHYA* subfamily, and one to the *PHYB* subfamily. We noted a transient increase of one of the *PHYA*-like genes, with an expression peak on the second time point and subsequent decline. The second *PHYA*-like gene and the *PHYB*-like gene were both downregulated across all five time points (data not shown). A homolog to *GIGANTEA* (*GI*) was also upregulated 5-fold during our study. Recently shown to be involved in the cold stress response in *Arabidopsis* (Shen et al. 2006), *GI* is better known as a key component of the photoperiodic flowering pathway (Mizoguchi et al. 2005). In the latter role, *GI* coordinates circadian expression of *CONSTANS* (*CO*) and *FLOWERING LOCUS T* (*FT*) (Mizoguchi et al., 2005). The upregulation of *GI* during cold acclimation in spruce suggests it is uncoupled from the *FT/CO* regulon, particularly in light of a recent study which showed that downregulation of *FT* is necessary for seasonal growth cessation in hybrid aspen (Bohlenius et al. 2006). In spite of the observed upregulation of *GI*, an *FT/TFL1* homolog identified in our EST collection was found to be strongly downregulated between time points one and five (see real-time PCR section). This sustained downregulation of *FT/TFL1* may reflect the latest stages of *FT* downregulation in response to critical nightlength (i.e., growth cessation), or it may suggest a specific role for *FT* in cold acclimation in conifers. In contrast to these results, Gyllenstrand et al. (2007) found a putative Norway spruce (*Picea abies* L. Karst.) *FT* homolog to be upregulated by short days. Further study is therefore needed to elucidate the role of this gene in growth cessation and possibly cold acclimation processes in conifers.

Elevation of cytosolic calcium is a first step in cold signaling in herbaceous annuals (Xiong et al. 2002). It has even been suggested that membrane calcium channels could be the elusive temperature sensor, as calcium influx occurs very early in the cold response, and changes in membrane fluidity as a function of decreased temperature could trigger these channels (Plieth et al. 1999; Xiong et al. 2002). As calcium plays a pivotal role in numerous other signaling cascades, and because multiple calcium transients may be necessary for cold signaling itself (Xiong et al. 2002), restoration of resting calcium levels following the initial signal is crucial. *CAX1*, a vacuolar $\text{Ca}^{2+}/\text{H}^{+}$ antiporter, performs this function in *Arabidopsis* (Catala et al. 2003). Upregulation of *CAX1*

following the initial calcium influx results in attenuation of the cold signaling machinery by downregulation of genes in the C-repeat binding factor/dehydration responsive element binding factor (CBF/DREB) pathway (Catala et al. 2003). Our data reveal periodic expression of a *CAX1* homolog, with expression peaks occurring at the third and fifth time points, and reduced expression at the fourth time point. This pattern would be congruent with a role for CAX1 in regulating both nightlength and freezing temperature-induced cold acclimation. Many calcium-binding proteins were similarly induced over the BC timecourse, including calcineurin B-like (CBL) genes and CBL-interacting protein kinases (CIPKs) (Table 2.2). CBL proteins interact specifically with CIPKs upon calcium influx (Shi et al. 1999), and in *Arabidopsis*, the CBL/CIPK signaling module is a crucial component of cold-stress signaling (Albrecht et al. 2003; Cheong et al. 2003). Other calcium-related genes upregulated in our study include a gene similar to *MAP KINASE 6* (*MPK6*), several annexins, a C2-domain containing gene, and two calmodulin genes.

2.4.6 Transcriptional regulation of gene expression during cold acclimation

Transcription factors are among the most studied cold-responsive genes. The CBF/DREB subfamily of APETELA (AP2) domain transcription factors, and their upstream regulator, ICE1, are essential for acclimation to low temperature in *Arabidopsis* (Stockinger et al. 1997; Jaglo-Ottosen et al. 1998; Chinnusamy et al. 2003). Although no CBF/DREB annotated transcription factors were induced over the course of our study at the two-fold threshold, several AP2-domain containing genes were upregulated, including ethylene response element binding factors (ERF) (Table 2.2). The role of ethylene in perennial cold acclimation is not well studied, but in winter rye, antifreeze proteins accumulate in response to ethylene treatment (Yu et al. 2001). We observed significant upregulation of four ERFs, one of which increased 9-fold between our first and fifth time points. Upregulation of ERF-like genes have also been observed during ecodormancy in leafy spurge (*Euphorbia esula*) (Horvath et al. 2006). Because the transcription factors regulating antifreeze gene expression have not been established, it will be interesting to investigate the possible role of these ERFs in regulating expression of spruce antifreeze genes.

2.4.7 Protein turnover and posttranscriptional gene silencing during cold acclimation

Molecular genetic studies of cold acclimation typically focus on genes that are induced. However, upon cold treatment, gene expression related to a range of cellular processes is repressed (Fowler and Thomashow 2002), presumably in part because resources normally allocated to expression of those traits are needed to mount the stress response. In addition, targeted breakdown of cellular proteins that inhibit cold signaling under normal growing conditions may be important to the

integration of abiotic cues governing cold acclimation. Indeed, expression of ubiquitin-conjugating enzymes is a crucial component of cold acclimation in *Arabidopsis* (Schwechheimer et al. 2002; Yan et al. 2003), and we observed induction of several such genes (Table 2.2). We also noted upregulation of *ARGONAUTE 1* (*AGO1*), which directs post-transcriptional gene silencing by cleaving target mRNAs after binding small RNAs (siRNAs and miRNAs) (Brodersen and Voinnet 2006). miRNAs are expressed in response to abiotic stresses including drought, cold and high salinity (Sunkar and Zhu 2004; Borsani et al. 2005; Brodersen and Voinnet 2006), and the induction of *AGO1* therefore suggests it may be worth investigating the role of small RNAs in spruce cold acclimation as well.

2.4.8 Genes with no homology

Although it is essential to view our results in light of functional data from model plants, it is also important to remember that many conifer expressed genes have no homology to genes in angiosperm sequence databases. Many of the spruce ESTs sequenced thus far fall into this category (Ralph et al., unpublished), and some of these were induced very strongly (up to 38-fold) during the course of our study (Table 2.2). In some cases, similarity to stress-inducible genes in other conifers was found through searches of the non-redundant collection of Genbank (e.g., WS00926_F21). However, in these cases little direct functional data exists. Given the ~380 million years of evolution separating gymnosperms and angiosperms, some of these may be conifer-specific cold hardiness genes, and though they are among the most difficult to study, they may also be novel and therefore among the most interesting.

2.4.9 Among-population variation in gene expression

Global gene expression profiles among genetically distinct taxa (populations or species) have been described for both plants and animals, and although widespread variation exists, methods to separate neutral from adaptive variation are in their infancy (Rifkin et al. 2003; Khaitovich et al. 2004; Yanai et al. 2004; Lemos et al. 2005; Lai et al. 2006). While some studies suggest the vast majority of expression divergence among taxa is the result of genetic drift (Khaitovich et al. 2004; Khaitovich et al. 2005), others have found extensive hallmarks of selection (Rifkin et al. 2003; Whitehead and Crawford 2006a; Gilad et al. 2008). One difficulty in evaluating these competing claims stems from the differences in evolutionary distance among the taxa compared within each study, where properties of among-taxa expression variation, and of the experimental method itself, depend strongly on the biological system (Whitehead and Crawford 2006b). It has therefore been suggested that comparisons will be most fruitful when made within a species or between closely related species (Whitehead and Crawford 2006b). In addition, closely related taxa are more likely to

share similar ecological contexts, which facilitates comparisons of expression variation with trait variation (Whitehead and Crawford 2006b; Whitehead and Crawford 2006a).

We have shown significant among-population differential gene expression along the latitudinal range of a single species, Sitka spruce. The clinal nature of this variation corresponds to a well studied genetic cline in cold hardiness, bud phenology and growth (Mimura and Aitken 2007c). Because cold hardiness comprises many component traits, differential gene expression could be expected to contribute to adaptive variation in cold hardiness in a variety of ways. For example, expression variation early in the fall could explain differences in timing of cold acclimation, whereas genes that are expressed more strongly in northern relative to southern populations in early winter could contribute to differences in maximum cold hardiness. We observed both of these patterns. Indeed, some of the genes already discussed were more highly expressed at the second time point in BC and AK populations than in the CA population. Surprisingly, many of the genes that were upregulated across the BC time-series did not exhibit population differences at either time point. There were, however, many genes that showed clinal expression patterns. There were 326 genes (among which 169 were annotated) more than two-fold more highly expressed in the AK compared to the CA population at the fourth time point (Figure 2.2b and Table 2.3). These genes also tended to be more highly expressed in BC compared to CA, particularly at the fourth time point, and did not vary to nearly the same extent between AK and BC as they did between CA and the other two populations.

Annotations for the genes that were more highly expressed in the AK relative to CA population suggest roles for some of them in cold tolerance. There were several genes involved in dehydrative stress tolerance, as well as genes with roles in both carbohydrate and lipid metabolism (Table 2.3). As in the temporal component of our study, many genes with no known homology were differentially expressed among-population. A subset of these can be found in Table 2.3. The variation in expression levels of these genes was particularly striking at the fourth time point, where they were 13-21-fold more abundant in the AK population relative to the CA population.

Finally, it is important to consider neutral divergence as a possible alternative explanation to the adaptive significance we have proposed for the among-population differential gene expression we observed. It has already been shown that population differentiation for cold hardiness in Sitka spruce ($Q_{st}=0.89$) is much stronger than for neutral microsatellite markers ($R_{st}=0.09$; $F_{st}=0.11$), reflecting strong divergent selection in spite of high gene flow (Mimura and Aitken 2007c). If genetic drift were driving gene expression divergence, we would expect to see the same strength and pattern of differentiation at neutral marker loci. However, R_{ST} values show that at neutral microsatellite loci, central British Columbia and northern California populations are substantially more similar to each other than to southern Alaska populations (Mimura and Aitken 2007c). Much of the among-

population gene expression differentiation we observed suggests more similar expression in BC and AK and divergent gene expression between these more northern provenances and their counterpart from California, particularly at the fourth timepoint. This is reflected in the late fall cold hardiness of these populations, where the LSmeans for cold injury for BC and AK (10.8% and 36.8%, respectively) suggest a relatively high degree of cold hardiness, whereas for CA an LSmean of 78.4% suggests significant cold injury at the test temperatures. Together these patterns provide additional evidence that the gene expression differences we observed are adaptive rather than the result of selectively neutral evolutionary processes.

2.4.10 Real-time PCR

The performance of the custom-built Treenomix *Picea* cDNA microarray used in this study has been previously evaluated using real-time PCR, and was found to be accurate even at moderate fold-differences (Ralph et al., unpublished). As this validation effort used tissue of clonal origin, but we employed seedlings grown from seed collected in wild populations, we chose to use real-time PCR to ensure the reliability of our results. Real-time PCR measurements of gene expression generally coincided well with our microarray results. Although microarrays can be susceptible to cross-hybridization, leading to artifactual results, each gene we validated using the more sensitive, quantitative technique of real-time PCR had an expression pattern similar to that measured on the array (Figure 2.4).

2.5 Conclusions

To the best of our knowledge, our study is the largest transcriptional profile of cold acclimation in a conifer, and provides many candidate genes for molecular dissection of this process. Until now, very little expression data existed in the literature connecting conifer genes to cold stress inducible homologs in model systems. As such, our study informs not only functional and population genomic studies of cold hardiness in spruce, but also those in other conifers for which substantial EST resources exist (e.g., *Pinus taeda*, *Pseudotsuga menziesii*). The candidate genes presented here will form the basis for a population genetic survey of nucleotide diversity, and ultimately lead to an association study linking genetic and phenotypic variation in cold hardiness. By undertaking targeted functional studies, and by characterizing the patterns of variation in the genes presented here, we can begin to understand the genetic networks and molecular adaptations that govern adaptation to local climate in general and cold hardiness in particular.

Table 2.1. Results of analyses of variance of index of injury (It) from freeze-testing of a) the British Columbia (BC) population on five test dates, and b) the California (CA), BC and Alaska (AK) populations on Oct. 18th and Dec. 1st.

Source of variation	DF	F	pr>F
BC population only			
Date	4	13.75	0.0001
Temperature (date)	5	2.66	0.0302
Pool (Date)	5	1.19	0.3257
Error	79		
CA, BC and AK populations			
Population	2	80.43	<0.0001
Date	1	3.15	0.0799
Population x Date	2	6.03	0.0037
Temperature (Date)	2	8.02	0.0007
Population x Temperature (Date)	4	3.69	0.0084
Pool (Population x Date)	6	0.67	0.6722
Error	78		

Table 2.2. Functional categorization of selected genes with $> 2.0x$ fold-change for at least one contrast within the BC time series and $Q < 0.05$. Fold-change (FC) values are given as ratios of the 2nd through 5th time points (BC2, Oct 18; BC3, Nov 22; BC4, Dec 1; and BC5, Dec 13) to time point 1 (BC1, Aug 30). Fold-change values given in italics have $Q < 0.05$, and those given in bold have $Q < 0.01$; all other fold-change values are not significant for the given contrast. Annotations are given for translated BLAST vs. *Arabidopsis thaliana* and vs. the non-redundant collection at the NCBI. In the case of the latter, the organism for each hit given in parentheses.

Clone ID	BLASTX vs. <i>Arabidopsis</i>	AGI code	BLASTX vs. NCBI (non-redundant)	NCBI accession	Fold-change values				
					BC2	BC3	BC4	BC5	
Stress responsive									
WS0046_E20	Dehydrin	At4g38410	Dehydrin [Picea abies]	AAX92687	7.6	21.1	20.6	25.6	
WS0047_B20	LEA	At1g01470	LEA [Pseudotsuga menziesii]	CAA10047	8.7	20.6	20.9	17.7	
WS0097_C01	Hevein-like	At3g04720	Pathogenesis-related [Capsicum chinense]	BAD11073	3.6	20.3	16.7	15.1	
WS0109_G23	ERD	At3g51250	Senescence-associated [Hemerocallis hybrid cultivar]	AAC34857	3.4	9.5	12.9	14.9	
WS00930_N09	Beta-1,3-glucanase	At2g01630	Glucanase-like protein [Thuja occidentalis]	AAV66572	3.1	8.2	13.2	13.9	
WS00923_I02	Osmotin	At4g11650	PR-5 thaumatin-like [Pseudotsuga menziesii]	AAQ84890	3.4	6.9	13.9	10.8	
WS00924_G17	Glutathione S-transferase	At2g30860	Glutathione S-transferase GST 22 [Glycine max]	AF243377	2.4	6.4	10.8	9.5	
WS00715_B10	Thaumatin family	At1g19320	Thaumatin-like [Pseudotsuga menziesii]	AAV74248	4.2	8.7	8.6	9.3	
WS00922_F20	Catalase	At4g35090	Catalase 1 [Zantedeschia aethiopica]	AF207906	4.8	7.0	7.3	7.9	
WS0024_K18	Chalcone synthase	At5g13930	Chalcone synthase [Abies alba]	ABD38596	2.0	5.3	8.7	7.7	
WS0034_J13	RCI	At3g05880	RCI2A [Arabidopsis thaliana]	NP_187239	7.4	9.0	10.0	7.3	
WS00922_B21	Basic chitinase	At3g12500	Endochitinase [Musa acuminata]	AF416677	2.1	4.4	5.5	6.5	
WS00928_M02	Isoflavone reductase	At1g75290	Isoflavone reductase-like [Cryptomeria japonica]	AAK27264	2.2	4.5	5.5	5.7	
WS00911_I09	Dirigent family	At1g64160	Dirigent-like protein [Tsuga heterophylla]	AF210071	1.7	4.1	6.0	4.7	
WS0075_N02	Peroxidase	At1g71695	Peroxidase [Picea abies]	CAD92858	2.4	3.3	3.4	4.2	
IS0014_D05	Glutathione peroxidase	At4g11600	Glutathione peroxidase-like [Spinacia oleracea]	O23814	1.8	4.0	5.4	3.8	
WS00810_O07	Leucine-rich repeat	At1g15740	Leucine-rich repeat family [Oryza sativa]	NP_001047418	1.2	2.4	2.4	3.1	
Carbohydrate metabolism									
WS00816_J13	Galactinol synthase	At1g60470	Galactinol synthase [Ajuga reptans]	CAB51533	2.1	11.4	18.0	15.0	
WS0064_L12	Galactosyltransferase	At2g26100	beta 1,3-glycosyltransferase-like [Oryza sativa]	CAD44839	2.1	8.8	10.3	9.4	
WS0083_A08	Sugar transporter	At4g02050	Monosaccharid transporter [Nicotiana tabacum]	CAA47324	1.7	4.9	6.9	5.5	
WS01028_M08	Sucrose synthase	At4g02280	Sucrose synthase [Pinus taeda]	ABR15470	2.1	4.0	3.7	4.3	
WS00915_H03	Glycosyl hydrolase family 1	At1g26560	Beta-glucosidase [Pinus contorta]	AAC69619	1.4	2.4	4.0	3.2	
WS01021_A02	Raffinose synthase	At5g20250	Raffinose synthase [Oryza sativa]	AA177910	1.4	2.3	3.1	2.7	
WS00917_K19	Alpha-amylase	At1g69830	Plasitid alpha-amylase [Actinidia chinensis]	AAX33233	1.5	2.5	2.2	2.2	

Table 2.2. Cont'd

Clone ID	BLASTX vs. <i>Arabidopsis</i>	AGI code	BLASTX vs. NCBI (non-redundant)	NCBI accession	Fold-change values				
					BC2	BC3	BC4	BC5	
<i>Lipid metabolism and cell wall-related</i>									
WS00939_C18	LTP (EARL11)	At1g62500	Plant lipid transfer [Medicago truncatula]	ABO83639	3.0	22.7	29.5	34.4	
WS0044_O08	PAL	At2g37040	PAL [Equisetum arvense]	AAW80639	3.8	12.6	21.1	27.8	
WS01035_P02	Squalene synthase	At4g34640	Squalene synthase [Panax notoginseng]	ABA29019	1.6	6.9	13.0	16.9	
WS00733_G19	COMT	At1g51990	O-methyltransferase, family 2 [Medicago truncatula]	ABD32718	5.1	8.1	9.0	9.6	
WS00946_M24	Expansin	At4g38400	Expansin-like [Solanum tuberosum]	ABC47126	1.4	2.6	3.8	5.8	
WS00930_N07	4-coumarate--CoA ligase	At3g21240	4-coumarate--CoA ligase [Pinus taeda]	P41636	1.7	2.9	4.0	4.2	
WS00824_K23	Delta-8 sphingolipid desaturase	At2g46210	Fatty acid desaturase [Marchantia polymorpha]	AAT85663	1.9	3.3	3.2	3.9	
<i>Calcium signaling</i>									
WS0079_I12	CAX1	At5g01490	Ca2+/H+ exchanger [Vigna radiata]	BAA25753	1.5	17.6	4.6	14.8	
WS01016_C12	Calcium-binding EF hand	At1g53210	Drought-induced protein [Oryza sativa]	BAD19956	3.0	4.1	5.1	4.3	
WS00824_I18	C2-domain containing	At5g11100	Ca2+-dependent lipid-binding protein [Oryza sativa]	NP_001061492	1.5	4.0	5.1	3.9	
WS00818_N13	Annexin	At5g10220	Annexin [Zea mays]	CAA66901	2.3	3.0	4.6	3.8	
WS0055_B09	Calmodulin	At3g22930	Calmodulin [Cryptomeria japonica]	BAF31994	1.6	1.8	2.6	3.1	
WS00925_D16	CBL	At5g55990	CBL [Populus trichocarpa]	ABO43662	0.9	1.7	1.9	2.4	
WS00819_N13	CIPK	At1g30270	CIPK [Populus trichocarpa]	ABJ91220	1.3	2.2	2.2	2.1	
<i>Phytohormone signaling</i>									
WS0074_E17	Dormancy/auxin associated	At1g28330	Auxin-repressed protein [Manihot esculenta]	AAX84677	4.0	15.9	12.2	13.4	
WS02610_F23	Auxin down-regulated	At3g22850	Hypothetical protein [Vitis vinifera]	CAN71784	2.6	9.7	9.2	7.6	
WS00733_N14	Auxin efflux carrier	At5g01990	Auxin efflux carrier [Oryza sativa]	BAD73344	2.8	5.6	7.6	6.5	
WS01011_O03	Auxin-responsive	At5g35735	Auxin-induced [Oryza sativa]	BAC75413	2.3	3.1	4.7	4.4	
WS00937_G05	Auxin-responsive	At5g54510	Auxin-induced [Pinus pinaster]	CAJ14972	1.4	4.1	3.2	4.4	
WS00918_P16	Gibberellin response modulator	At1g14920	Gibberellic acid-insensitive protein [Vitis vinifera]	Q8S4W7	1.3	4.4	3.4	3.6	
<i>Transcription regulation</i>									
WS00811_F18	NAM (NAC domain)	At1g52890	NAM protein [Oryza sativa]	NP_001049997	4.8	12.2	13.2	15.0	
WS00925_L16	ERF	At1g50640	ERF [Nicotiana tabacum]	BAA76734	1.9	5.8	12.8	9.0	
WS00728_K02	MYB family	At5g47390	MYB transcription factor [Glycine max]	ABH02845	1.9	6.1	6.6	5.8	
WS00917_I23	AREB	At3g56850	AREB [Oryza sativa]	AAT77290	1.9	2.6	2.7	2.7	

Table 2.2. Cont'd

Clone ID	BLASTX vs. <i>Arabidopsis</i>	AGI code	BLASTX vs. NCBI (non-redundant)	NCBI accession	Fold-change values				
					BC2	BC3	BC4	BC5	
Post-transcriptional regulation and protein turnover									
WS00930_P10	Serine carboxypeptidase	At1g15000	Serine carboxypeptidase [Oryza sativa]	BAA94235	1.2	10.1	22.4	14.0	
WS0076_I24	C3HC4-type zinc finger	At1g72310	RING-H2 [Populus alba x Populus tremula]	AAW33880	1.9	2.1	2.9	3.1	
WS00716_J15	Ubiquitin-conjugating enzyme	At1g45050	Ubiquitin-conjugating enzyme [Medicago truncatula]	ABE89644	1.3	2.0	1.6	2.8	
WS00112_G14	Aspartyl protease family	At1g62290	Aspartic protease [Oryza sativa]	BAA06876	1.5	2.0	1.8	2.1	
WS0091_H14	Kelch repeat-containing	At1g15670	Kelch repeat-containing F-box-like [Oryza sativa]	BAD25000	0.9	1.9	1.9	2.0	
WS0084_C03	Argonaute (AGO1)	At1g31280	Piwi domain containing [Oryza sativa]	AAS01930	1.3	1.7	1.6	2.0	
Miscellaneous									
WS01012_F14	AAA-type ATPase	At4g28000	AAA ATPase [Medicago truncatula]	ABE83074	2.1	12.6	15.8	30.6	
WS0063_I19	Phospholipase A2	At2g06925	Phospholipase A2 [Nicotiana tabacum]	BAD90927	2.3	11.6	13.6	11.5	
WS0094_H16	ABC transporter	At1g67940	Multidrug resistance protein [Cicer arietinum]	BAA76420	1.9	8.0	14.2	10.6	
WS0106_F09	Glutamate receptor family	At2g17260	Ligand gated channel-like [Brassica napus]	AF109392	1.7	5.8	5.9	4.8	
WS0033_C24	Gigantea	At1g22770	Putative gigantea [Picea abies]	CAK26425	2.7	4.1	4.3	4.9	
WS01039_D02	MAP kinase 6 (MPK6)	At2g43790	Mitogen-activated protein kinase 2 [Glycine max]	AF329506	1.7	2.5	2.4	2.8	
WS0021_H24	FK506-binding protein 15	At3g25220	Putative immunophilin [Hordeum vulgare]	CAD42633	1.6	1.9	2.1	2.1	
WS0101_F05	Phytochrome A	At1g09570	Phytochrome N [Pinus sylvestris]	CAC11136	2.2	1.6	1.4	1.3	
No hit									
WS0108_H04	No significant hit	N/A	N/A	N/A	5.2	34.5	39.9	30.4	
WS00914_D20	No significant hit	N/A	N/A	N/A	5.2	38.8	37.0	28.9	
WS0102_I01	No significant hit	N/A	N/A	N/A	3.3	22.4	31.4	26.2	
WS00926_F21	No significant hit	N/A	N/A	N/A	8.9	27.2	28.4	25.5	
WS0107_M21	No significant hit	N/A	N/A	N/A	9.3	20.2	26.3	24.5	
WS0041_B09	No significant hit	N/A	N/A	N/A	4.9	16.5	16.5	22.4	
Abbreviations: ERD, early response to dehydrative stress; RCI, rare cold-inducible; NAM, no apical meristem; ERF, ethylene response factor; AREB, ABA response element binding factor; COMT, Caffeic acid O-methyltransferase; LTP, lipid transfer protein; PAL, phenylalanine ammonia lyase; CBL, calcineurin B-like protein; CIPK, CBL interacting protein kinase; ABC, ATP-binding cassette; LEA, late embryogenesis abundant									

Table 2.3. Functional categorization of selected genes with AK4/CA4 > 2.0 and Q < 0.05; Untransformed fold-change (FC) values are given as ratios of northern (AK and BC) to southern (CA) population at the 2nd (Oct. 18, 2004) and 4th (Dec. 1, 2004) timepoints. Fold-change values given in italics have Q < 0.05, and those given in bold have Q < 0.01; all other fold-change values are not significant for the given contrast.

Clone ID	BLASTX vs. <i>Arabidopsis</i>	AGI code	BLASTX vs. NCBI (non-redundant)	NCBI Accession	Fold-change values			
					AK2/CA2	AK4/CA4	BC2/CA2	BC4/CA4
Stress responsive								
WS00927_G12	No significant hit	N/A	LEA protein [Picea glauca]	AAB01550	2.4	12.8	1.3	4.7
WS00923_A21	Terpene synthase/cyclase	At3g14490	(-)-limonene synthase [Picea abies]	AAS47694	6.8	10.5	4.4	7.9
WS00922_F20	Catalase 2	At4g35090	Catalase [Prunus persica]	CAD42909	7.4	7.9	5.3	8.0
WS0109_G23	ERD	At3g51250	Senescence-associated [Medicago truncatula]	ABE77914	7.3	3.7	4.2	2.6
WS0046_E20	Dehydrin	At4g38410	Dehydrin 1 [Picea abies]	AAX92687	9.2	2.7	7.4	2.2
WS00935_H19	Heat shock protein	At1g56410	Cytosolic heat shock protein [Spinacia oleracea]	AAB88132	2.2	2.2	1.8	2.1
WS00811_K09	Leucine-rich repeat	At1g80630	Cyclin-like F-box [Medicago truncatula]	ABE81084	3.2	2.2	2.0	1.8
Metabolism								
WS00931_D24	Lipase class 3	At3g14360	Lipase [Ricinus communis]	AAV66577	1.9	11.1	0.9	1.0
WS0064_L12	Galactosyltransferase	At2g26100	Galactosyltransferase [Hordeum vulgare]	ABL11234	2.3	3.4	1.6	3.0
WS00928_N23	UDP-glucosyl transferase	At2g36970	UDP-glucuronosyl/UDP-glucosyltransferase [Medicago truncatula]	ABE90468	3.7	2.9	1.6	1.9
WS00733_G19	COMT	At1g51990	O-methyltransferase [Medicago truncatula]	ABD32718	3.4	2.6	2.1	2.4
WS0099_O10	Malate oxidoreductase	At2g13560	Putative malate dehydrogenase [Oryza sativa]	NP_001059700	3.3	2.2	2.8	2.1
WS0044_O07	Family II lipases	At1g71120	Lipolytic enzyme [Medicago truncatula]	ABO83318	2.0	2.1	1.0	1.6
Signal transduction								
WS0106_F09	Glutamate receptor family	At2g17260	Ligand gated channel-like protein [Brassica napus]	AF109392_1	2.4	5.0	1.6	2.3
WS0063_I19	Phospholipase A2	At2g06925	Phospholipase A2 [Nicotiana tabacum]	BAD90927	3.1	3.3	1.6	3.4
IS0012_C06	DAG acyltransferase	At3g51520	DAG acyltransferase [Medicago truncatula]	ABO83262	2.0	2.7	1.3	1.7
WS00928_I04	AMP-dependent synthetase	At3g16910	AMP-binding enzyme [Oryza sativa]	ABF95517	2.9	2.5	1.4	1.7
WS0086_C05	Calcineurin B-like protein	At2g31800	Ankyrin protein kinase [Brassica napus]	AAT94403	1.6	2.3	1.1	1.7
Transcriptional regulation and protein turnover								
WS0101_P19	Serine carboxypeptidase	At5g08260	Putative carboxypeptidase D [Oryza sativa]	BAD25095	5.3	4.7	2.6	3.2
WS00112_G14	Aspartyl protease family	At1g62290	Aspartic protease [Oryza sativa]	BAA06876	1.7	2.4	1.5	1.7
WS0109_C08	Metalloprotease	At1g17870	Hypothetical protein [Vitis vinifera]	CAN73523	2.4	2.3	1.5	1.7
WS00933_K20	Metallo-beta-lactamase	At4g33540	Beta-lactamase-like [Anabaena variabilis]	YP_321165	2.2	2.2	2.0	1.9

Table 2.3. Cont'd

Clone ID	BLASTX vs. <i>Arabidopsis</i>	AGI code	BLASTX vs. NCBI (non-redundant)	NCBI Accession	Fold-change values			
					AK2/CA2	AK4/CA4	BC2/CA2	BC4/CA4
<i>Miscellaneous</i>								
WS01011_B24	ACT domain-containing	At1g12420	Amino acid-binding ACT [Medicago truncatula]	ABE90521	1.6	9.1	1.3	3.8
WS01012_F14	AAA-type ATPase	At4g28000	AAA ATPase [Medicago truncatula]	ABE83074	8.1	6.4	3.0	4.7
WS00946_G15	PPR	At3g02650	PPR repeat-containing protein-like [Oryza sativa]	BAC99540	3.6	6.2	2.0	1.2
WS00937_G05	Auxin-responsive	At5g54510	Auxin-induced [Pinus pinaster]		2.9	4.2	1.5	2.1
WS0078_P14	LOB domain	At2g28500	LOB domain protein 1 [Solanum demissum]	AAT40528	2.6	3.6	1.8	2.2
WS00922_D24	Acyl-CoA binding protein	At5g53470	Membrane acyl-CoA binding protein [Agave americana]	AA.T81164	3.5	3.0	1.8	2.1
WS0071_O08	ABA responsive	At5g13200	ABA-responsive [Oryza sativa]	ABA98234	4.0	2.7	2.1	1.8
WS0101_F10	Cytochrome P450 family	At5g36110	Cytochrome P450 [Medicago truncatula]	ABC59076	2.2	2.5	1.2	2.3
<i>No hit</i>								
WS00923_F16	No significant hit	N/A	N/A	N/A	11.9	13.6	1.1	1.8
WS0107_L16	No significant hit	N/A	N/A	N/A	5.5	21.3	1.2	1.8
Abbreviations: LEA, late embryogenesis abundant; ERD, early response to dehydrative stress; COMT, Caffeic acid O-methyltransferase; TPR, Tetratricopeptide repeat; PPR, Pentatricopeptide repeat; LOB, Lateral organ boundaries; DAG, Diacylglycerol								

Figure 2.1. Distribution of Sitka spruce, origins of study populations, and location of common garden (a). Daily maximum and minimum temperatures measured at field site, located at Vancouver, British Columbia, Canada. Dashed lines indicate 5°C and 0°C (b). Mean percent cold injury at -10°C for study populations, measured on each of five sampling dates (c).

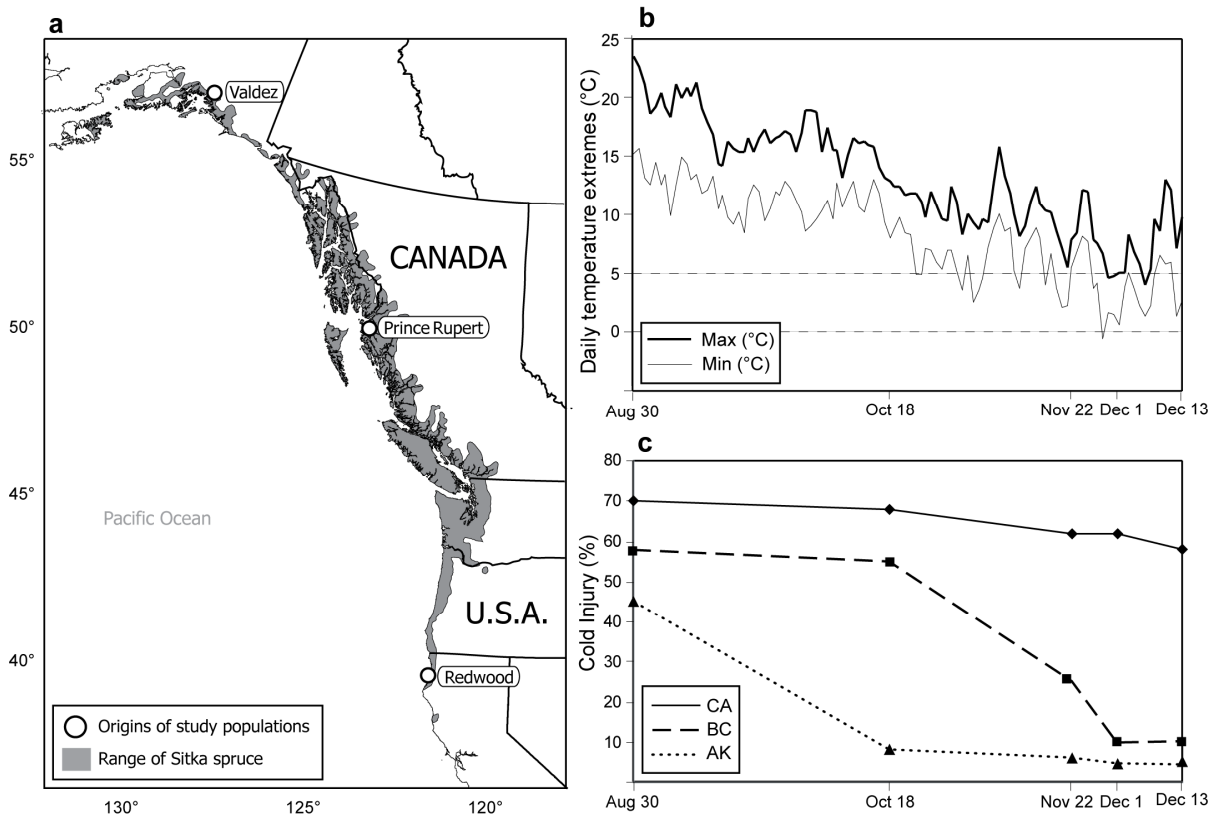


Figure 2.2. Illustration of experimental design. Arrows indicate hybridizations that were made among time points (a) and populations (b). Bold and italicized numbers indicate genes that were two-fold up- and downregulated ($Q < 0.05$), respectively, between the time point at which the arrow originates and the time point at which it terminates.

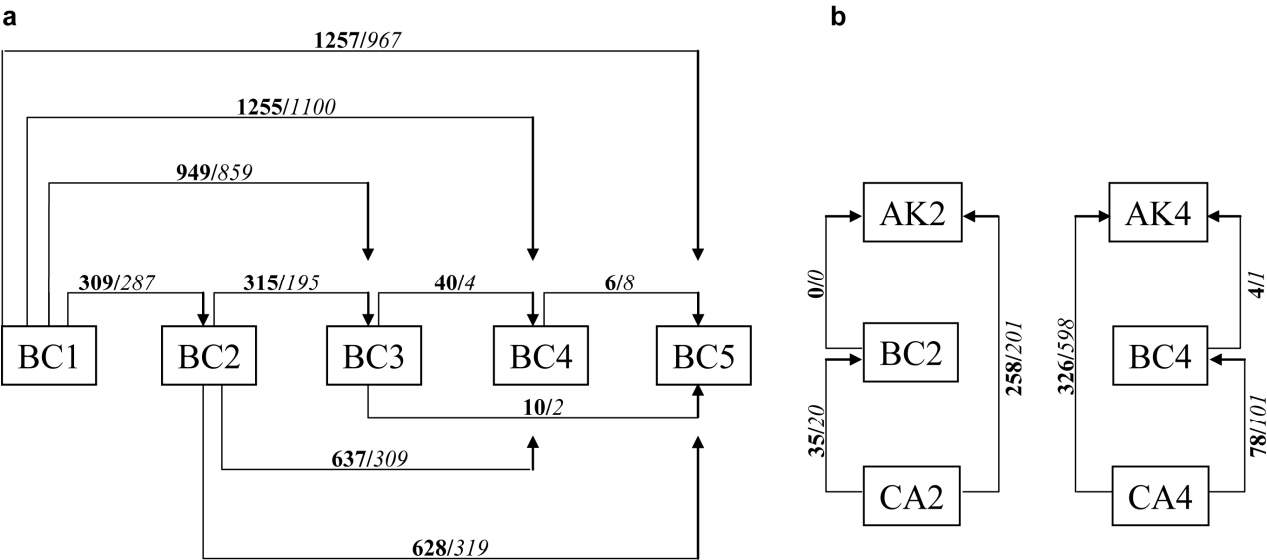


Figure 2.3. Hierarchies of Gene Ontology terms statistically overrepresented (FDR < 0.05) among genes upregulated two-fold between BC1 and BC5 (a) and genes downregulated two-fold between BC1 and BC5 (b). The size of each circle indicates the relative number of genes in each category and the colour scale indicates FDR adjusted P-values. GO terms with no shading were not significant, and were included to illustrate the hierarchy for child terms that were significant.

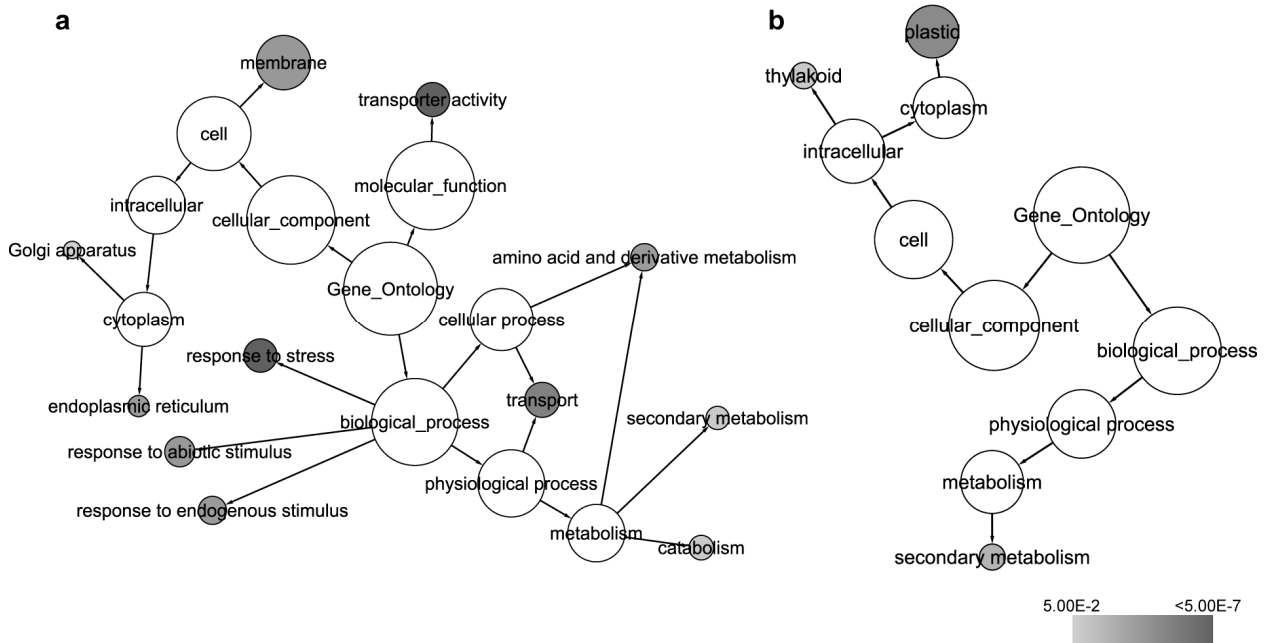
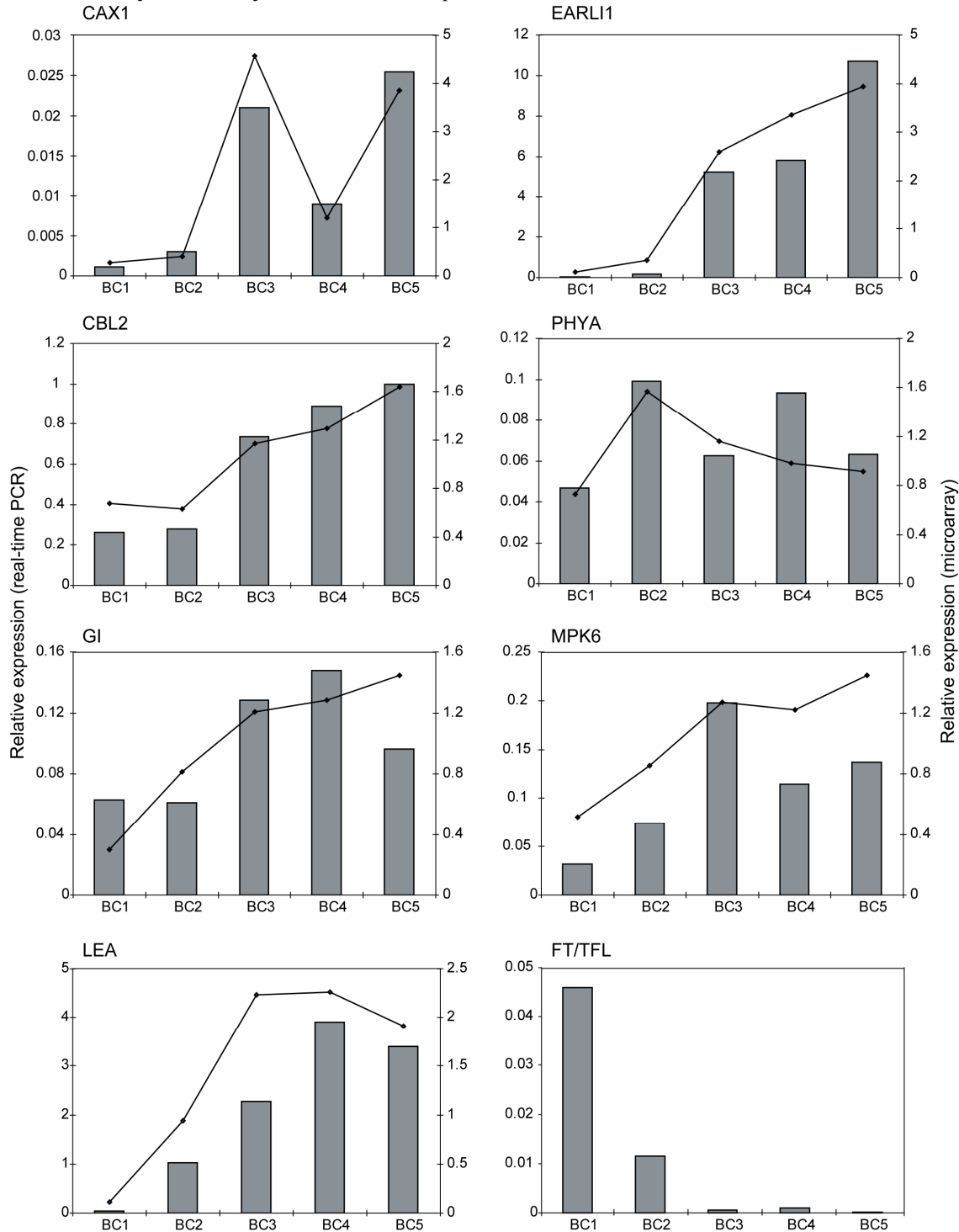


Figure 2.4. Real-time PCR analysis of gene expression among five time points in the BC population for *CAX1* (WS0079_I12), *EARL1* (WS00939_C18), *CBL2* (WS00925_D16), *PHYA* (WS01021_F05), *GIGANTEA* (WS0033_C24), *MPK6* (WS01039_DO2), *LEA* (WS0047_B20), and *FT/TFL1* (WS02738_CO3). Bars indicate real-time PCR expression measurements (relative to the housekeeping gene, TIF5A) and lines indicate corresponding microarray data (given as natural log difference from mean of the five BC time points). Note that because *FT/TFL1* was not represented on the microarray, it has only real-time PCR expression measurements.



2.6 Literature cited

- Ahuja, M. R., and D. B. Neale. 2005. Evolution of genome size in conifers. *Silvae Genetica* **54**:126-137.
- Aitken, S. N. 1999. Conserving adaptive variation in forest ecosystems. *Journal of Sustainable Forestry* **8**:1-10.
- Aitken, S. N., and W. T. Adams. 1996. Genetics of fall and winter cold hardiness of coastal Douglas-fir in Oregon. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere* **26**:1828-1837.
- Albrecht, V., S. Weinl, D. Blazevic et al. 2003. The calcium sensor CBL1 integrates plant responses to abiotic stresses. *Plant Journal* **36**:457-470.
- Alscher, R. G., R. G. Amundson, J. R. Cumming, S. Fellows, J. Fincher, G. Rubin, P. Vanleuken, and L. H. Weinstein. 1989. Seasonal changes in the pigments, carbohydrates and growth of red spruce as affected by ozone. *New Phytologist* **113**:211-223.
- Bohlenius, H., T. Huang, L. Charbonnel-Campaa, A. M. Brunner, S. Jansson, S. H. Strauss, and O. Nilsson. 2006. CO/FT regulatory module controls timing of flowering and seasonal growth cessation in trees. *Science* **312**:1040-1043.
- Borsani, O., J. Zhu, P. E. Verslues, R. Sunkar, and J. K. Zhu. 2005. Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell* **123**:1279-1291.
- Brazma, A., P. Hingamp, J. Quackenbush et al. 2001. Minimum information about a microarray experiment (MIAME) - toward standards for microarray data. *Nature Genetics* **29**:365-371.
- Brodersen, P., and O. Voinnet. 2006. The diversity of RNA silencing pathways in plants. *Trends in Genetics* **22**:268-280.
- Cannell, M. G. R., and L. J. Sheppard. 1982. Seasonal changes in the frost hardiness of provenances of *Picea sitchensis* in Scotland. *Forestry* **55**:137-153.
- Cannell, M. G. R., P. M. Tabbush, J. D. Deans, M. K. Hollingsworth, L. J. Sheppard, J. J. Philipson, and M. B. Murray. 1990. Sitka spruce and Douglas-Fir seedlings in the nursery and in cold storage - Root growth potential, carbohydrate content, dormancy, frost hardiness and mitotic index. *Forestry* **63**:9-27.
- Catala, R., E. Santos, J. M. Alonso, J. R. Ecker, J. M. Martinez-Zapater, and J. Salinas. 2003. Mutations in the Ca²⁺/H⁺ transporter CAX1 increase CBF/DREB1 expression and the cold-acclimation response in Arabidopsis. *Plant Cell* **15**:2940-2951.
- Cheong, Y. H., K. N. Kim, G. K. Pandey, R. Gupta, J. J. Grant, and S. Luan. 2003. CBL1, a calcium sensor that differentially regulates salt, drought, and cold responses in Arabidopsis. *Plant Cell* **15**:1833-1845.
- Chinnusamy, V., M. Ohta, S. Kanrar, B. H. Lee, X. H. Hong, M. Agarwal, and J. K. Zhu. 2003. ICE1: a regulator of cold-induced transcriptome and freezing tolerance in Arabidopsis. *Genes & Development* **17**:1043-1054.
- Clapham, D. H., I. Ekberg, C. H. Little, and O. Savolainen. 2001. Molecular Biology of Conifer Frost Tolerance and Potential Applications to Tree Breeding. Pp. 187-219 in F. J. Bigras, and S. J. Colombo, eds. *Conifer Cold Hardiness*. Kluwer Academic Press, Dordrecht.
- Close, T. J. 1997. Dehydrins: A commonality in the response of plants to dehydration and low temperature. *Physiologia Plantarum* **100**:291-296.
- Fowler, S., and M. F. Thomashow. 2002. Arabidopsis transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. *Plant Cell* **14**:1675-1690.
- Gilad, Y., S. A. Rifkin, and J. K. Pritchard. 2008. Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends in Genetics* **24**:408-415.
- Griffith, M., and M. W. Yaish. 2004. Antifreeze proteins in overwintering plants: a tale of two activities. *Trends in Plant Science* **9**:399-405.

- Gyllenstrand, N., D. Clapham, T. Kallman, and U. Lagercrantz. 2007. A Norway spruce FLOWERING LOCUS T homolog is implicated in control of growth rhythm in conifers. *Plant Physiology* **144**:248-257.
- Hannerz, M., S. N. Aitken, J. N. King, and S. Budge. 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.
- Hiilovaara-Teijo, M., A. Hannukkala, M. Griffith, X. M. Yu, and K. Pihakaski-Maunsbach. 1999. Snow-mold-induced apoplastic proteins in winter rye leaves lack antifreeze activity. *Plant Physiology* **121**:665-674.
- Hon, W. C., M. Griffith, A. Mlynarz, Y. C. Kwok, and D. S. Yang. 1995. Antifreeze proteins in winter rye are similar to pathogenesis-related proteins. *Plant Physiology* **109**:879-889.
- Horvath, D. P., J. V. Anderson, W. S. Chao, and M. E. Foley. 2003. Knowing when to grow: signals regulating bud dormancy. *Trends in Plant Science* **8**:534-540.
- Horvath, D. P., J. V. Anderson, M. Soto-Suarez, and W. S. Chao. 2006. Transcriptome analysis of leafy spurge (*Euphorbia esula*) crown buds during shifts in well-defined phases of dormancy. *Weed Science* **54**:821-827.
- Howe, G. T., S. N. Aitken, D. B. Neale, K. D. Jermstad, N. C. Wheeler, and T. H. H. Chen. 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.
- Howe, G. T., G. Gardner, W. P. Hackett, and G. R. Furnier. 1996. Phytochrome control of short-day-induced bud set in black cottonwood. *Physiologia Plantarum* **97**:95-103.
- Huber, W., A. von Heydebreck, H. Sultmann, A. Poustka, and M. Vingron. 2002. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18**:S96-104.
- Huq, E., J. M. Tepperman, and P. H. Quail. 2000. GIGANTEA is a nuclear protein involved in phytochrome signaling in Arabidopsis (vol 97, pg 9789, 2000). *Proceedings of the National Academy of Sciences of the United States of America* **97**:11673-11673.
- Jaglo-Ottosen, K. R., S. J. Gilmour, D. G. Zarka, O. Schabenberger, and M. F. Thomashow. 1998. Arabidopsis CBF1 overexpression induces COR genes and enhances freezing tolerance. *Science* **280**:104-106.
- Joosen, R. V. L., M. Lammers, P. A. Balk, P. Bronnum, M. Konings, M. Perks, E. Statten, M. F. Van Wordragen, and A. H. M. van der Geest. 2006. Correlating gene expression to physiological parameters and environmental conditions during cold acclimation of *Pinus sylvestris*, identification of molecular markers using cDNA microarrays. *Tree Physiology* **26**:1297-1313.
- Kendall, E. J., and B. D. McKersie. 1989. Free-Radical and Freezing-Injury to Cell-Membranes of Winter-Wheat. *Physiologia Plantarum* **76**:86-94.
- Kenrick, P., and P. R. Crane. 1997. The origin and early evolution of plants on land. *Nature* **389**:33-39.
- Kerr, M. K., M. Martin, and G. A. Churchill. 2000. Analysis of variance for gene expression microarray data. *Journal of Computational Biology* **7**:819-837.
- Khaitovich, P., S. Paabo, and G. Weiss. 2005. Toward a neutral evolutionary model of gene expression. *Genetics* **170**:929-939.
- Khaitovich, P., G. Weiss, M. Lachmann, I. Hellmann, W. Enard, B. Muetzel, U. Wirkner, W. Ansorge, and S. Paabo. 2004. A neutral model of transcriptome evolution. *PLoS Biol* **2**:E132.
- Kinlaw, C. S., and D. B. Neale. 1997. Complex gene families in pine genomes. *Trends in Plant Science* **2**:356-359.
- Koag, M. C., R. D. Fenton, S. Wilkens, and T. J. Close. 2003. The binding of maize DHN1 to lipid vesicles. Gain of structure and lipid specificity. *Plant Physiology* **131**:309-316.
- Kolosova, N., B. Miller, S. Ralph, B. Ellis, C. Douglas, K. Ritland, and J. Bohlmann. 2004. Isolation of high-quality RNA from gymnosperm and angiosperm trees. *Biotechniques* **36**:821-824.

- Lai, Z., B. L. Gross, Y. Zou, J. Andrews, and L. H. Rieseberg. 2006. Microarray analysis reveals differential gene expression in hybrid sunflower species. *Molecular Ecology* **15**:1213-1227.
- Lemos, B., C. D. Meiklejohn, M. Caceres, and D. L. Hartl. 2005. Rates of divergence in gene expression profiles of primates, mice, and flies: Stabilizing selection and variability among functional categories. *Evolution* **59**:126-137.
- Levitt, J. 1980. Responses of plants to environmental stress. Academic Press, Inc., New York.
- Lynch, D. V., and P. L. Steponkus. 1987. Plasma membrane lipid alterations associated with cold acclimation of winter rye seedlings (*Secale cereale*). *Plant Physiology* **83**:761-767.
- Maere, S., K. Heymans, and M. Kuiper. 2005. BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. *Bioinformatics* **21**:3448-3449.
- McKersie, B. D. 1991. The role of oxygen free radicals in mediating freezing and desiccation stress in plants. Pp. 107-118 in E. Pell, and K. Steffen, eds. *Active Oxygen/Oxidative Stress and Plant Metabolism*. Am. Soc. of Plant Physiology.
- Mimura, M., and S. N. Aitken. 2007. Adaptive gradients and isolation-by-distance with postglacial migration in *Picea sitchensis*. *Heredity*.
- Mizoguchi, T., L. Wright, S. Fujiwara et al. 2005. Distinct roles of GIGANTEA in promoting flowering and regulating circadian rhythms in *Arabidopsis*. *Plant Cell* **17**:2255-2270.
- Morgenstern, E. K. 1996. Geographic variation in forest trees: genetic basis and application of knowledge in silviculture. University of British Columbia Press, Vancouver.
- Olsen, J. E., O. Junttila, J. Nilsen, M. E. Eriksson, I. Martinussen, O. Olsson, G. Sandberg, and T. Moritz. 1997. Ectopic expression of oat phytochrome A in hybrid aspen changes critical daylength for growth and prevents cold acclimatization. *Plant Journal* **12**:1339-1350.
- Oquist, G., P. Gardestrom, and P. A. Huner. 2001. Metabolic Changes During Cold Acclimation and Subsequent Freezing and Thawing. Pp. 137-163 in F. J. Bigras, and S. J. Colombo, eds. *Conifer Cold Hardiness*. Kluwer Academic Press, Dordrecht.
- Plieth, C., U. P. Hansen, H. Knight, and M. R. Knight. 1999. Temperature sensing by plants: the primary characteristics of signal perception and calcium response. *Plant Journal* **18**:491-497.
- Pomeroy, M. K., Siminovi. D, and F. Wightman. 1970. Seasonal biochemical changes in living bark and needles of red pine (*Pinus resinosa*) in relation to adaptation to freezing. *Canadian Journal of Botany* **48**:953-960.
- Ralph, S. G., H. Yueh, M. Friedmann et al. 2006. Conifer defence against insects: microarray gene expression profiling of Sitka spruce (*Picea sitchensis*) induced by mechanical wounding or feeding by spruce budworms (*Choristoneura occidentalis*) or white pine weevils (*Pissodes strobi*) reveals large-scale changes of the host transcriptome. *Plant Cell and Environment* **29**:1545-1570.
- Rifkin, S. A., J. Kim, and K. P. White. 2003. Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nat Genet* **33**:138-144.
- Rohde, A., and R. P. Bhalerao. 2007. Plant dormancy in the perennial context. *Trends in Plant Science* **12**:217-223.
- SAS, I. 1989. SAS/STAT(R) User's Guide, Version 6, Fourth Edition. SAS-Institute-Inc, Cary, NC.
- Schrader, J., R. Moyle, R. Bhalerao, M. Hertzberg, J. Lundeberg, P. Nilsson, and R. P. Bhalerao. 2004. Cambial meristem dormancy in trees involves extensive remodeling of the transcriptome. *Plant J* **40**:173-187.
- Schwechheimer, C., G. Serino, and X. W. Deng. 2002. Multiple ubiquitin ligase-mediated processes require COP9 signalosome and AXR1 function. *Plant Cell* **14**:2553-2563.
- Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**:2498-2504.
- Shen, Y. Y., X. F. Wang, F. Q. Wu et al. 2006. The Mg-chelatase H subunit is an abscisic acid receptor. *Nature* **443**:823-826.

- Shi, J. R., K. N. Kim, O. Ritz, V. Albrecht, R. Gupta, K. Harter, S. Luan, and J. Kudla. 1999. Novel protein kinases associated with calcineurin B-like calcium sensors in *Arabidopsis*. *Plant Cell* **11**:2393-2405.
- Stockinger, E. J., S. J. Gilmour, and M. F. Thomashow. 1997. *Arabidopsis thaliana* CBF1 encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. *Proceedings of the National Academy of Sciences of the United States of America* **94**:1035-1040.
- Storey, J. D., and R. Tibshirani. 2003. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America* **100**:9440-9445.
- Sunkar, R., and J. K. Zhu. 2004. Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. *Plant Cell* **16**:2001-2019.
- Tuskan, G. A.S. DiFazio S. Jansson et al. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**:1596-1604.
- Uemura, M., R. A. Joseph, and P. L. Steponkus. 1995. Cold acclimation of *Arabidopsis thaliana* - Effect on plasma membrane lipid composition and freeze induced lesions. *Plant Physiology* **109**:15-30.
- Uemura, M., and P. L. Steponkus. 1994. A contrast of the plasma membrane lipid composition of oat and rye leaves in relation to freezing tolerance. *Plant Physiology* **104**:479-496.
- Weiser, C. J. 1970. Cold resistance and injury in woody plants. *Science* **169**:1269-1278.
- Whitehead, A., and D. L. Crawford. 2006a. Neutral and adaptive variation in gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **103**:5425-5430.
- Whitehead, A., and D. L. Crawford. 2006b. Variation within and among species in gene expression: raw material for evolution. *Molecular Ecology* **15**:1197-1211.
- Wilkosz, R., and M. Schlappi. 2000. A gene expression screen identifies EARL1 as a novel vernalization-responsive gene in *Arabidopsis thaliana*. *Plant Molecular Biology* **44**:777-787.
- Xiong, L., K. S. Schumaker, and J.-K. Zhu. 2002. Cell signaling during cold, drought, and salt stress. *Plant Cell* **14**:S165-183.
- Yan, J., J. Wang, Q. Li, J. R. Hwang, C. Patterson, and H. Zhang. 2003. AtCHIP, a U-box-containing E3 ubiquitin ligase, plays a critical role in temperature stress tolerance in *Arabidopsis*. *Plant Physiol* **132**:861-869.
- Yanai, I., D. Graur, and R. Ophir. 2004. Incongruent expression profiles between human and mouse orthologous genes suggest widespread neutral evolution of transcription control. *Omics-a Journal of Integrative Biology* **8**:15-24.
- Yu, X. M., M. Griffith, and S. B. Wiseman. 2001. Ethylene induces antifreeze activity in winter rye leaves. *Plant Physiol* **126**:1232-1240.
- Ziegler, P., and O. Kandler. 1980. Tonoplast stability as a critical factor in frost injury and hardening of spruce (*Picea abies* L. Karst). *Z. Pflanzenphysiol. Bd.* **99**:393-410.
- Zwiazek, J. J., S. Renault, C. Croser, J. Hansen, and E. Beck. 2001. Biochemical and biophysical changes in relation to cold hardiness. Pp. 165-186 in F. J. Bigras, and S. J. Colombo, eds. *Conifer Cold Hardiness*. Kluwer Academic Publishers, Dordrecht.

3. Evidence for Positive and Diversifying Selection Among 174 Cold Hardiness-related Candidate Genes in Sitka spruce (*Picea sitchensis*)²

3.1 Introduction

Natural selection to conserve or alter protein function leaves its signature in nucleotide sequence data. In the absence of selection, nucleotide diversity within-population is governed by the effective population size (N_e) and the per generation per base pair mutation rate (μ); linkage disequilibrium is affected by N_e , μ , and the per generation recombination rate between adjacent sites (r); and the level of synonymous and non-synonymous nucleotide diversity is expected to be roughly equal. Positive selection may increase the proportion of derived SNPs segregating at high frequency, increase local linkage disequilibrium (LD), or increase the proportion of non-synonymous polymorphisms (Nielsen 2005). Whereas gene flow tends to homogenize allele frequencies among populations, spatially heterogeneous selection may lead to variation in allele frequencies across a species range for genes related to local adaptation. Genes that appear to be targets of natural selection are of interest for association studies in non-model species, where a suite of candidate genes must be identified in advance of SNP genotyping in large mapping populations. More generally, understanding patterns of nucleotide diversity, linkage disequilibrium, and population structure are prerequisite to the successful execution of association studies.

By synchronizing their annual cycle of growth and dormancy to environmental cues, primarily daylength and temperature, temperate and boreal conifers maximize growth in a particular climatic niche while minimizing damage caused by freezing temperatures (Weiser 1970). Since the last glacial maxima approximately 18,000 years ago, conifer species have expanded to fill ranges that in some cases span >20 degrees of latitude, and clinal variation in climate-related traits has been observed in many of these species (Morgenstern 1996; Savolainen et al. 2007). In the context of climate change, maintenance of genetic variation directly related to local climatic adaptation is of increasing concern, particularly as it has become apparent that neutral genetic markers are not good proxies for conservation of the molecular genetic determinants of quantitative traits (Reed and Frankham 2001).

² A version of this chapter is being revised for re-submission for publication. Holliday, J.A., Ritland, K., and Aitken, S.N. Evidence for Positive and Diversifying Selection Among 174 Cold Hardiness-related Candidate Genes in Sitka spruce (*Picea sitchensis*) (Genetics).

Sitka spruce (*Picea sitchensis* (Bong.) Carr.) is an ideal species for genomic dissection of local adaptation to climate due to strong clinal variation in associated traits, such as timing of budset and cold hardiness. When grown in a common garden, seedlings originating from the northern edge of the species range (southern Alaska, USA) set bud in early July, whereas southern provenances (originating from northern California, USA) often do not do so until November (Mimura and Aitken 2007a). Similar patterns of variation in autumn cold hardiness have also been observed. Accordingly, the proportion of variation in these quantitative traits that is distributed among populations, Q_{ST} , has been estimated at ~ 0.9 , whereas F_{ST} , which is the analog to Q_{ST} for genetic markers, is comparatively low (~ 0.1 for microsatellite markers) (Mimura and Aitken 2007a).

Can specific genes be identified that contribute to this adaptive differentiation? The task is not as simple as it seems, as recent studies in Norway spruce (*Picea abies*) suggest that some of this quantitative trait variation in cold hardiness and bud phenology may be explained by epigenetics (Johnsen et al. 2005a; Johnsen et al. 2005b). Nevertheless, investigating a comprehensive set of candidate genes for local adaptation is now much easier with the advent of genomics resources for conifers, and such studies provide insights into the possible role of these genes in adaptive phenotypes. To this end, surveys of nucleotide diversity have been carried out in several conifer species, including Norway spruce (*Picea abies*) (Heuertz et al. 2006), loblolly pine (*Pinus taeda*) (Brown et al. 2004), Douglas-fir (*Pseudotsuga menziesii*) (Krutovsky and Neale 2005), and Scots pine (*Pinus sylvestris*) (Pyhajarvi et al. 2007). Between 15 and 22 candidate genes were analyzed in these studies, which revealed lower than expected nucleotide diversity and generally low levels of LD. Hallmarks of background or positive selection were found in several genes. In addition, two recent studies employed outlier detection approaches to identify signatures of divergent selection. Eveno et al. (2008) assessed genetic differentiation across the range of maritime pine (*Pinus pinaster*) and found evidence for divergent selection in five out of 11 tested genes. Namroud et al. (2008) sampled six adjacent populations of white spruce (*Picea glauca*) that were largely panmictic ($F_{ST}=0.006$), and employed high-throughput genotyping to successfully score 534 SNPs across 345 expressed genes (Namroud et al. 2008). The SNPs genotyped in the latter study were also mapped, and outlier loci were fairly evenly distributed across the genome.

One of the largest expressed sequence tag (EST) resources for a non-model plant is available for *Picea* spp., and a spruce cDNA microarray comprised of 21,840 unique elements has also been developed (Ralph et al. 2006; Ralph et al. 2008). In conifers there have been few functional studies of genes, and candidate gene selection therefore relies on homology searches of angiosperm databases and gene expression studies. To enhance selection of candidate genes with possible functional roles in cold hardiness-related traits, we previously employed the spruce 21.8k cDNA microarray to select

174 local adaptation-related candidate genes (Holliday et al. 2008). In the current study we have resequenced these candidate genes in a panel of 24 Sitka spruce genotypes to characterize nucleotide diversity and linkage disequilibrium, and to identify targets of positive and diversifying selection.

3.2 Materials and methods

3.2.1 *Plant material*

Foliage for RNA extraction was obtained from 6-yr-old Sitka spruce, which were grown from seed collected from natural populations spanning the species range in a raised-bed outdoor common garden in Vancouver, BC, Canada (49° N, 123°W) (Figure 3.1) (Mimura & Aitken, 2007). Needle tissue samples (~1g) were flash frozen in an N₂ vapor tank upon collection, and subsequently stored at -80°C until processing. Four seedlings were sampled from each of six geographic populations: Redwood, California (41°N), Columbia River, Oregon (47°N), Vancouver, British Columbia (49°N), Prince Rupert, British Columbia (53°N), Valdez, Alaska (62°N), and Kodiak Island, Alaska (57°N). It should be noted that although the Kodiak Island population is farther south than Valdez, it is the farthest population from the southern limit of the species range from the perspective of gene flow, and represents the leading edge of migration of the species range (Mimura and Aitken 2007). These populations were chosen to be roughly evenly distributed geographically, to be from low elevation, and to encompass most of the latitudinal range of the species, which spans approximately 3400km (although this does not correspond to a true north-south cline as the Pacific coast runs southeast to northwest). The sampled populations also represent the range of diverse climates inhabited by Sitka spruce (and hence the variation in cold hardiness-related traits), for which mean annual temperature varies from a high of 11.6°C for the Redwood population, to a low of 3.5°C for the Valdez population.

3.2.2 *Candidate gene selection and primer design*

Candidate genes were selected on the basis of their previously characterized autumn expression profiles (Holliday et al. 2008), or in a few cases only the function of their nearest *Arabidopsis* homolog. ESTs of interest from the microarray were queried against our in-house contig assembly and the largest contig was selected for use in primer design. In some cases ESTs were present in the database as singletons. To design primers, we first used BLASTN to identify possible paralogous genes among all spruce ESTs. These sequences were aligned to the target and primers were designed using Primer3 (Rozen and Skaletsky 2000) software to maximize product length, minimize potential cross-amplification of paralogs, and give an annealing temperature of ~60°C.

3.2.3 Sequencing

To identify SNPs present in protein coding sequences, we used cDNA as a template for PCR. Total RNA was extracted by a previously published protocol (Kolosova et al. 2004), and cDNA synthesized using Superscript II reverse transcriptase (Invitrogen, Carlsbad, CA, USA) with an oligo dT₁₂₋₁₈ primer. Candidate genes were amplified by PCR using Taq DNA polymerase (Roche, Indianapolis, IN, USA) and those giving single products (as evidenced by a single band on a 1% agarose gel) were purified using AMPure magnetic beads (Agencourt Bioscience, Beverly, MA, USA) and sequenced directly using BigDye v.3.1 ready reaction mix on an ABI3730xl in both the forward and reverse direction in the 24 Sitka spruce genotypes (Applied Biosystems, Foster City, CA, USA). Bases were called and trace files assembled using CodonCode Aligner software (CodonCode Corporation, Dedham, MA, USA). Sequence reads were retained if they had at least 200 bp with an average PHRED score greater than 20. Low quality ends were trimmed manually. Although we generally obtained quality sequence for most if not all individuals sequenced, a minimum cutoff of 15 individuals with quality sequence (30 haploid genomes) per gene was set for further analysis. Putative SNPs were identified using the mutation detection algorithm embedded in CodonCode Aligner, and traces at putative homozygous and heterozygous SNPs were reviewed manually. Primer pairs exhibiting evidence of amplification of paralogous genes were discarded. This was ascertained based on three criteria. First, possible low-similarity paralogs that were indistinguishable from poor sequence were eliminated at the trace alignment phase. Second, finished alignments with greater than 5% divergence among individuals were assumed to include paralogs and discarded. As a final filter, the program HaPLoSNPer (<http://www.bioinformatics.nl/tools/haPLoSnper/>) was used to calculate the standard deviation of the number of SNPs among haplotypes (the D-value) (Tang et al. 2008). The intuition behind this approach is that for sequenced amplicons from a given pair of primers, haplotypes with a large number of SNP variants relative to other haplotypes likely represent paralogous genes. Specifically, as the D-value exceeds 0.6 the likelihood of an alignment comprising paralogs increases dramatically (Tang et al. 2006). As such, in the few cases where this was observed, the candidate genes were discarded.

3.2.4 Sequence analysis

Sequence alignments for the 24 diploid individuals in the study (with ambiguous bases at heterozygous sites), were split into 48 pseudo-haplotypes. Haplotype were reconstructed using the ELB algorithm in Arlequin software (version 3.11) (Excoffier et al. 2003; Excoffier et al. 2005). To determine open reading frames (ORFs), consensus sequences for each candidate gene were used as BLASTX queries to search the *Arabidopsis* genome. Where the query sequence was translated in the

same reading frame for alignment to the top three hits, this was assumed to be the correct ORF. Although we biased our selection of candidate genes toward those with homology to genes in other species, we included seven genes with no homology that were upregulated strongly during autumn cold acclimation in the previous study. In these cases, sequence reads were translated in all six reading frames, and the longest of these ORFs was used. It should be noted that although our sequencing of cDNA excludes introns, there can still be substantial non-coding regions (5' and 3' untranslated regions (UTRs)) in these sequences. Candidate genes were annotated according to their top BLASTX match to the *Arabidopsis* genome, and given unique identifiers according to this hit. To reflect uncertain orthology in conducting such searches, we generally used the gene family of the best hit for the annotation, rather than the specific gene. For example, where three candidate genes were respectively annotated as members of three different glycosyl hydrolase subfamilies, they were simply given names that reflected their similarity to glycosyl hydrolases (e.g., *gh1*, *gh2*, *gh3*). Although there may be more appropriate species for annotation purposes with respect to evolutionary proximity (e.g., *Pinus taeda*) or availability of a genome sequence (e.g., *Populus trichocarpa*), annotations for these species are more often than not ‘daisy-chained’ back to *Arabidopsis*, and therefore usually do not provide further insight as to the possible function of gene products associated with our candidate genes.

3.2.5 Nucleotide diversity, linkage disequilibrium and selective neutrality tests

Standard diversity parameters (number of segregating sites, number of haplotypes, nucleotide diversity (π and θ), as well as non-synonymous and synonymous nucleotide diversity) were calculated using DnaSP software (version 4.5.3) (Rozas et al. 2003). LD was characterized for each candidate gene by calculating the descriptive statistic r^2 using DnaSP. To determine the average decay of LD with physical distance, we used the NLS function in R (<http://www.r-project.org/>) to fit the r^2 expectation of Hill and Weir (1988)

$$E(r^2) = \left[\frac{10 + \rho X}{(2 + \rho X)(11 + \rho X)} \right] \left[1 + \frac{(3 + \rho X)(12 + 12\rho X + (\rho X)^2)}{n(2 + \rho X)(11 + \rho X)} \right],$$

where X is the distance between polymorphic sites in base-pairs, n is the sample size, and ρ is the least squares estimate of the population recombination parameter (where $\rho = 4N_e r$).

To test the null hypothesis of neutral evolution for each of our candidate genes, the selective neutrality tests d_N/d_S , Tajima’s D , and Fay and Wu’s H were calculated using DnaSP (Tajima 1989; Fay and Wu 2000). We used the Z-test implemented in MEGA software (version 4.024) to test the null hypothesis that $d_N = d_S$, by comparing the average number of nonsynonymous substitutions per nonsynonymous site (d_N) to the average number of synonymous substitutions per synonymous site

(d_S) [$Z = (d_N - d_S) / \sqrt{\text{Var}(d_N) - \text{Var}(d_S)}$] (Tamura et al. 2007). Values of the test statistic that are significantly less than zero indicate purifying selection, i.e., a deficiency of non-synonymous substitution, whereas values significantly greater than zero indicate positive directional selection, i.e., an excess of non-synonymous substitutions. Standard error was computed by the bootstrap with 500 replicates. Tajima's D quantifies skew in the frequency spectrum of mutations by comparing the mean number of substitutions between pairs of sequences (π) to the total number of segregating sites (S). Under neutrality, these should be equal. Excesses of low frequency mutations suggest background selection or selective sweeps, whereas excesses of medium frequency mutations suggest balancing selection. Fay and Wu's H uses an outgroup sequence to identify high-frequency derived mutations characteristic of genetic hitchhiking associated with a selective sweep. For the latter, outgroup sequences were identified using BLASTN from our in-house white spruce (*Picea glauca*) contig build. Statistical significance of Tajima's D and Fay and Wu's H was assessed using standard coalescent simulations without recombination (1000 realizations), and two-tailed probability values were adjusted for the false-discovery rate using the 'qvalue' package in R (Storey and Tibshirani 2003). As a compromise between false-positives and false-negatives, we used a Q-value cutoff of 0.1.

To test for decoupling of nucleotide diversity from divergence, we applied the HKA test to those candidate genes for which outgroup sequences were available (Hudson et al. 1987). First, the multilocus HKA test distributed by Hey (<http://lifesci.rutgers.edu/~heylab/index.html>) was used to determine whether the dataset as a whole fits the neutral expectation for diversity vs. divergence. This program performs coalescent simulations (2000 replicates in our case) based on parameters observed in the data and compares the fit of the observed values for diversity and divergence with the distribution of values simulated under a neutral model. A maximum likelihood HKA test (MLHKA, available at http://labs.eeb.utoronto.ca/wright/Stephen_I._Wright/Programs.html) (Wright and Charlesworth 2004) was then used to test specific loci that had the greatest marginal contribution to observed deviations. MLHKA uses a Markov-chain Monte Carlo (MCMC) approach to evaluate the fit of the data to two models, one assuming neutrality of all loci (where the selection parameter, k , in the model, is fixed at $k=1$ for all loci) and one allowing for selection on a subset of loci (for which k is allowed to vary). A likelihood ratio test of the two models is then used to test for deviation from neutral expectations (where twice the difference in log likelihood of the models is approximately chi-squared distributed with degrees of freedom equal to either the number of genes tested or to the difference in number of selected loci for nested tests between two selection models). We applied this test to a subset of 16 of our candidate genes: 10 that deviated little from neutral expectation and were therefore included as a 'neutral' test set, and a further six genes that exhibited high divergence relative to diversity. Several run lengths were tested and we found that runs of one million steps

converged on very similar likelihoods, so this run length was used for each model and the final log likelihoods were averaged across runs for the purposes of the likelihood ratio test. Initial runs were conducted assuming a neutral model, followed by runs allowing for selection on the six candidates exhibiting deviations from the neutral expectation on the basis of the initial HKA test. Subsequent runs were also conducted that excluded one gene with an estimate for k close to one, suggesting it was not subject to selection.

High F_{ST} for a particular candidate gene or SNP can be indicative of diversifying selection, and we therefore sought to characterize the patterns of population subdivision in Sitka spruce for our candidate genes. F_{ST} was first calculated using all SNPs for each candidate gene (Hudson et al. 1992). To identify specific SNPs exhibiting outlier behavior, we used the approach of Beaumont and Balding (2004), implemented in the program 'BayesFst' (available at <http://www.reading.ac.uk/Statistics/genetics/software.html>) which employs MCMC to regress $\log(F_{STij}/1 - F_{STij})$ onto a locus effect (α_i), a population effect (β_j) and a locus by population interaction term (γ_{ij}). This model allows us to separate locus-specific effects (i.e., mutation, selection) from population-specific effects (i.e., N_e , migration). The interaction term encompasses selection that is restricted to a single or a small number of populations. As our sample sizes are relatively small, we have little power to detect the latter effects, and initial runs revealed no significant γ_{ij} effects. We therefore tested only the simplified model as inclusion of γ_{ij} greatly increases computation time. Gaussian priors with means of 0 and -2.0 were placed on α_i and β_j , respectively, with a standard deviation of 1.0 for each. We performed multiple runs to ensure consistency, with 50,000 iterations of the Markov chain discarded as burn-in, followed by 500,000 iterations post-burn-in, of which 2000 uncorrelated outputs were sampled to estimate the posterior distribution for each parameter. Locus and population effects were considered significant if the one-tailed 100(1-P)% posterior interval excluded the prior mean. It should be noted that in this case F_{ST} refers to the probability that two randomly chosen chromosomes in a population have a common ancestor within that population (Beaumont and Balding 2004). In contrast, the method of Hudson (1992) used for characterizing F_{ST} for the whole sequence of a given gene is $1 - (H_w/H_b)$, where H_w is the average number of differences between individuals in the same population and H_b is the average number of differences between individuals from different populations.

As a complement to these model-based tests, we tested whether particular candidate genes exhibited outlier behavior with respect to our empirical distribution of nucleotide diversity and F_{ST} (Akey et al. 2002). We assumed our 174 candidate genes represented a genome-wide sample of loci, and calculated percentiles of the distributions for these two statistics. Genes exhibiting apparently reduced nucleotide diversity or enhanced population differentiation were compared to this

distribution, and considered outliers if the test statistic fell below the 2.5th percentile or outside the 97.5th percentile for π and F_{ST} , respectively (i.e., an α level of 0.05). This approach has the advantage that it is not subject to model assumptions (e.g., with regard to mutation, migration, etc.). However, because all distributions have values in their tails, it by definition assumes that some loci (i.e., the outliers) are subject to selection (Kelley and Swanson 2008).

Finally, as population structure resulting from unequal gene flow across a species range can have a significant impact on the conclusions drawn from future genetic association studies in Sitka spruce, we also characterized genome-wide patterns of population structure. For this purpose, pairwise F_{ST} values were calculated using all SNPs in all genes for all possible contrasts between populations. These estimates were also used to assess isolation by distance along the latitudinal range of Sitka spruce, by correlating genetic and geographic distance matrices using a Mantel test with 1000 permutations.

3.3 Results

We obtained quality sequence for a total of 174 genes, with a combined length of 107.2 kb (Table 3.1). On average, 22 of 24 individuals per gene returned >200bp of quality sequence (44 out of 48 haploid genomes), and 86% of genes returned quality sequence for ≥ 20 individuals (40 haploid genomes). A total of 1829 parsimony informative SNPs were observed across all candidate genes. After trimming low-quality ends, the average length of the re-sequenced candidate gene alignments and number of segregating sites was 616bp and 11, respectively (Table 3.1). This gives an average of approximately one SNP every 55bp. The average nucleotide diversity (π) across all sites was 3.5×10^{-3} , and within coding regions the synonymous nucleotide diversity exceeded the non-synonymous diversity by 3.4 fold (6.8×10^{-3} vs. 2.1×10^{-3}) (Table 3.1).

3.3.1 Patterns of nucleotide diversity

Substantial variation was observed in nucleotide diversity among candidate genes, with some having almost no variation (e.g., $\pi = 0.0001$ for *pat1*, a gene similar to *PHYTOCHROME A SIGNAL TRANSDUCTION 1*) and others with nearly an order of magnitude more variation than the mean (e.g., $\pi = 0.022$ for a putative basic helix-loop-helix transcription factor, *bhlh2*). In addition to *pat1*, several genes putatively involved in light signal transduction, chosen for their strong potential for involvement in local adaptation based on functional studies of putative orthologs in model systems, had very little diversity. These included two cryptochrome-like genes (*cry1*, $\pi = 0.0003$; *cry2*, $\pi = 0.0004$), a gene similar to *GIGANTEA* (*gi*, $\pi = 0.0005$), a *FLOWERING LOCUS T*-like gene (*ft*, $\pi = 0.0007$), and a *CONSTANS*-like gene (*co*, $\pi = 0.0008$). A gene similar to a recently described abscisic

acid (ABA) receptor also exhibited very limited diversity (*chlhl1*, $\pi = 0.0007$) (Shen et al. 2006), as did a putative ortholog to the flowering time control gene, *FCA* (*fca*, $\pi = 0.0002$). Two of these candidate genes, *pat1* and *fca*, had nucleotide diversity estimates that fell below the 2.5th percentile of the empirical distribution of nucleotide diversity. Indeed, *pat1* had the lowest diversity among all of our candidate genes.

3.3.2 Evidence for purifying and directional selection

The mean d_N/d_S ratio comparing the level of nonsynonymous with synonymous substitutions was 0.40 (Table 3.1). A total of nine genes had d_N/d_S significantly less than one using the Z-test (Table 3.2). By contrast, evidence for positive selection to alter protein function using this test was observed for only one candidate, a putative auxin-responsive gene (*air*; $Z=3.53$, $Q<0.05$) (Table 3.2). Although not statistically significant, four others had $d_N/d_S>2$, including a putative glutamate receptor (*glr*), putative dormancy/auxin-related gene (*drm2*), a gene similar to the circadian clock-associated myb transcription factor *LATE ELONGATED HYPOCOTYL* (*lhy*), and a putative glutathione S-transferase (*gst4*) (data not shown).

Average values for Tajima's D and Fay and Wu's H were negative (-0.56 and -0.85, respectively) (Table 3.1). The majority of candidate genes had negative Tajima's D values (Figure 3.2), 13 of which were statistically significant, indicating an excess of low frequency mutations suggestive of background selection against deleterious mutations or selective sweeps. In contrast, only two genes (*bhlh2* and *int2*) had significantly positive Tajima's D values indicating an excess of medium frequency mutations characteristic of balancing selection. Where white spruce outgroup sequences were available (148 out of 174 candidate genes), we also calculated Fay and Wu's H, which is most sensitive to high-frequency derived mutations indicative of recent selective sweeps. Although the mean value for Fay and Wu's H was more negative than for Tajima's D, this was driven by large negative values for a few genes. Most values for Fay and Wu's H were clustered around zero, with a median of -0.05, whereas the median Tajima's D was -0.73 (Figure 3.2). Six genes had significantly negative values for Fay and Wu's H after correction for multiple testing, including a histone H1-like gene (*hta2*; $H=-21.78$, $Q<0.1$), a gene putatively involved in RNA splicing (*swap*; $H=-8.50$, $Q<0.05$), a putative $\Delta 8$ -sphingolipid desaturase (*sld*; $H=-6.40$, $Q<0.1$), a late embryogenesis abundant (LEA) gene (*lea1*; $H=-5.57$, $Q<0.1$), a basic helix-loop-helix family transcription factor (*bhlhl1*; $H=-5.22$, $Q<0.1$), and an isoflavone reductase-like gene (*ifr4*; $H=-3.27$, $Q<0.1$) (Table 3.2).

A statistically significant and positive relationship was observed between nucleotide diversity (θ) and divergence (D_{xy}) for the candidate with available outgroup sequences ($R^2=0.27$, $P<0.001$). To test whether these estimates fit the neutral expectation, we used a multilocus HKA test and found a

significant departure from neutrality ($\chi^2=108.01$, $P<0.05$). Six genes had a deviation from the expected value for diversity or divergence greater than ~ 3 , including a leucine rich repeat-like gene (*lrr6*), a phosphatidylinositol-4-phosphate 5-kinase family gene (*pip5k*), a raffinose synthase like gene (*rs1*), a putative ATP-binding cassette transporter (*abc5*), a gene with no significant hit (*nsh1*), and one of the candidate ABA-receptors (*chlh1*). The program MLHKA revealed that the selection model fit the data significantly better than the neutral model (Table 3.3), providing evidence for positive selection on these genes. Although all of these genes had selection parameter estimates less than one, one gene, *nsh1*, had an estimate for k approaching one, suggesting it is less likely to have been a target of selection than the other five genes. We therefore tested a reduced model allowing for selection on only five genes, which was also fit the data significantly better than the neutral model. A comparison between the two selection models suggests that the six-gene model is not a significant improvement over the five-gene model, and we therefore conclude that only the subset of five genes have been targets of positive selection (Table 3.3).

3.3.3 Detection of empirical and model-based F_{ST} outlier loci

Mean F_{ST} across all genes was 0.12, with a standard deviation of 0.13. Estimates were highly variable among genes, ranging from close to zero to as high as 0.62. To identify spatially heterogeneous selection across the climatically diverse range of Sitka spruce, we first assessed population subdivision by averaging F_{ST} for each gene across all available segregating sites. Among the genes exhibiting high levels of population differentiation were two endochitinase-like genes (*chib1* and *chib2*; $F_{ST}=0.42$ and 0.55 , respectively), a thaumatin-like gene (*tlp1*; $F_{ST}=0.59$), two auxin-related genes (*air* and *drm2*; $F_{ST}=0.52$ and 0.62 , respectively), a putative no apical meristem family transcription factor (*nam2*; $F_{ST}=0.60$), and a peroxidase-like gene (*per3*; $F_{ST}=0.45$). Five of these genes were outside the 97.5th percentile ($F_{ST}>0.50$) of the empirical distribution of F_{ST} , suggesting possible diversifying selection.

Using the model-based approach of BayesFst, thirty-five SNPs within 14 candidate genes were identified whose 90% posterior intervals were positively shifted from the prior mean of zero, and are therefore candidates for diversifying selection (Figure 3.4a). Among these significant outliers were a total of four SNPs within the two endochitinase-like genes noted above (*chib1* and *chib2*), 13 within the peroxidase-like gene (*per3*), and one within the NAM-like transcription factor (*nam2*) (Table 3.4). The list of outliers also contained a notable excess of oxidative stress-related genes and carbohydrate metabolism. In addition to *per3*, two glutathione peroxidase-like genes (*gp1* and *gp2*) contained outlier SNPs, as did another peroxidase-like gene (*per2*) and an isoflavone reductase-like gene (*ifr2*). Two glycosyl hydrolase-like genes also exhibited outlier SNPs (*gh2* and *gh4*), as did a

putative xyloglucan:xyloglucosyl transferase (*xth1*). In contrast, only two genes (*ugt2* and *abc5*) with markedly low F_{ST} reached the 10% significance threshold (that is, negatively shifted posterior distributions) suggestive of spatially homogenizing selection. This is not surprising as the method of Beaumont and Balding (2004) is not particularly sensitive to such outliers when the genome-wide mean F_{ST} is ~ 0.1 or less.

3.3.4 Genome-wide patterns of population structure

The Bayesian regression method implemented in BayesFst accounts for variable migration rates by explicitly modeling those population effects in the β_j term. Positive shifts from the prior mean are suggestive of a particular population receiving less gene flow relative to other populations, or of having a smaller N_e , whereas negative shifts imply the opposite. The posterior distributions for the six populations included in this study are given in figure 4b. Three of the six populations have posterior distributions positively shifted from the prior mean of -2.0: Kodiak Island, AK ($P < 0.001$), Redwood, CA ($P < 0.01$), and Valdez, AK ($P < 0.07$). Interestingly, the strongest signals were from the Kodiak Island and Redwood populations, which are at the leading and trailing edge of the species range, respectively. Although Valdez is within the continuous portion of the range, it is also near the northern limit. In contrast to these populations apparently receiving less gene flow, no population had a significant negatively shifted posterior distribution of β_j indicative of greater gene flow or effective population size.

In agreement with the results from BayesFst, the highest pairwise F_{ST} values when all SNPs were included were for those involving the disjunct Kodiak Island population. Conversely, the lowest pairwise contrasts were those involving the Columbia River population (except when compared with Kodiak Island) (Table 3.4). The largest pairwise F_{ST} values were for the Columbia River – Kodiak Island and Redwood – Kodiak Island contrasts ($F_{ST} = 0.146$ and 0.145 , respectively), which are the most geographically distant populations. The most genetically homogeneous pair of populations was Columbia River and Vancouver ($F_{ST} = 0.035$), which are geographically proximal. Significant isolation by distance was observed when all SNPs were pooled to estimate F_{ST} ($r = 0.71$, $P < 0.01$) (Figure 3.5).

3.3.5 Linkage disequilibrium

LD decayed very rapidly, on average. At an r^2 of 0.2, the distance over which LD extended was, on average, less than 100bp (Figure 3.6). Even over very short distances, many pairwise r^2 estimates were close to zero. The nonlinear regression of r^2 on distance depicted in Figure 3.6 gives a least-squares estimate of the recombination parameter, ρ , of 0.049. In spite of this average rapid decay of LD, there were many SNPs that remained highly correlated at distances greater than 500bp.

3.4 Discussion

This study describes the patterns of nucleotide polymorphism, natural selection, and linkage disequilibrium in a widely distributed, economically important conifer, Sitka spruce. This species occupies highly diverse climatic niches across its ~3600km latitudinal range, with substantial variation in traits related to local adaptation to climate segregating both within and among-population. We used several types of analyses to identify which of our candidate genes may have been subject to natural selection. Although such a study is inevitably incomplete, it provides testable hypotheses about specific candidate genes that can be carried forward to larger population samples, further assays of phenotypic traits, or functional studies.

3.4.1 Lack of variation in genes putatively involved in light and ABA perception

The overall nucleotide diversity in the panel of 174 nuclear genes reported here (3.49×10^{-3}) is approximately average for a conifer (Savolainen and Pyhajarvi 2007), and wide variation in diversity was observed among genes. Several genes chosen due to their putative involvement in ABA and daylength perception and signaling, two processes fundamental to cold acclimation in perennials, exhibited among the lowest level of variation of all the candidate genes we sequenced. Among these was a candidate ABA receptor, a cryptochrome-like gene, and genes similar to *CO*, *FT* and *GI*, and *FCA*. One of these genes, *chlhl*, had low diversity relative to divergence suggestive of positive selection. In addition, *fca* and *patl* had nucleotide diversity estimates that fell below the 2.5th percentile of the empirical distribution. A lower mutation rate in these genes may explain this result, but it is also possible that the reduced diversity is due to either strong purifying selection or a recent, complete selective sweep. Purifying selection appears unlikely given the lack of both synonymous and non-synonymous substitutions in both the coding and non-coding (UTR) regions of both of these genes. Each of these genes have only two polymorphisms with allele frequencies < 5%, and one would expect an average level of silent site diversity if the primary selective constraint is against amino acid changing mutations. This level of diversity could be consistent with a very strong selective sweep, but we would expect an average level of divergence with the white spruce outgroup sequence, which is not the case (there are only two fixed differences between white spruce and Sitka spruce for each of these genes). As Sitka and white spruce freely hybridize along two river drainages on the central coast of British Columbia, it is possible that the pattern of diversity and divergence observed for *patl* and *fca* is the result of a selective sweep that crossed the species boundary. Re-sequencing of the entire coding region of these genes in a larger panel of individuals from both white and Sitka spruce will be necessary to determine if, and in what way, selection is acting.

3.4.2 Positive selection or demography? Contrasting patterns between neutrality tests based on the frequency spectrum

The majority of genes we re-sequenced had negative Tajima's *D* values. As Tajima's *D* is influenced by neutral demographic processes, this suggests a genome-wide departure from equilibrium conditions in Sitka spruce. Overall negative values for Tajima's *D* have been observed in most studies of nucleotide diversity in forest trees, and in recent years efforts have been made to model a variety of historical demographic events conditional on diversity parameters estimated from re-sequencing datasets (Heuertz et al. 2006; Pyhajarvi et al. 2007; Ingvarsson 2008). These studies found the model most consistent with the data is a bottleneck scenario dating to between ~250 and ~750 KYA. Although there is substantial uncertainty as to the timing, and such a scenario undoubtedly oversimplifies the demographic history of these species (Ingvarsson 2008), these studies provide reasonably strong evidence that caution must be exercised when using the standard neutral model as a null hypothesis for frequency spectrum-based neutrality tests. Given the observed negative mean value for Tajima's *D* in the current study, it is therefore possible that some of the significantly negative values for this statistic are false discoveries, and that some positive outliers have been missed. Whereas Tajima's *D* is most sensitive to an excess of low frequency variants that may be indicative of either recent selective sweeps or population expansion after a bottleneck, Fay and Wu's *H* tests for an excess of high frequency derived variants, a metric that is not as readily confounded by demography. Consistent with this, most of the values we observed for Fay and Wu's *H* were close to zero, as was the median, in contrast to Tajima's *D*, for which most genes had negative values, and both the mean and median are negative. We found six genes with significant values for *H* after correcting for multiple testing, including three putatively involved in transcriptional regulation (*hta2*, *bhlh1*, and *swap*), an isoflavone reductase-like gene (*ifr4*), a dehydrin-like gene (*lea1*), and a putative $\Delta 8$ -sphingolipid desaturase (*sld*). The latter two are most interesting from the perspective of climatic adaptation. Dehydrins are induced by both drought and cold (Close 1997; Richard et al. 2000; Welling et al. 2004; Holliday et al. 2008), and this dual role likely reflects the dehydration common to both stressors. Although evidence is somewhat circumstantial, it is thought that dehydrins act to stabilize cellular macromolecules (Close 1997). Several previous studies of nucleotide diversity in conifers have included one or more dehydrin-like genes, and one found patterns of polymorphisms consistent with divergent selection (Eveno et al. 2008). The putative $\Delta 8$ -sphingolipid desaturase was included in this study due to its induction under short days in Sitka spruce (Holliday et al. 2008), and because substitution of saturated for unsaturated phospholipids is an important component of cold acclimation (Uemura and Steponkus 1994; Uemura et al. 1995). Indeed, *cis* double bonds at the $\Delta 8$

position of sphingolipids have been associated with chilling and freezing resistance in plants (Lynch and Dunn 2004), and this gene is therefore very interesting as a candidate for functional studies.

3.4.3 *Departures from neutrality for diversity vs. divergence*

We observed a significant departure from neutral expectation among our 148 candidate genes using the multilocus HKA test. We further probed six of the genes that appeared to deviate most from neutrality, and found that five of these genes may have been subjects of positive selection. Two particularly interesting loci among these are *rs1* and *chlh1*. The former is a putative raffinose synthase, an enzyme that performs the final step in the biosynthesis of the well-characterized cold hardiness-related trisaccharide sugar, raffinose (Zwiazek et al. 2001), and the latter is a candidate ABA receptor. The significant reduction in diversity relative to divergence in *chlh1* suggests that ABA sensing may have been a target of positive selection in Sitka spruce.

3.4.4 *Purifying selection against replacement substitutions is pervasive in Sitka spruce*

Many more synonymous than non-synonymous substitutions were found in our candidate genes, which is reflected in a mean d_N/d_S ratio of 0.40. The same pattern was found across a substantial panel of SNPs in white spruce (Pavy et al. 2006), which suggests genome-wide selective pressure to maintain existing protein function in spruce species. There is some debate as to how within-species d_N/d_S ratios should be interpreted, however. Recently developed theory suggests that strong negative selection is expected to produce a lower d_N/d_S ratio between divergent lineages than within lineages (Kryazhimskiy and Plotkin 2008). Positive selection is still expected to produce ratios greater than one under many circumstances (and importantly, $d_N/d_S > 1$ still generally implies positive selection), but as the strength of selection increases, the relationship reverses and values again approach zero. These results imply that positive selection may be more prevalent in our dataset than the d_N/d_S estimates would suggest, and may also explain why evidence for positive selection using frequency spectrum-based tests does not generally correlate with positive d_N/d_S ratios. However, the model of Kryazhimskiy and Plotkin (2008) relies on the assumption of free recombination among segregating sites, which is unrealistic and the result may be very different if linkage is taken into account. Interestingly, an empirical study comparing within- and between-species d_N/d_S values found a strong correlation between the two (Liu et al. 2008). Clearly, further research is needed to understand the behavior of d_N/d_S within-species. More generally, d_N/d_S ratios, whether within or between species, have long been understood to provide an incomplete picture of the action of selection, as positive selection may be restricted to a particular domain of a given gene, with

purifying selection against non-synonymous substitutions dominating the rest of the sequence. The consequence of pooling the entire sequence in estimating d_N/d_S is that the positive selection is masked, which is why $d_N/d_S > 1$ is rarely seen. Although we observed several examples of significantly negative d_N/d_S ratios, which in light of the above discussion could imply either positive or negative selection, only one gene, *air*, had a significantly positive value for this test, which was the result of 26 non-synonymous substitutions in a window of 530bp, forming a distinct haplotype that only occurred in the most southern population in our study (Redwood, CA) (data not shown). These substitutions were highly homozygous in the two individuals from this population in which they were observed. As apparent heterozygotes are expected where paralogous genes have been amplified, it is unlikely that this is the cause of the large number of non-synonymous SNPs observed for this gene. *air* most closely resembles members of an auxin-induced gene family in *Arabidopsis*. As insensitivity to auxin is one of the features of the dormant cambial meristem (Little and Bonga 1974; Ugglä et al. 1996), positive selection in this gene suggests a possible role for auxin in the variation observed in the growth and dormancy cycle in Sitka spruce.

3.4.5 Functional roles of genes exhibiting outlier behavior

Although the preceding tests highlight potential targets of positive selection among our cold hardiness-related candidate genes, they do not distinguish between globally and locally adaptive variation. High population differentiation provides the strongest evidence that a candidate gene or SNP is locally adaptive. Simulation studies have revealed that the distribution of F_{ST} is not as susceptible to demographic history and patterns of migration (e.g., island vs. stepping-stone models) as originally thought (Beaumont and Nichols 1996), particularly as the number of sampled populations increases (Whitlock 2008). Potential distortions due to variation in migration rates among sampled subpopulations can be accounted for with computational approaches such as the Bayesian model implemented in the program BayesFst. Using BayesFst, we identified 35 SNPs within 14 candidate genes that have unusually high F_{ST} suggestive of the actions of divergent selection among populations. Although our panel of 174 candidate genes is quite diverse with respect to putative functions of the respective gene products, more than a third of the outlier candidate genes fell into the oxidative stress category (Table 3.4). This result is intriguing as radical oxygen species present one of the primary stresses associated with freezing temperatures. Although genes involved in mitigating oxidative stress are often observed to be upregulated during cold acclimation (Holliday et al. 2008), their role in freezing tolerance is not well understood and their importance may be underappreciated. In contrast, apoplastic antifreeze proteins are well studied. These secreted proteins adhere to the surface of ice, preventing its propagation and the concomitant dehydrative and mechanical stress

(Griffith and Yaish 2004). The associated genetic loci are homologous to pathogenesis-related genes, and we observed two such genes, *chib1* and *chib2*, that contained outlier SNPs suggestive of divergent selection. Although we cannot rule out the possibility that these genes may be involved in local adaptation to fungal pathogens, they nevertheless warrant further study as these genes are also strongly upregulated during cold acclimation in Sitka spruce (Holliday et al. 2008).

Among the multifaceted response plants mount to freezing stress, carbohydrate remodeling has been well characterized. Typically, an increase in di- and oligosaccharides is observed in conjunction with starch breakdown, which is thought to facilitate water deficit resistance, among other possible roles (Zwiazek et al. 2001). Three candidate sugar metabolism genes contained outlier SNPs, suggesting a role for carbohydrate metabolism in adaptation to local climate. Collectively, these results support the role of three gene classes known to be vital to cold hardiness, as also potentially being important to adaptive variation in this trait. Targeting these genes in future association studies provides a means to directly determine test their importance to locally adaptive traits. In addition, placement of the SNPs onto a linkage map for Sitka spruce will enhance our understanding of the genomic architecture of local adaptation. Namroud et al. (2008) found that outlier loci related to local adaptation in white spruce were approximately evenly spread across the genome. Understanding the generality of this phenomenon in outcrossing species with high gene flow, such as Sitka and white spruce, will enhance our ability to predict the likelihood of adaptation under climate change. Genetic hitchhiking may facilitate adaptation if adaptive loci are clustered in the genome, and the speed of adaptation would be expected to increase. However, to the extent that local adaptation is driven by standing genetic variation, if the results of Namroud et al. are commonplace, we would expect the rate of adaptation to a changing climate to be substantially slower.

3.4.6 Unequal contributions of peripheral populations to subdivision in Sitka spruce

When we pooled the data for all SNP loci detected in our candidate genes, we found a roughly equivalent estimate for F_{ST} (0.12) to that recently reported for a set of five microsatellite markers (0.11) (Mimura and Aitken 2007a). These values are higher than those from previous studies in Sitka spruce which used isozymes and expressed sequence tag polymorphism (ESTP) markers (Yeh and El-Kassaby 1980; Gapare et al. 2005). Isolation by distance (IBD) was significant in our study, which agrees with Mimura and Aitken (2008), although most pairwise F_{ST} estimates were less than ~0.1. The highest pairwise F_{ST} estimates were those involving the Kodiak Island population (Table 3.4), which is likely because this population was only recently established (less than 400 years

ago) and is therefore still affected by a founder effect (Mimura and Aitken 2007a). This result is also evidenced by the significant population effect (β_j) observed for Kodiak using BayesFst. At the southern edge of the range, the Redwood population did not exhibit striking pairwise F_{ST} values, though they were moderately higher than for populations in the centre of the species range (e.g., Columbia River, Vancouver, Prince Rupert). However, the Redwood population did have a significant positively shifted population effect using BayesFst suggesting lower gene flow or lower effective population size than most populations. Taken together, these results complement previous studies that used neutral markers to study peripheral populations of Sitka spruce, which were found to harbor more spatial genetic structure (Gapare and Aitken 2005) and have higher inbreeding levels than populations within the core of the species range (Mimura and Aitken 2007).

3.4.7 Very low linkage disequilibrium in Sitka spruce

The distance over which LD extends in the genome is an important consideration in designing genetic association studies. In large populations, LD is shaped by mating system (selfing versus outcrossing), the degree of population admixture, and bottlenecks (Flint-Garcia et al. 2003). Population structure can also increase LD in populations with limited gene flow and increased levels of inbreeding. In species with relatively extensive LD in natural populations (e.g., *Arabidopsis thaliana*, *Homo sapiens*), genome-wide scans for associations are possible. Previous studies of the patterns of nucleotide diversity in conifers have found little evidence for extended LD, and our results confirm this. Indeed, our study suggests even lower LD may exist in Sitka spruce than in loblolly pine (Brown et al. 2004). Our estimate for the population recombination parameter ρ (0.049) is more than an order of magnitude larger than that observed for loblolly (where $\rho=0.00115$). Prior to European colonization, the range of loblolly pine was characterized by significantly smaller and relatively isolated populations in wet areas. Contemporary fire suppression regimes and industrial forestry have expanded the current distribution of loblolly to include much of the historical upland, fire-prone range of shortleaf pine (*Pinus palustris*) (Van Lear et al. 2005). The extent of LD in loblolly may therefore reflect historically smaller effective population sizes and restricted gene flow. In contrast, the range of Sitka spruce is continuous over long distances along the west coast, a distribution which has not been fundamentally altered since colonization. The very low levels of LD in Sitka spruce means that dense SNP genotyping will be necessary to have sufficient power to detect associations between genotype and phenotype, and that the haplotype-tagging approach applied with success in human genomics may not be effective in spruce.

3.5 Conclusions

By targeting the coding regions of 174 genes that span a diverse array of putative biological functions, we have characterized nucleotide diversity and linkage disequilibrium for the Sitka spruce genome, and have begun to elucidate the functional categories of genes which have been the targets of positive and diversifying selection. We found that genes putatively involved in light sensing and ABA perception, which were *a priori* strong candidates for local adaptation, tended to be depauperate of nucleotide variation. Although we were not able to reject the null hypothesis of neutral evolution, two of these genes were empirical outliers with respect to nucleotide diversity, suggesting the possibility of non-neutral evolution. In addition, one of the candidate ABA receptors exhibited a significant reduction in nucleotide diversity relative to divergence. We observed substantial population differentiation for several genes putatively involved in oxidative stress and carbohydrate metabolism, as well as two pathogenesis-related loci that may be antifreeze genes. Therefore, diversifying selection on these genes may in part explain the strong clinal variation in locally adaptive traits in Sitka spruce. Future association studies, making use of the resources presented here, will provide a suite of ecologically relevant genetic markers to better understand the clinal variation in locally adaptive traits observed in Sitka spruce, and an important resource for genetic conservation strategies in a changing climate.

Table 3.1. Summary of nucleotide diversity and tests of selective neutrality for 174 candidate genes related to adaptation to local climate in Sitka spruce; values for π , θ , π_{rep} , and π_{silent} given as $\times 10^{-3}$

Totals		Means									Tajima's D	Fay and Wu's H	F _{ST}
Candidate genes	L	S	L	S	H	π	θ	π _{rep}	π _{silent}	d _N /d _S			
174	107,191	1,829	616	10.5	9.5	3.5	4.2	2.1	6.8	0.4	-0.56	-0.85	0.12

L, length of trimmed alignment; S, number of segregating sites; H, number of haplotypes; π, nucleotide diversity; θ, d_N, average number of nonsynonymous substitutions per nonsynonymous site; d_S, average number of synonymous substitutions per synonymous site; π_{rep}, nucleotide diversity for nonsynonymous sites; π_{silent}, nucleotide diversity for synonymous sites; ND, not defined

Table 3.2. Functional categorization of candidate genes with significant tests of selective neutrality; values for π , θ , dN and dS given as x 103; values for selective neutrality tests with single and double stars have Q less than 0.1 and 0.05, respectively, based on p-values obtained by coalescent simulations without recombination for D and H, and based on the Z-test described in the methods for dN/dS.

Gene	BLASTX vs. <i>Arabidopsis</i>	L	S	H	π	θ	π_{rep}	π_{silent}	dN/dS	Tajima's D	Fay and Wu's H
Stress responsive											
<i>lrr1</i>	disease resistance protein	584	27	23	14.5	12.3	8.7	32.2	0.26*	0.62	3.24
<i>lea1</i>	late embryogenesis abundant	638	13	10	1.9	4.3	1.1	1.7	0.30	-1.73	-5.57*
<i>lrr9</i>	leucine-rich repeat	389	4	5	0.6	2.3	0.9	0.0	ND	-1.66**	0.24
<i>tlp2</i>	thaumatin-like	642	37	20	14.7	13.7	7.0	40.9	0.17**	0.25	1.35
<i>gp2</i>	glutathione peroxidase	759	14	12	4.1	7.1	0.3	16.1	0.03*	-0.94	-0.27
<i>gst6</i>	glutathione S-transferase	615	18	17	6.4	6.6	2.5	18.9	0.13**	-0.07	2.86
<i>glp</i>	germin-like protein	662	21	14	5.6	7.4	1.9	17.1	0.11*	-0.77	-0.70
Metabolism											
<i>chs1</i>	chalcone synthase	901	24	22	4.7	6.2	1.5	14.6	0.10**	-0.93	-2.84
<i>sld</i>	$\Delta 8$ sphingolipid desaturase	501	8	8	1.6	3.6	0.6	5.4	0.09	-1.50	-6.40*
<i>gh1</i>	glycosyl hydrolase	1445	10	11	0.5	1.6	0.3	1.2	0.25	-2.01**	-1.14
<i>gh3</i>	glycosyl hydrolase	560	11	7	1.1	4.4	1.2	1.4	0.71	-2.30**	-1.40
<i>gh8</i>	glycosyl hydrolase	712	9	9	1.0	2.9	0.7	2.6	0.27	-1.84*	NA
<i>ifr4</i>	isoflavone reductase	655	10	9	0.9	3.4	0.4	3.2	0.13	-2.14**	-3.27*
<i>pal1</i>	phenylalanine ammonia-lyase	690	21	23	5.5	7.0	0.4	22.4	0.02**	0.72	0.33
Phytohormone-related											
<i>aec1</i>	auxin efflux carrier	765	14	15	1.7	4.4	1.1	3.7	0.29	-1.92*	1.12
<i>air</i>	auxin-responsive	1820	53	16	5.6	7.0	9.2	3.4	2.76**	-0.65	-4.30
<i>arg1</i>	auxin/aluminum-responsive	773	28	29	8.7	8.2	3.4	17.3	0.20*	0.22	2.41
<i>arg4</i>	auxin down-regulated	503	3	3	0.4	1.4	0.1	1.2	0.08	-1.57*	0.18
Signal transduction											
<i>cipk2</i>	CBL-interacting protein kinase	847	9	5	0.8	2.5	0.6	1.0	0.60	-1.93*	0.67
<i>cbl1</i>	calcineurin B-like	457	8	7	1.4	4.0	1.3	1.9	0.63	-1.80*	0.61
<i>cry1</i>	cryptochrome	509	3	4	0.3	1.3	0.2	1.0	0.20	-1.58*	0.17
Transcriptional regulation											
<i>hta2</i>	histone H1-3	860	34	26	5.5	8.9	3.2	11.6	0.28	-1.30	-21.78*
<i>bhlh1</i>	basic helix-loop-helix	669	9	8	1.1	3.0	0.4	0.4	1.00	-1.80*	-5.22*
<i>bhlh2</i>	basic helix-loop-helix	243	12	5	22.0	12.5	12.2	50.4	0.32	2.50*	1.32
<i>nam1</i>	no apical meristem family	604	7	6	1.9	2.6	0.0	8.6	0.00*	-0.74	-2.18
<i>swap</i>	SWAP-domain containing	835	9	6	1.1	2.7	0.9	2.0	0.56	-0.78	-8.50**
<i>hct</i>	ubiquitin-transferase	750	6	6	0.5	1.8	0.2	0.9	0.22	-1.94**	-1.53
Miscellaneous											
<i>abc4</i>	ABC transporter	559	5	6	0.6	2.2	0.4	0.0	ND	-1.89**	NA
<i>int2</i>	inositol transporter	429	6	3	6.9	3.4	NA	NA	0.30	2.87**	NA

L, length of trimmed alignment; S, number of segregating sites; Hap, number of haplotypes; π , nucleotide diversity; θ , π_{rep} , nucleotide diversity for nonsynonymous sites; π_{silent} , nucleotide diversity for synonymous sites; ND, not defined; NA, no outgroup available to perform test

Table 3.3. Model comparison of maximum-likelihood HKA test between a neutral model, and models allowing selection on 6 and 5 loci. Likelihoods and selection parameters, k , given as average across three model runs

Model	Description	ln L	Model Comparison	Likelihood-ratio statistic (d.f.)	P	k					
						<i>lrr6</i>	<i>pip5k</i>	<i>rs1</i>	<i>abc5</i>	<i>nsh1</i>	<i>chlh1</i>
A	Neutral	-73.94									
B	Selection, 5 candidate genes	-68.1	A vs. B	11.7 (5)	0.039	0.16	0.57	0.07	0.43	1	0.26
C	Selection, 6 candidate genes	-67.41	A vs. C	13.1 (6)	0.042	0.45	0.30	0.25	0.34	0.79	0.3696
			B vs. C	1.4 (1)	0.237						

Table 3.4. SNP outliers detected using 'BayesFst'

Contig or EST ID	BLASTX vs. <i>Arabidopsis</i>	Significant SNPs	F_{ST} (SNP) ^a	F_{ST} (Gene) ^b	α_i ^a	P-value ^a
<i>Diversifying selection</i>						
<i>gp1</i>	glutathione peroxidase	4	0.38	0.21	1.21	0.053
<i>gp2</i>	glutathione peroxidase	1	0.37	0.19	0.85	0.098
<i>per2</i>	peroxidase	4	0.39	0.15	1.78	0.061
<i>per3</i>	peroxidase	13	0.41	0.45	1.30	0.043
<i>chib1</i>	basic endochitinase	3	0.34	0.42	0.93	0.069
<i>chib2</i>	basic endochitinase	1	0.40	0.55	1.26	0.052
<i>gh2</i>	glycosyl hydrolase	1	0.35	0.35	1.08	0.075
<i>gh4</i>	glycosyl hydrolase	1	0.33	0.20	0.51	0.082
<i>abc2</i>	ABC transporter	1	0.35	0.06	0.85	0.090
<i>gram</i>	GRAM domain-containing	2	0.42	0.23	1.74	0.033
<i>exl1</i>	extensin	1	0.31	0.25	0.98	0.099
<i>xth1</i>	xyloglucan:xyloglucosyl transferase	1	0.37	0.33	1.05	0.056
<i>ifr2</i>	isoflavone reductase	1	0.35	0.21	1.08	0.078
<i>nam2</i>	no apical meristem family	1	0.34	0.60	0.94	0.094
<i>Homogenizing selection</i>						
<i>ugt2</i>	UDP-glucuronosyl/UDP-glucosyl transferase	2	0.08	-0.03	-0.94	0.093
<i>abc5</i>	ABC transporter	2	0.08	-0.15	-0.90	0.083
^a F_{ST} , α_i and probability values given as averages for genes with greater than one significant SNP						
^b F_{ST} for entire gene sequence from Table S1						

Table 3.5. Pairwise F_{ST} estimates based on all SNP loci for the six geographic populations represented in this study

	RW	CR	VA	PR	VD	KD
RW	*					
CR	0.0931	*				
VA	0.0767	0.0352	*			
PR	0.0783	0.0524	0.0526	*		
VD	0.1095	0.0761	0.0999	0.0922	*	
KD	0.1454	0.1459	0.1349	0.1352	0.1229	*

Origins of study populations: RW, Redwood, CA, USA; CR, Columbia River, OR, USA; VA, Vancouver, BC, Canada; PR, Prince Rupert, BC, Canada; VD, Valdez, AK, USA; KD, Kodiak Island, AK, USA

Figure 3.1. Natural range of Sitka spruce and origins of study populations. Samples were collected from plants growing in a common garden in Vancouver.

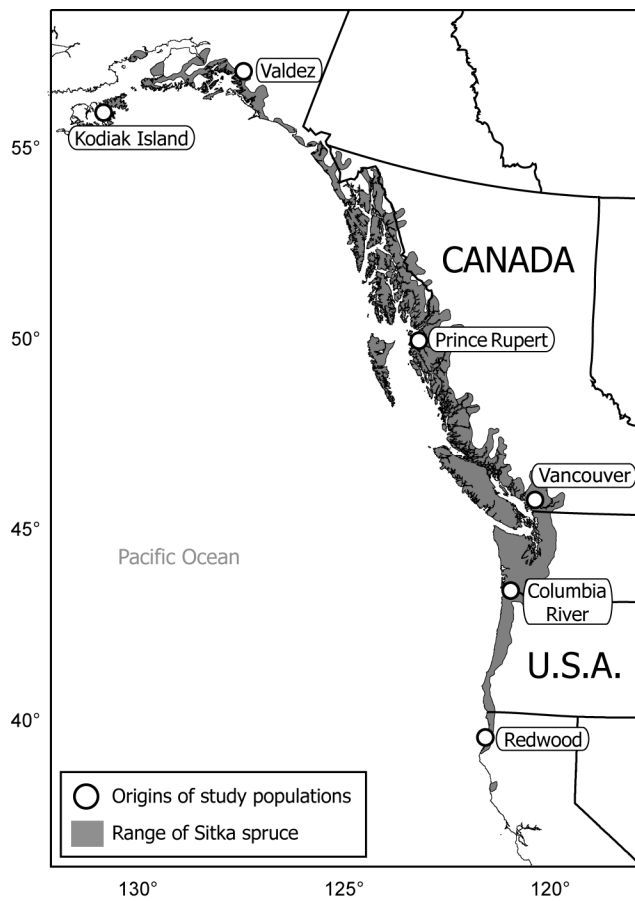


Figure 3.2. Values for Tajima's D and Fay and Wu's H (only the 148 genes for which an outgroup was available) for candidate genes described in this study. Single asterisk indicates $Q < 0.1$ and double asterisk indicates $Q < 0.05$.

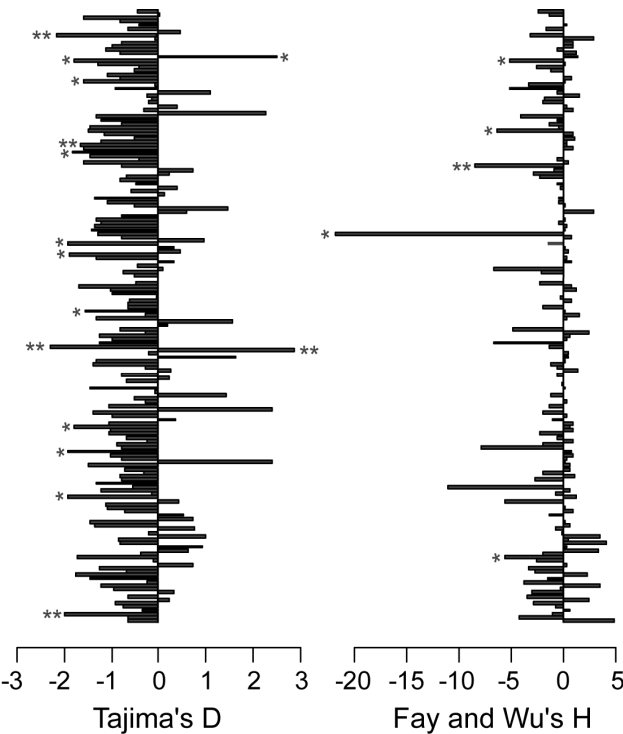


Figure 3.3. Comparison of within-species nucleotide diversity (θ) and average total between-species divergence (D_{xy}) for 148 candidate genes with available white spruce outgroup sequences. Highlighted points were tested for evidence of selection using MLHKA test against a neutral test set for which no deviation was observed (see Table 3.3). Correlation between diversity and divergence (solid line) was significant ($R^2=0.26$, $P<0.001$).

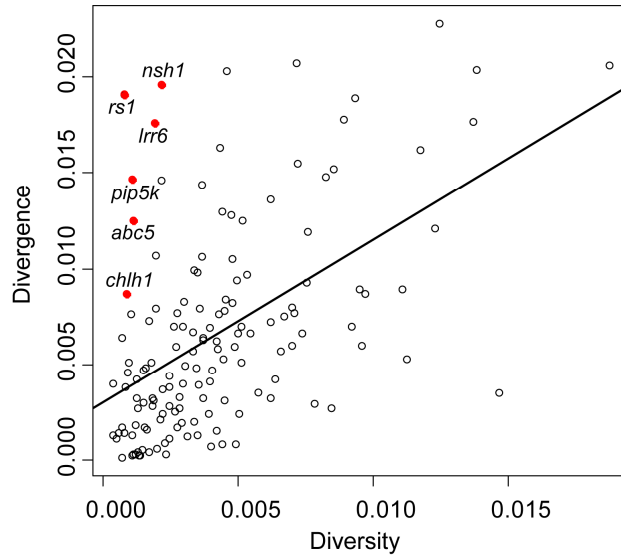


Figure 3.4. Results of Bayesian outlier detection. (a) Estimate of F_{ST} plotted against transformed p-values. The dashed and solid vertical lines indicate the 10% and 5% Bayesian probability levels, respectively. Hence, points highlighted in blue were significant at the 10% level, while those highlighted in red were significant at the 5% level (b) Posterior density curves for the population parameter β_j for each of six populations used in this study (RW, Redwood, CA, USA; CR, Columbia River, OR, USA; VA, Vancouver, BC, Canada; PR, Prince Rupert, BC, Canada; VD, Valdez, AK, USA; KD, Kodiak Island, AK, USA). The prior mean was set at -2.0.'

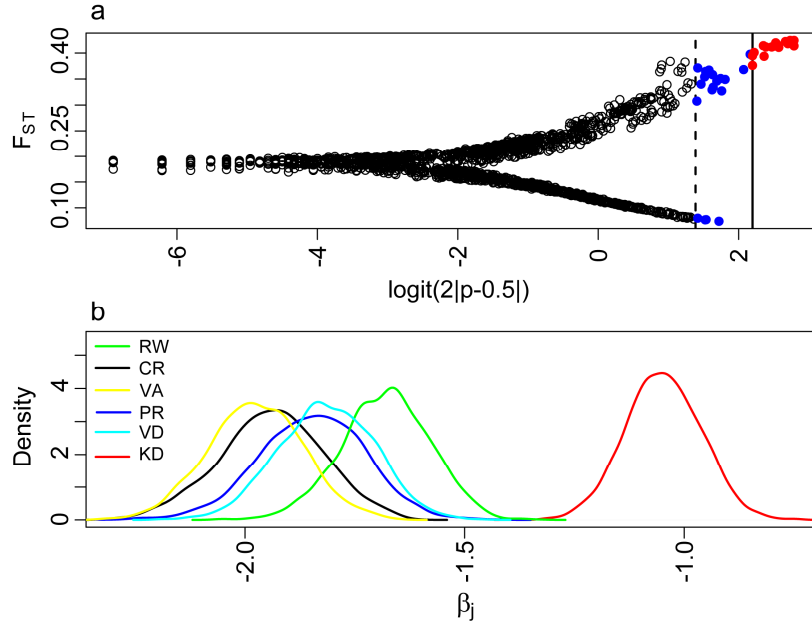


Figure 3.5. Mantel test of isolation by distance for six populations of Sitka spruce based on pairwise F_{ST} estimates using the 1829 SNPs described in this study.

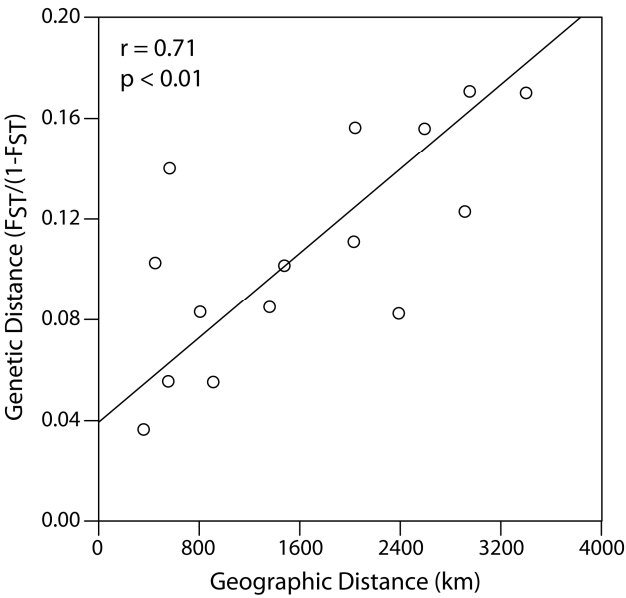
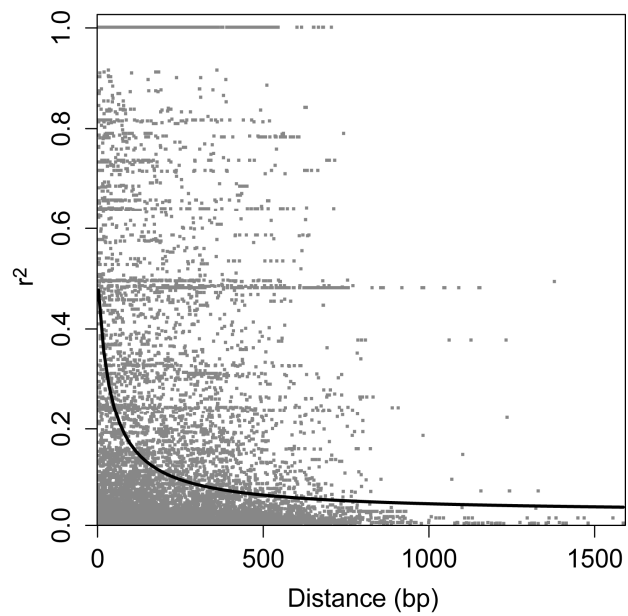


Figure 3.6. Pairwise within-gene linkage disequilibrium estimates for the 1829 SNPs found in 174 candidate genes described in this study. Best fit curve based on Hill and Weir (1988).



3.6 Literature cited

- Akey, J. M., G. Zhang, K. Zhang, L. Jin, and M. D. Shriver. 2002. Interrogating a high-density SNP map for signatures of natural selection. *Genome Research* **12**:1805-1814.
- Beaumont, M. A., and D. J. Balding. 2004. Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13**:969-980.
- Beaumont, M. A., and R. A. Nichols. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B-Biological Sciences* **263**:1619-1626.
- Brown, G. R., G. P. Gill, R. J. Kuntz, C. H. Langley, and D. B. Neale. 2004. Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proceedings of the National Academy of Sciences of the United States of America* **101**:15255-15260.
- Close, T. J. 1997. Dehydrins: A commonality in the response of plants to dehydration and low temperature. *Physiologia Plantarum* **100**:291-296.
- Eveno, E., C. Collada, M. A. Guevara et al. 2008. Contrasting patterns of selection at *Pinus pinaster* Ait. drought stress candidate genes as revealed by genetic differentiation analyses. *Molecular Biology and Evolution* **25**:417-437.
- Excoffier, L., G. Laval, and D. Balding. 2003. Gametic phase estimation over large genomic regions using an adaptive window approach. *Human Genomics* **1**:7-19.
- Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**:47-50.
- Fay, J. C., and C. I. Wu. 2000. Hitchhiking under positive Darwinian selection. *Genetics* **155**:1405-1413.
- Flint-Garcia, S. A., J. M. Thornsberry, and E. S. Buckler. 2003. Structure of linkage disequilibrium in plants. *Annual Reviews of Plant Biology* **54**:357-374.
- Gapare, W. J., and S. N. Aitken. 2005. Strong spatial genetic structure in peripheral but not core populations of Sitka spruce (*Picea sitchensis* (Bong.) Carr.) *Molecular Ecology* **14**:2659-2667.
- Gapare, W. J., S. N. Aitken, and C. E. Ritland. 2005. Genetic diversity of core and peripheral Sitka spruce (*Picea sitchensis* (Bong.) Carr.) populations: implications for conservation of widespread species. *Biol Cons* **123**:113-123.
- Griffith, M., and M. W. Yaish. 2004. Antifreeze proteins in overwintering plants: a tale of two activities. *Trends in Plant Science* **9**:399-405.
- Heuertz, M., E. De Paoli, T. Kallman, H. Larsson, I. Jurman, M. Morgante, M. Lascoux, and N. Gyllenstrand. 2006. Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [*Picea abies* (L.) Karst]. *Genetics* **174**:2095-2105.
- Hill, W. G., and B. S. Weir. 1988. Variances and covariances of squared linkage disequilibria in finite populations. *Theoretical Population Biology* **33**:54-78.
- Holliday, J. A., S. G. Ralph, R. White, J. Bohlmann, and S. N. Aitken. 2008. Global monitoring of autumn gene expression within and among phenotypically divergent populations of Sitka spruce (*Picea sitchensis*). *New Phytologist* **178**:103-122.
- Hudson, R. R., M. Kreitman, and M. Aguade. 1987. A Test of Neutral Molecular Evolution Based on Nucleotide Data. *Genetics* **116**:153-159.
- Hudson, R. R., M. Slatkin, and W. P. Maddison. 1992. Estimation of Levels of Gene Flow from DNA-Sequence Data. *Genetics* **132**:583-589.
- Ingvarsson, P. K. 2008. Multilocus Patterns of Nucleotide Polymorphism and the Demographic History of *Populus tremula*. *Genetics* **180**:329-340.
- Johnsen, O., O. G. Daehlen, G. Ostreng, and T. Skroppa. 2005a. Daylength and temperature during seed production interactively affect adaptive performance of *Picea abies* progenies. *New Phytologist* **168**:589-596.

- Johnsen, O., C. G. Fossdal, N. Nagy, J. Molmann, O. G. Daehlen, and T. Skroppa. 2005b. Climatic adaptation in *Picea abies* progenies is affected by the temperature during zygotic embryogenesis and seed maturation. *Plant Cell and Environment* **28**:1090-1102.
- Kelley, J. L., and W. J. Swanson. 2008. Positive selection in the human genome: From genome scans to biological significance. *Annual Review of Genomics and Human Genetics* **9**:143-160.
- Kolosova, N., B. Miller, S. Ralph, B. Ellis, C. Douglas, K. Ritland, and J. Bohlmann. 2004. Isolation of high-quality RNA from gymnosperm and angiosperm trees. *Biotechniques* **36**:821-824.
- Krutovsky, K. V., and D. B. Neale. 2005. Nucleotide diversity and linkage disequilibrium in cold-hardiness- and wood quality-related candidate genes in Douglas fir. *Genetics* **171**:2029-2041.
- Little, C. H. A., and J. M. Bonga. 1974. Rest in Cambium of *Abies balsamea*. *Canadian Journal of Botany-Revue Canadienne De Botanique* **52**:1723-1730.
- Liu, J., Y. Zhang, X. Lei, and Z. Zhang. 2008. Natural selection of protein structural and functional properties: a single nucleotide polymorphism perspective. *Genome Biology* **9**:R69.
- Lynch, D. V., and T. M. Dunn. 2004. An introduction to plant sphingolipids and a review of recent advances in understanding their metabolism and function. *New Phytologist* **161**:677-702.
- Mimura, M., and S. N. Aitken. 2007. Adaptive gradients and isolation-by-distance with postglacial migration in *Picea sitchensis*. *Heredity* **99**:224-232.
- Morgenstern, E. K. 1996. Geographic variation in forest trees: genetic basis and application of knowledge in silviculture. University of British Columbia Press, Vancouver.
- Namroud, M. C., J. Beaulieu, N. Juge, J. Laroche, and J. Bousquet. 2008. Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Molecular Ecology* **17**:3599-3613.
- Nielsen, R. 2005. Molecular signatures of natural selection. *Annual Review of Genetics* **39**:197-218.
- Pavy, N., L. S. Parsons, C. Paule, J. MacKay, and J. Bousquet. 2006. Automated SNP detection from a large collection of white spruce expressed sequences: contributing factors and approaches for the categorization of SNPs. *BMC Genomics* **7**.
- Pyhajarvi, T., M. R. Garcia-Gil, T. Knurr, M. Mikkonen, W. Wachowiak, and O. Savolainen. 2007. Demographic history has influenced nucleotide diversity in European *Pinus sylvestris* populations. *Genetics* **177**:1713-1724.
- Ralph, S. G., H. J. E. Chun, N. Kolosova et al. 2008. A conifer genomics resource of 200,000 spruce (*Picea* spp.) ESTs and 6,464 high-quality, sequence-finished full-length cDNAs for Sitka spruce (*Picea sitchensis*). *BMC Genomics* **9**.
- Ralph, S. G., H. Yueh, M. Friedmann et al. 2006. Conifer defence against insects: microarray gene expression profiling of Sitka spruce (*Picea sitchensis*) induced by mechanical wounding or feeding by spruce budworms (*Choristoneura occidentalis*) or white pine weevils (*Pissodes strobi*) reveals large-scale changes of the host transcriptome. *Plant Cell and Environment* **29**:1545-1570.
- Reed, D. H., and R. Frankham. 2001. How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution* **55**:1095-1103.
- Richard, S., M. J. Morency, C. Drevet, L. Jouanin, and A. Seguin. 2000. Isolation and characterization of a dehydrin gene from white spruce induced upon wounding, drought and cold stresses. *Plant Molecular Biology* **43**:1-10.
- Rozas, J., J. C. Sanchez-DelBarrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**:2496-2497.
- Rozen, S., and H. J. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. Pp. 365-386 in S. Krawetz, and S. Misener, eds. *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ.
- Savolainen, O., and T. Pyhajarvi. 2007. Genomic diversity in forest trees. *Current Opinion in Plant Biology* **10**:162-167.
- Savolainen, O., T. Pyhajarvi, and T. Knurr. 2007. Gene flow and local adaptation in trees. *Annual Review of Ecology Evolution and Systematics* **38**:595-619.

- Shen, Y. Y., X. F. Wang, F. Q. Wu et al. 2006. The Mg-chelatase H subunit is an abscisic acid receptor. *Nature* **443**:823-826.
- Storey, J. D., and R. Tibshirani. 2003. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America* **100**:9440-9445.
- Tajima, F. 1989. Statistical-Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics* **123**:585-595.
- Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**:1596-1599.
- Tang, J., B. Vosman, R. Voorrips, C. G. van der Linden, and J. Leunissen. 2006. QualitySNP: a pipeline for detecting single nucleotide polymorphisms and insertions/deletions in EST data from diploid and polyploid species. *BMC Bioinformatics* **7**:438.
- Tang, J. F., J. A. M. Leunissen, R. E. Voorrips, C. G. van der Linden, and B. Vosman. 2008. HaPLoSNPer: a web-based allele and SNP detection tool. *Bmc Genetics* **9**.
- Uemura, M., R. A. Joseph, and P. L. Steponkus. 1995. Cold acclimation of *Arabidopsis thaliana* - Effect on plasma membrane lipid composition and freeze induced lesions. *Plant Physiology* **109**:15-30.
- Uemura, M., and P. L. Steponkus. 1994. A contrast of the plasma membrane lipid composition of oat and rye leaves in relation to freezing tolerance. *Plant Physiology* **104**:479-496.
- Uggla, C., T. Moritz, G. Sandberg, and B. Sundberg. 1996. Auxin as a positional signal in pattern formation in plants. *Proceedings of the National Academy of Sciences of the United States of America* **93**:9282-9286.
- Van Lear, D. H., W. D. Carroll, P. R. Kapeluck, and R. Johnson. 2005. History and restoration of the longleaf pine-grassland ecosystem: Implications for species at risk. *Forest Ecology and Management* **211**:150-165.
- Weiser, C. J. 1970. Cold resistance and injury in woody plants. *Science* **169**:1269-1278.
- Welling, A., P. Rinne, A. Vihera-Aarnio, S. Kontunen-Soppela, P. Heino, and E. T. Palva. 2004. Photoperiod and temperature differentially regulate the expression of two dehydrin genes during overwintering of birch (*Betula pubescens* Ehrh.). *J. Exp. Bot.* **55**:507-516.
- Whitlock, M. C. 2008. Evolutionary inference from Q(ST). *Molecular Ecology* **17**:1885-1896.
- Wright, S. I., and B. Charlesworth. 2004. The HKA test revisited: A maximum-likelihood-ratio test of the standard neutral model. *Genetics* **168**:1071-1076.
- Yeh, F. C., and Y. A. El-Kassaby. 1980. Enzyme variation in natural populations of Sitka spruce (*Picea sitchensis*). 1. Genetic variation patterns among trees from 10 IUFRO provenances. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere* **10**:415-422.
- Zwiazek, J. J., S. Renault, C. Croser, J. Hansen, and E. Beck. 2001. Biochemical and Biophysical Changes in Relation to Cold Hardiness. Pp. 165-186 in F. J. Bigras, and S. J. Colombo, eds. *Conifer Cold Hardiness*. Kluwer Academic Publishers, Dordrecht.

4. Association mapping of autumn cold hardiness and budset timing in Sitka spruce (*Picea sitchensis*)³

4.1 Introduction

Adaptation to local climate in temperate and boreal forest trees is determined by a suite of component traits that together determine the relative susceptibility of local populations to freezing temperatures, while enabling sufficient annual growth. As growth and cold hardiness are incompatible, this susceptibility is a function of the timing of growth initiation in spring and growth cessation in fall, timing of cold acclimation in fall, and, the maximum cold hardiness achieved in mid-winter (Weiser 1970). Quantitative trait loci (QTL) mapping suggests that the genomic underpinnings of these traits are complex, and involve many genes with typically small individual effect sizes (Jermstad et al. 2001a; Jermstad et al. 2001b; Wheeler et al. 2005). Although the QTL mapping approach enables a genome-wide appraisal of the trait of interest, the associations uncovered are not transferable among mapping populations or to natural stands (Neale and Savolainen 2004). Hence, the usefulness of the QTL approach, for example as a tool for marker aided selection or conservation genetics, is limited. Recent advances in forestry genomics, including the generation of large numbers of expressed sequence tags (ESTs) and full-length cDNAs (Ralph et al. 2006b; Ralph et al. 2008a; Ralph et al. 2008b) have made it feasible to extend the association mapping approach, applied successfully in numerous model systems, to conifers.

Although conifers are a difficult experimental system for most applications, they are well-suited to genetic association studies. Temperate and boreal tree species in particular typically have large, random mating populations that remain for the most part undomesticated. This contrasts with the situation in other species of economic importance such as crop plants and livestock, which have gone through domestication bottlenecks. Association mapping is similar to QTL mapping in that both methods track the co-segregation of marker alleles and phenotypic variants, but whereas QTL mapping uses family pedigrees with relatively few meioses and hence high linkage disequilibrium (LD), association mapping uses random-mating populations with complex, unknown familial relationships and low levels of LD (Lander and Schork 1994; Neale and Savolainen 2004). QTL mapping therefore has limited usefulness for breeding applications and genetic conservation strategies because the associations cannot be transferred between populations.

³ A version of this chapter will be submitted for publication. Holliday, J.A. and Aitken, S.N. Association mapping of autumn cold hardiness and budset timing in Sitka spruce (*Picea sitchensis*)

Conifer breeding programs that make use of phenotypic selection are for the most part in their second or third generation of breeding. Although this approach has realized significant genetic gains in economically important traits, the potential exists to integrate molecular markers into the process. Identification of markers that govern trait variation may accelerate the breeding process for traits that are not readily assayed phenotypically until the plants are mature (e.g., wood quality, pest resistance), as well as provide for the infusion of novel alleles, captured from wild populations, into the germplasm of breeding programs. The latter is of increasing concern under climate change, where rapid warming is expected to dramatically alter climate landscapes. For the same reason, maintenance of standing variation in natural populations will be crucial with climate change. This goal is normally advanced by using anonymous selectively neutral marker alleles to determine, for example, the extent of genetic diversity in existing natural populations, and the appropriate size of a protected area to maintain that variation. The tacit assumption is that genetic diversity encapsulated by these neutral marker surveys does well to describe the standing variation in heritable fitness-related traits. However, meta-analyses conclude this assumption may not be well founded (Reed and Frankham 2001). Hence, the search for ecologically relevant genetic markers, i.e., those that directly determine the traits or are linked to the loci that do, has taken on additional importance beyond the more general effort to understand the genomic basis and population structure underlying the phenotypic variation in complex traits.

Association studies are currently being pursued in both gymnosperm and angiosperm tree species, but are in their infancy. To date only a few studies have been published describing genetic association with phenotypic traits (Gonzalez-Martinez et al. 2007; Gonzalez-Martinez et al. 2008; Ingvarsson et al. 2008). Here we report the application of association mapping to two dormancy-related traits, namely, cold hardiness and budset timing. We have assayed 339 single nucleotide polymorphisms in a mapping population of 410 individuals. The results of this study provide a suite of genetic markers that can be applied in both marker-aided selection (MAS) and conservation genetics, and represent a substantial forward step in our understanding of the genomic underpinnings of complex adaptive traits in conifers.

4.2 Materials and Methods

4.2.1 Plant material and tissue sampling

A common garden was established in 2003 consisting of ~1000 open-pollinated offspring of 10-13 seed parents per population from 17 geographic populations (Mimura and Aitken 2007a). Due to size limitations, this common garden was thinned in 2005, leaving 14 populations and a total of

410 plants (Figure 4.1 and Table 4.1). We subsequently vegetatively propagated each of the plants by detaching and rooting current year lateral shoots in January, 2006. Lower needles were removed and a cuttings were dipped in a 0.01% solution of indole-3-butyric acid prior to planting in a 3:2 mixture of peat:perlite. Cuttings were kept in a greenhouse at ambient temperature with supplemental bench heating to keep the soil temperature between 16°C and 20°C throughout the rooting process. The ramets were misted twice daily for 30min to limit transpirational water loss, and once rooted, were moved to an outdoor patio, and subsequently transplanted to a permanent outdoor common garden. Tissue for genomic DNA extraction was sampled from newly flushing lateral buds from the plants in the original common garden, in May, 2006.

4.2.2 Selection of single nucleotide polymorphisms

Single nucleotide polymorphisms (SNPs) to be genotyped in the mapping population were chosen from the panel of 174 candidate genes described in chapter 3 of this thesis. In addition, a further 28 genes were also screened for SNPs. The latter did not have sufficient quality sequence reads for population genetic analysis, but nevertheless had enough acceptable sequences to detect common SNPs. Three criteria were used to select SNPs for GoldenGate assay design. First, the minor allele frequency (MAF) threshold in the discovery panel (Chapter 3, this thesis) was set at 5%. Second, a minimum spacing of ~60 base-pairs (bp) was set due to limitations of the GoldenGate platform (see below). Finally, nonsynonymous (i.e., amino acid changing) SNPs were preferred over synonymous SNPs, as the former are more likely to have a functional consequence for the gene product that could then be a target of selection. It should be noted, however, that positive selection may increase local linkage disequilibrium and therefore synonymous substitutions adjacent to a positively selected site are expected to in some cases associate with the phenotype due to this linkage. This is the basis for the haplotype-tagging approach applied successfully in humans and other model systems. We therefore included many synonymous substitutions. However, we chose not to employ haplotype-tagging explicitly, as LD decays very rapidly in Sitka spruce (Chapter 3, this thesis), and as a result the decrease in genotyping effort would not be that substantial compared to the quasi-saturated approach described above. In addition, some potentially informative SNPs are inevitably lost through haplotype-tagging.

4.2.3 Phenotyping for association tests

Prior to transplanting of the rooted cuttings described above from the outdoor patio to the common garden, cold hardiness for each individual in the common garden was measured on October 1 and December 1, 2007, using the artificial freeze test described in Chapter 2 of this thesis (Hannerz

et al. 1999). To account for increasing cold hardiness during the fall, different temperatures were tested on the two dates: -5°C, -10°C, and -15°C on October 1st, and -5°C, -10°C, and -20°C on December 1st. It should be noted that ambient temperature fell below freezing on several dates prior to the December 1st sampling, but not prior to the October 1st sampling. Budset timing was assessed in fall 2008, by recording the Julian date on which bud scales became visible on the apical bud of each rooted cutting. This was done on a weekly basis beginning July 7th, 2008, through December 1st, 2008.

4.2.4 Illumina GoldenGate genotyping

SNP genotyping was carried out using the Illumina bead array platform in conjunction with the GoldenGate allele specific assay in a 96-well, 768 SNP format (Fan et al. 2003; Shen et al. 2005). This is a highly multiplexed assay, and involves adherence of genomic DNA to a solid support, followed by annealing of a locus-specific oligo (LSO) and two allele-specific oligos (ASO) for each SNP to be assayed. Allele-specific extension is then carried out, followed by amplification of the resulting PCR template using three primers specific to the ASOs and LSOs. These primers carry fluorescent moieties which enable detection of alternative homozygotes as well as heterozygotes. 'Address sequences' unique to each set of PCR primers then enable hybridization of the PCR products to a standard array, which carries 30-fold replication of beads for each SNP. SNP calls are thus made on the basis of the clustering of fluorescent signals from replicate beads, and call quality is evaluated on the basis of the separation between homozygous and heterozygous clusters. The minimum threshold for this 'GenTrain' score was set at 0.25 for the current assay. In addition, SNPs were only accepted that had call rates of greater than 90%, though most SNPs retained for further analysis had call rates in excess of 95%. Although we selected SNPs from the discovery panel with minor allele frequencies (MAFs) above 5%, following GoldenGate genotyping it was nevertheless found that in the total mapping population, some SNPs nevertheless had MAFs < 5%. Although there is little power to detect associations for these, alleles of large effect size may give significant results, and as a compromise we chose a minimum threshold of 1% for a SNP to be included in the association tests.

4.2.5 Data analysis

As population structure can increase type I error in association studies due to the stochasticity of genetic drift in isolated subpopulations (Neale and Savolainen 2004), we explored population structure in our mapping population using the program STRUCTURE (Pritchard et al. 2000). This method assigns individuals to one or more genetic clusters on the basis of their multilocus genotype.

To avoid confounding our inferences about demography with selection on our candidate SNPs, we used a separate cohort of 98 SNPs for the purposes of running STRUCTURE. These SNPs were included on the Illumina array for the purposes of genetic mapping in a separate population not part of this study, and were not *a priori* candidates for association tests. The same thresholds for quality were applied to these SNPs as those to be tested for associations with phenotypes. Models with a putative number of clusters (k) from one to 15 were tested. 50,000 and 100,000 iterations were used for the pre- and post-burn in periods, respectively.

Association analysis was carried out using the mixed model framework of Yu et al. (2006), implemented in the program TASSEL, version 2.1 (<http://www.maizegenetics.net>). A mixed linear model (MLM) was fitted for each marker and phenotype. In fitting the model, we log transformed October cold hardiness to achieve a more normal distribution, but this was not necessary for either budset or December cold hardiness (October cold hardiness was bimodally distributed due to the northern populations having begun cold acclimation, whereas the southern populations had not).

To account for the confounding effects of population structure, the relative assignment of each individual to each genetic cluster (the ‘Q’ matrix of STRUCTURE, where k=3) was incorporated into the model as a covariate. Second, pairwise kinship coefficients were modeled as random effects. For a largely outcrossing species with a continuous range and high gene flow, such as Sitka spruce, highly variable levels of kinship are expected. As such, the ability to account for multiple levels of relatedness further reduces the incidence of spurious associations, particularly as we have a rangewide sample of genotypes in this study. To incorporate kinship into our association analysis, we used the method Lynch and Ritland (1999), as implemented in the program SPAGeDI (Hardy and Vekemans 2002) (<http://www.ulb.ac.be/sciences/ecoevol/spagedi.html>), which was estimated using the same 98 SNPs used for estimating population structure. The mixed model can be expressed as

$$y = S\alpha + Qv + Z\mu + \varepsilon$$

where y is vector of phenotypic observations, α is a vector of SNP effects, v is a vector of population effects, μ is a vector of kinship effects, ε is a vector of residuals, and S, Q and Z are incidence matrices of 1’s and 0’s relating y to α , v, μ , respectively. Probability values obtained from this model were adjusted for multiple testing using the false discovery rate (FDR) implemented in the ‘qvalue’ package in R (<http://www.r-project.org/>) (Storey and Tibshirani 2003).

4.3 Results

4.3.1 Phenotypic trait variation

Highly significant clinal variation was observed between study population origin and each of the three measured phenotypic traits (Figure 4.2). Distance from the southern limit of the species range explained 71%, 91%, and 81% of the variation in budset, October cold hardiness, and December cold hardiness, respectively ($p < 0.001$ for each). Although variation in each of these traits exists within and among the Alaska populations, the clines tended to flatten in this northern portion of the range (Figure 4.2). As the coastline runs more east-west in this area, we hypothesized that latitude may be a better predictor of phenotype for these populations. However, regressions of each trait on latitude for only the Alaska populations revealed no significant relationships (data not shown).

4.3.2 Genotyping

Of the 768 SNPs assayed using the GoldenGate platform, a total of 476 had call rates above 90%, GenTrain quality scores above 0.25, and were polymorphic. This success rate is comparable with other studies in non-model systems. For example, a recent 768 SNP GoldenGate assay in white spruce gave 534 successful SNPs that were polymorphic (Namroud et al. 2008). Among the successful SNPs in the current study were the 98 to be used for assessing population structure and kinship, and hence these were excluded from the association tests. An additional 39 had MAFs less than 1% and were also excluded. The remaining 339 SNPs to be tested for association were distributed across 153 genes, giving an average of 2.2 SNPs per gene.

4.3.4 Population structure and kinship

Clustering of multilocus genotypes using STRUCTURE revealed no clear solution to the number of genetic populations present in our rangewide collection of Sitka spruce. That is, the log-likelihood for successive models (from $k=1$ to $k=15$) increased steadily, without an obvious plateau. However, a large increase in the likelihood was observed between $k=2$ and $k=3$ (mean $\Delta \text{LnP(D)} = 467.5$), followed by generally smaller increases for models assuming $k > 3$ (mean $\Delta \text{LnP(D)}$ between ~ 25 and 75). This suggested that $k=3$ may be the best solution. Simulation studies have revealed that this rate of change between successive values for k provides an accurate way to assess the ‘true’ number of clusters in a metapopulation in the absence of a clear answer based on the likelihood values (Evanno et al. 2005). Indeed, by applying this method we found that $k=3$ provides the best solution. Figure 4.3 gives a visual representation of this scenario, which splits the 14 population studied into three clusters, the first including all populations in California, Oregon, and British

Columbia, the second including all populations in Alaska except Kodiak Island, and the final cluster comprising the Kodiak Island population alone. It should be noted that these splits are robust to the number of clusters assumed. We therefore included the Q-matrix from STRUCTURE, where $k=3$, as the covariate in our model described in the methods.

As our mapping population consists of open-pollinated families collected from wild populations, we expect complex familial relationships to be present, and this is evidenced by the continuously descending kinship estimates (Figure 4.4). Though most were close to zero, indicating unrelatedness, ~5% had values between 0.2 and 0.3, suggesting either half-sib or parent-offspring relationships, and a few had values of ~0.5, suggesting full-sibs.

4.3.5 Associations between phenotype and genotype

Forty-five significant associations ($Q < 0.1$) in 28 genes were detected out of the 339 SNPs and the three phenotypic traits tested. Among these, 16 SNPs were unique to budset timing, nine were unique to December cold hardiness, and 10 were shared between these two traits (Figure 4.5). Surprisingly, after multiple test correction, no significant associations were detected for October cold hardiness. SNPs associated with budset explained between 1.0% and 4.3% of the phenotypic variance in this trait, whereas those associated with December cold hardiness had values ranging from 0.7% to 5.4% (Table 4.2). Cumulatively, these associations describe 34.4% of the phenotypic variation in budset and 28.1% of the variation in December cold hardiness in our mapping population.

The largest effect sizes were found for SNPs within a xyloglucan:xyloglucosyl transferase (*xth1*) and a peroxidase (*per6*), which also had relatively high percentage of variance explained (PVE) for both budset and December cold hardiness. A synonymous SNP at position 350 in *xth1* had the highest PVE of any SNP in our study, at 5.4% for December cold hardiness and 4.3% for budset. A neighboring SNP had PVE of 4.2% and 3.1% for each of these traits, respectively (Table 4.2). In total, four SNPs were associated with both budset and December cold hardiness in this gene. *per3* had the second highest PVE among associated SNPs in this study, at 5.1% for budset and 2.6% for December cold hardiness. This gene was one of six putatively involved in mitigation of oxidative stress that harbored SNPs associated either or both of the phenotypic traits we assayed (the others being *gp4*, *gp5*, *gst4*, *ifr2*, and *per6*).

Two other functional categories contained several genes with significant associations with phenotypes, namely, light signaling and auxin-related genes. Several canonical dormancy-related genes fell in the former category, including a gene similar to *PHYTOCHROME A* (*phya*), a *CONSTANS*-like gene (*co*), a *GIGANTEA*-like gene (*gi*), and two genes similar to *CRYPTOCHROME 1* (*cry2* and *cry3*). Among the auxin-related genes for which associations were

found, was a putative auxin efflux carrier (*aec3*) and two putative auxin-responsive (*arg1* and *aux1*) genes that explained 1.1%, 1.6%, and 1.9%, respectively, of the variation in budset timing, and an auxin-associated gene (*drm2*) that explained 0.7% of the variation in December cold hardiness. Finally, *chlh2*, a gene similar to the recently described abscisic acid (ABA) receptor Mg Chelatase H, was found to be associated with December cold hardiness.

4.3.6 Clinal variation in genotype frequencies

Divergent selection due to variation in climate across a species range is expected to enhance levels of population differentiation at sites that are the targets of selection. We have previously identified SNP loci with F_{ST} levels suggestive of the actions of divergent selection (Chapter 3, this thesis). In addition to identifying associations between genotype and phenotype while removing effects related to population of origin, as we have done above, the north-south cline in budset and cold hardiness and associated climatic variables presents an additional means by which to assess the effects of genotype on these traits. For SNP loci with additive effects on phenotypes (or linked to loci with such effects), we would expect the homozygote for the nucleotide with a positive effect on cold hardiness or early budset to present at higher frequencies in the north of the range, the heterozygous state to be more frequent in the centre of the range, and the alternative homozygote to be more frequent in the south. Such a pattern is not readily generated by genetic drift, which is the primary concern in controlling for population structure in association studies. Strong clinal variation ($R^2 > 0.5$) was found for nine of the SNPs associated with either budset timing or cold hardiness, or both (Figure 4.6). Among these were two of the four SNPs in *xth1* (positions 199 and 350), as well as the SNP in *per6*. As was noted above, SNPs within these genes had the largest effect sizes of all the SNPs we tested. The candidate light-signaling gene *co* also exhibited clinal variation. Strong genotypic effects associated with each of the above SNPs were also observed. For example, alternative homozygotes for the A/G mutation in *per6* had average budset dates and index of injury scores on December 1st that varied by ~75 days and 30%, respectively (Figure 4.7). Similar effects were also observed for the A/C SNP found at position 350 in *xth1*. By contrast, many SNPs that had significant associations with assayed phenotypes did not vary across the range in a clinal fashion, suggesting the homogenizing effects of gene flow outweighed the adaptive advantage of these loci.

4.4 Discussion

Association genetics is now widely recognized to be a powerful tool for complex trait dissection, whether that trait is a human disease state or an economically or ecologically important character of a plant species. The association approach has been successfully applied in a few cases to

forest trees, whereby the targeting of a small number of candidate genes has led to detection of allelic variants contributing to trait variation for budset (Ingvarsson et al. 2008), wood properties (Gonzalez-Martinez et al. 2007), and drought tolerance (Gonzalez-Martinez et al. 2008). Several larger association studies are ongoing in conifers. For example, the ADEPT2 project has re-sequenced ~8000 candidate genes in loblolly pine in an effort to elucidate the genomic architecture of economically important traits such as wood properties, and adaptive traits such as drought tolerance (D. Neale, pers. comm.). Here we report one of the largest association studies thus far in a non-model plant. After correcting for the effects of population structure and relative kinship in our rangewide collection of open-pollinated Sitka spruce genotypes, numerous significant associations were detected, which individually explained between approximately one and five percent of the variance in budset and cold hardiness in our mapping population, but cumulatively explained 34.4% and 28.1% of the variation in these traits, respectively. These relatively small effect sizes for individual loci are in line with other association studies in forest trees (Gonzalez-Martinez et al. 2007; Gonzalez-Martinez et al. 2008; Ingvarsson et al. 2008), as well as other taxa to which association genetics has been applied (Aranzana et al. 2005; Altshuler et al. 2008), and generally support the infinitesimal model underlying classical quantitative genetic analysis of such traits.

4.4.1 Patterns of gene flow in Sitka spruce

Using 98 SNPs assayed in 410 individuals collected from across the species range, in conjunction with the genetic clustering program STRUCTURE, we found that three genetic populations provides the best solution to population subdivision in Sitka spruce. Although it was predictable that the Kodiak Island population would be genetically distinct given previous work on gene flow in Sitka spruce (Mimura and Aitken 2007a) (Chapter 3, this thesis), the delineation between Prince Rupert and Icy Bay are more clear than we might have expected given the continuous nature of the range in this area (Figure 4.3). This result suggests a step cline for Sitka spruce, with three possible reasons for its establishment. The most obvious explanation is a geographical barrier. This is clearly the case for the Iniskin – Kodiak Island split, with ~50 km of ocean separating Kodiak Island from the continuous portion of the range. This demarcation line is exacerbated by founder effects leading to local inbreeding on Kodiak (Gapare and Aitken 2005; Gapare et al. 2005; Mimura and Aitken 2007b). Such a barrier does not exist between Prince Rupert and Icy Bay, however, and although these two populations are separated by ~900km, this distance does not appear to be sufficient to explain their genetic differentiation. For example, the distance between Prince Rupert and Redwood is even greater (~1600km), but these populations do not appear differentiated. Two interesting explanations are therefore possible for the breakpoint at Prince Rupert. The first is that a

glacial refugium was present in southeast Alaska, perhaps on one of the offshore islands. Expansion from both northern and southern refugia could explain the genetic differences between mainland Alaska populations of Sitka spruce and those comprising the southern portion of the range. Another possibility is that the populations in SE Alaska have introgressed with white spruce, as hybridization is rampant in this area as one moves away from the coast. Although the populations we sampled were very close to the coast and are therefore thought to be ‘pure’ Sitka, F2 or F3 hybrids may nevertheless be contributing white spruce alleles to these populations. Interestingly, although the ‘blue’ mainland Alaska cluster is largely absent from the more southern populations, it is higher in the Vancouver population, which is somewhat surprising since this population is continuous with surrounding ones. One possible reason for this, which would be consistent with the introgression hypothesis, is that gene flow from white spruce down the Fraser River valley from the interior of British Columbia has provided white spruce alleles to the Vancouver population as well. Indeed, one of the few populations of Sitka spruce that is resistant to the shoot tip weevil (*Pissodes strobi*) is found near Vancouver (King et al. 2004), and resistance to this pest is a characteristic of white spruce. This explanation, however, is not consistent with the fact that hybridization also occurs along two major river drainages on the north coast of British Columbia, near Prince Rupert. In validating our association results in a separate mapping population, we will also genotype a number of white spruce individuals. Doing so will help to disentangle these two competing hypotheses to explain the step-cline observed here.

4.4.2 SNP associations and their possible functional consequences

The putative functions of many of the associations uncovered in this study provides tantalizing clues to the genetic underpinnings of adaptive traits in forest trees. Although reasonable hypotheses can be generated about genes that may be involved in local adaptation in forest trees based on functional studies in model systems, casting the net wide in candidate gene-based association mapping provides the greatest chance of successfully identifying genetic variants associated with phenotypic trait variation. We therefore chose a wide variety of genes for the current study, and the results demonstrate both successful *a priori* identification of genes that are putative homologs to genes in other systems that are known to be involved in climate-related traits, and more unexpected results that reinforce the need for even larger panels of candidates. The former is demonstrated starkly by the association of numerous genes putatively involved in light perception and signaling. These include a homolog to *PHYTOCHROME A* (*phyA*), the most upstream element in the pathway leading to seasonal growth cessation and budset (Howe et al. 1996; Olsen et al. 1997), as well as downstream genes similar to *CO* (*co*) and *GI* (*gi*). Circadian expression of *CO* was recently shown to not only be crucial to growth cessation, but also to explain much of the variation in timing

of growth cessation along a cline in European aspen (*Populus tremula*) (Bohlenius et al. 2006), and *GI*, though not involved in this pathway as it relates to dormancy, is a key component of the photoperiodic flowering pathway that also includes *CO* and *FLOWERING LOCUS T (FT)*. Two cryptochrome-like genes were also associated with either budset or cold hardiness. Cryptochromes are blue light receptors that regulate the transition to flowering in annual plants (Li and Yang 2007). Although a role for cryptochromes in bud dormancy has not been established, given the apparent overlap between flowering and dormancy-related light signaling pathways, and the associations we have shown here, the role of cryptochromes in bud dormancy and cold hardiness in trees may be underappreciated.

In contrast to the well-characterized role of the above genes in plant dormancy, the gene harboring the largest number of SNPs exhibiting phenotypic associations, and with the largest PVE for those associations (*xth1*), was an unexpected find. We tested four SNPs in a putative xyloglucan:xyloglucosyl transferase (also known as xyloglucan endo-transglycosylase (XET)), and each of these variants was associated with both budset timing and cold hardiness. Two of these SNPs (positions 289 and 350) had very low Q-values and among the highest PVE of all SNPs tested (4.3% and 5.4% for budset and cold hardiness, respectively). As these were all synonymous substitutions, it may be that the true functional variant in this gene is in linkage disequilibrium with the SNPs we detected, particularly considering the fact that all four SNPs we assayed in this gene are associated with both traits. However, it is possible that selection on translational efficiency due to codon bias has been a target of selection in this gene. It is difficult to hypothesize about the possible relationship between this enzyme and the traits we assayed, though the literature points to at least one possibility. Xyloglucan is a cell wall polysaccharide that provides strength and can be covalently linked by XETs to other polysaccharides, including cellulose (Hrmova et al. 2007). Such linkages may increase the flexibility of the cell wall, which could be advantageous during cold acclimation as the protoplast dehydrates, putting stress on connections between the cell wall and the plasma membrane. This hypothesis is supported by another highly significant association between an expansin-like gene (*expl2*) and cold hardiness. Expansins are best known for their role in cell wall loosening during growth, but have been shown to be expressed in response to abiotic stress, most notably drought (Dongsu Choi 2006). Indeed, expansin activity is correlated with drought-induced cellular dehydration and an associated increase in cell wall folding (Jones and McQueen-Mason 2004). As freezing temperatures lead to cellular dehydration, it is likely that the cell wall must adjust in much the same way as in times of drought, and the associations we have shown between a putative XET and expansin may reflect this. This hypothesis is also supported by the upregulation of *xth1* during autumn cold acclimation in Sitka spruce (Holliday et al. 2008).

4.4.3 Do selective neutrality tests correlate with associations between genotype and phenotype?

Using various tests against the neutrality of mutations, we have previously shown that several of our candidate genes appear to have been the targets of positive selection (Chapter 3, this thesis). As neutrality tests and association mapping both evaluate the hypothesis of neutral evolution of a gene, we would expect the results of the two methods to align. This will not be true in all cases, however, as neutrality tests do not provide any information about the phenotypic trait that has been the target of selection, and subsequent association mapping may fail to assay the appropriate trait. Nevertheless, it is interesting to compare these results to determine the extent to which genes harboring SNPs associated with a phenotype of interest can also be detected as evolving non-neutrally through neutrality tests that rely on the site frequency spectrum, the relative frequency of synonymous and non-synonymous substitutions, the ratio of diversity to divergence, or the level of variation within and between geographic populations.

Although re-sequencing data for many of the genes associated with either budset or cold hardiness suggested neutral evolution, evidence for positive selection was observed for a few. These include *xth1*, which in addition to being strongly associated with both cold hardiness and budset, exhibited enhanced population differentiation as detected using the program BayesFst (Beaumont and Balding 2004). Interestingly, two SNPs in this gene also had significant clinal variation in the current study (Figure 4.6). Two genes putatively involved in oxidative stress, *ifr2* and *per3*, also harbored outlier SNPs detected using this method. Indeed, the latter had 13 outlier SNPs, though this high number of outliers in a single gene is likely due to the effects of linkage disequilibrium surrounding a single selected SNP. Other genes with both phenotypic associations and significant neutrality tests included *swap*, a putative RNA splicing factor, which had a significantly negative value for Fay and Wu's H suggestive of a selective sweep, and *pip5k*, a gene putatively involved in phosphoinositide signaling, for which the Hudson-Kreitman-Aguade test suggested a selective sweep.

4.4.4 Clinal variation in allele frequencies in the presence of high gene flow

Population geneticists have long sought to resolve an apparent paradox that exists for outcrossing plant species that inhabit diverse habitats connected by gene flow. Common garden experiments in such species frequently demonstrate among-population variation in quantitative traits related to local adaptation, but molecular analyses often reveal low or non-existent genetic differentiation. Under these circumstances, it is assumed that while gene flow tends to homogenize populations at the neutral loci (e.g., isozymes, AFLPs, microsatellites), geographically-restricted selection on loci important to local adaptation will be sufficient to maintain population differentiation

in the genomic vicinity of these loci. In this study we have identified nine SNPs associated with climate-related complex trait variation that also exhibited clinal variation in allele frequency along the north-south range of Sitka spruce, suggesting that selection is sufficiently strong at these loci so as to resist the homogenizing effects of gene flow. By contrast, gene flow appears to have prevented or disrupted differentiation in allele frequencies at many other SNPs that were associated with adaptive trait variation. It should be noted that while clinal variation in allele frequency provides strong evidence for the adaptive importance of a particular mutation, epistasis likely also plays an important role in local adaptation (Phillips 2008), and particular combinations of alleles may have as much or more to do with divergence in quantitative traits as the component individual allelic effects.

4.4.5 Implications for marker-aided selection and conservation genetics

Phenotypic selection in conifers has led to substantial genetic gains in growth and yield since its relatively recent inception, and one of the anticipated outcomes of association mapping is that functional genetic markers associated with trait variation may augment the breeding process. Advantages of marker-aided selection (MAS) include the introduction of novel alleles from natural populations not currently present in breeding programs, as well as the ability to genotype seedlings for informative markers rather than waiting for phenotypes that may not be possible to assay for years (e.g., in the case of some wood phenotypes, insect pest resistance). Evidence from quantitative trait loci (QTL) studies over the past two decades point to numerous mutations of small effect size, which agrees with results from association studies presented here and elsewhere. These small effect sizes are sometimes interpreted as an argument against MAS, since implementation of such a program would require genotyping many markers to determine which crosses to make. However, the decreasing cost and relative technical ease of modern high-throughput genotyping makes these problems negligible. It is also well to remember that a genetic gain of even 10%, obtained through MAS, would be substantial, and such a gain may in principle be achieved with only a handful of markers already available, assuming the estimated effect sizes of these markers are accurate.

In addition to the possibility of using the markers presented here for the purposes of MAS, these markers represent a valuable tool previously unavailable to conservation geneticists. Although it is typically assumed that assaying a handful of neutral markers reasonably approximates the genetic diversity, and hence the trait diversity, present in a population or species, meta-analyses suggest that this may not be the case (Reed and Frankham 2001). As such, determining the appropriate reserve size to capture most of the standing genetic variation, on the basis of neutral markers, is unlikely to successfully capture the desired proportion of complex trait variation segregating within a local population. For traits related to climatic adaptation, this problem becomes acute as species climate

envelopes are expected to shift dramatically in the coming century (Hamann and Wang 2006; Wang et al. 2006), and standing adaptive variation provides one means by which populations can track their climate (Aitken et al. 2008). Therefore, the markers we have described here, augmented by those found in future studies, could be a key tool in assaying the potential of our forests to adapt to rapidly changing climatic conditions. Future studies of the transferability of these markers among species may also extend their usefulness beyond Sitka spruce.

4.5 Conclusions

This study represents a significant step toward the goal of a comprehensive description of the genomic basis of local adaptation in widely distributed plant species in general, and conifer trees in particular. The identification of a suite of markers that collectively explain approximately one third of the phenotypic variation in budset and cold hardiness in Sitka spruce provides an important resource for marker-aided selection, the conservation of genetic resources present in wild populations of spruce species, and prediction of ability to adapt to new climates. Before applying these findings, it will be important to validate the associations in an independent, panmictic mapping population, and this work is ongoing.

Table 4.1. Origins of study populations, sample sizes, and associated climatic variables.

Population (two letter code)	State/Prov.	n	Geographic and climatic variables						
			Lat./Long.	Dist.	MAT	MWMT	MCMT	DD	Precip.
Fort Bragg (FB)	CA	3	39°N-123°W	0	11.8	14.7	9.1	2479	1041
Redwood (RW)	CA	38	42°N-124°W	292	11.6	14.8	8.8	2395	968
Columbia River (CR)	OR	13	47°N-124°W	741	10.6	16.0	5.8	2019	1705
Vancouver (VA)	BC	39	49°N-123°W	1099	10.0	17.1	3.6	2027	1277
Vancouver Island (VI)	BC	46	49.5°N-125°W	1240	9.7	17.6	3.0	1960	1179
Ocean Falls (OF)	BC	48	52.5°N-128°W	1585	8.0	16.8	-1.0	1652	1702
Queen Charlotte Islands (QC)	BC	37	53°N-132°W	1793	8.3	15.0	3.2	1462	1398
Prince Rupert (PR)	BC	33	53.5°N-130°W	1914	7.1	13.5	1.3	1226	2594
Icy Bay (IB)	AK	15	60.5°N-141°W	2812	4.2	12.0	-3.4	721	4074
Valdez (VL)	AK	38	62°N-146°W	3129	3.5	12.9	-5.6	815	1712
Montague (MI)	AK	12	61°N-147°W	3157	3.9	12.5	-4.1	786	2445
Rocky Bay (RB)	AK	35	58°N-151°W	3378	4.1	12.7	-2.2	735	1706
Iniskin (IN)	AK	15	60°N-153°W	3512	1.0	11.7	-10.5	765	1956
Kodiak Island (KD)	AK	38	57°N-153°W	3575	4.7	12.8	-1.3	769	1914

CA, California, USA; OR, Oregon, USA; BC, British Columbia, Canada; AK, Alaska, USA; MAT, mean annual temperature (°C); Dist., Distance of origin of study population from southernmost point in species range (Fort Bragg, CA, USA); MWMT, mean warmest month temperature (°C); MCMT, mean coldest month temperature (°C); DD, degree days (>5°C); Prec, annual precipitation (mm).

Table 4.2. Associations between SNPs and two phenotypic traits: budset and December 1st cold hardiness. R² value only given for statistically significant effects (Q<0.1)

Locus	Position	Type	BLASTX vs. <i>Arabidopsis</i>	Marker effect for budset				Marker effect for cold hardness			
				F	P	Q	R ²	F	P	Q	R ²
<i>abc3</i>	318	S	ABC transporter	6.11	0.002	0.047	0.013	8.60	2.E-04	0.006	0.017
<i>aec3</i>	475	NS	auxin efflux carrier	5.44	0.005	0.066	0.011	--	--	--	--
<i>ago</i>	180	S	argonaute	5.24	0.006	0.066	0.011	--	--	--	--
<i>arg1</i>	646	NS	auxin/aluminum-responsive	7.93	4.E-04	0.014	0.016	--	--	--	--
<i>aux1</i>	567	S	auxin-responsive	9.77	7.E-05	0.004	0.019	6.22	0.002	0.032	0.012
<i>cb13</i>	828	S	calcineurin B-like	--	--	--	--	5.65	0.004	0.053	0.011
<i>chlH2</i>	504	S	Mg chelatase H	--	--	--	--	7.56	0.001	0.013	0.015
<i>co</i>	206	S	constans-like	5.25	0.006	0.066	0.011	--	--	--	--
<i>cry2</i>	188	NS	cryptochrome	5.27	0.006	0.066	0.011	--	--	--	--
	264	S		--	--	--	--	7.26	0.001	0.014	0.015
<i>cry3</i>	123	NS	cryptochrome	--	--	--	--	9.13	1.E-04	0.004	0.018
	207	S		--	--	--	--	9.18	1.E-04	0.004	0.018
	579	S		5.28	0.005	0.066	0.011	--	--	--	--
<i>drm2</i>	438	S	dormancy/auxin associated	--	--	--	--	7.28	0.007	0.091	0.007
<i>erf1</i>	455	NS	ethylene-responsive element-binding factor	5.82	0.003	0.058	0.012	7.37	0.001	0.014	0.015
<i>expl2</i>	199	NS	expansin	--	--	--	--	23.47	2.E-06	1.E-04	0.023
<i>gi</i>	374	NS	gigantea	4.89	0.008	0.083	0.010	--	--	--	--
<i>gp4</i>	98	NS	glutathione peroxidase	--	--	--	--	8.21	3.E-04	0.008	0.016
<i>gp5</i>	98	NS	glutathione peroxidase	--	--	--	--	13.58	2.E-06	1.E-04	0.026
<i>gst4</i>	395	NS	glutathione S-transferase	4.86	0.008	0.083	0.010	--	--	--	--
<i>hta3</i>	523	S	histone h2A	4.74	0.009	0.091	0.010	--	--	--	--
	585	S		6.61	0.002	0.032	0.013	7.23	0.001	0.014	0.014
<i>ifr2</i>	260	NS	isoflavone reductase	12.62	4.E-04	0.014	0.013	10.06	0.002	0.025	0.010
<i>lrr3</i>	424	S	leucine-rich repeat	11.10	2.E-05	0.001	0.022	--	--	--	--
<i>pal3</i>	363	S	phenylalanine ammonia lyase	5.63	0.004	0.062	0.011	--	--	--	--
<i>per3</i>	570	NS	peroxidase	5.73	0.004	0.059	0.011	--	--	--	--
<i>per6</i>	573	S	peroxidase	27.83	5.E-12	1.E-09	0.051	13.29	3.E-06	1.E-04	0.026
<i>phya</i>	441	S	phytochrome A	5.17	0.006	0.067	0.010	--	--	--	--
<i>pip5k</i>	319	S	phosphatidylinositol kinase	7.35	0.001	0.021	0.015	--	--	--	--
<i>skip2</i>	182	NS	F-box family	6.81	0.001	0.028	0.014	--	--	--	--
<i>swap</i>	441	S	SWAP domain-containing	5.41	0.005	0.066	0.011	--	--	--	--
<i>xth1</i>	39	S	xyloglucan:xyloglucosyl transferase	6.88	0.001	0.028	0.014	12.67	5.E-06	2.E-04	0.025
	199	S		8.36	3.E-04	0.012	0.016	5.39	0.005	0.064	0.011
	289	S		16.13	2.E-07	2.E-05	0.031	22.87	4.E-10	5.E-08	0.042
	350	S		23.06	3.E-10	4.E-08	0.043	30.01	7.E-13	2.E-10	0.054

Figure 4.1. Distribution of Sitka spruce and origins of study populations for this study. Population names corresponding to two-letter codes can be found in Table 4.1.

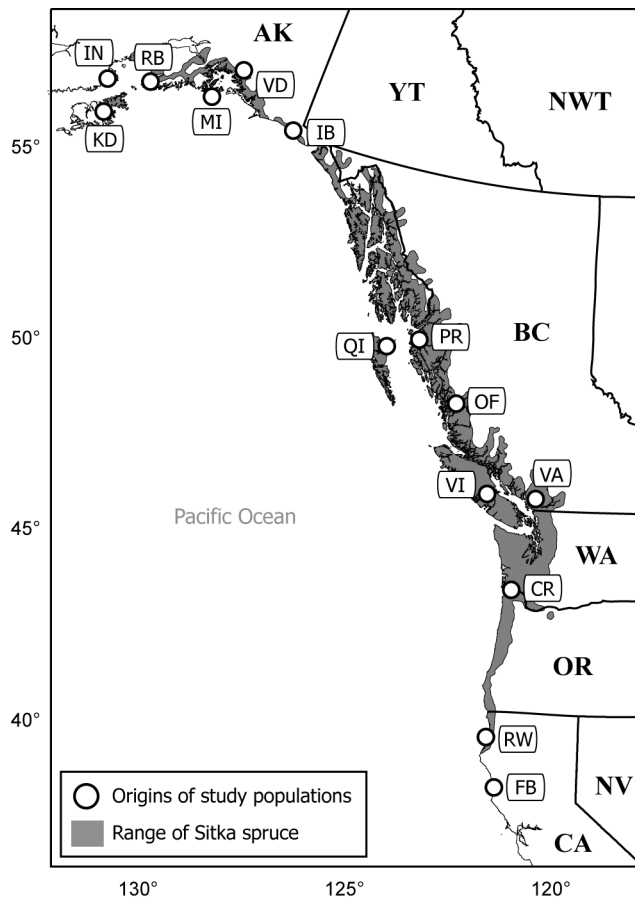


Figure 4.2. Regressions of mean budset and cold hardiness (measured on two dates) on distance from the southern limit of the species range at Fort Bragg, CA, USA. Budset was monitored weekly starting July 1, 2008. Cold hardiness was measured using the electrolyte leakage assay described in methods section.

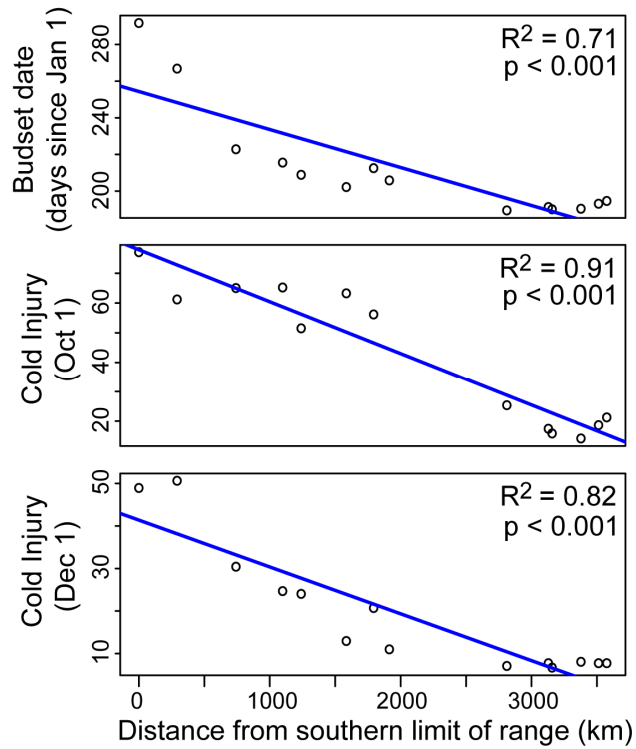


Figure 4.3. Genetic clustering of 410 individuals from across the species range of Sitka spruce using the program STRUCTURE. Vertical represent individuals and colours represent proportional membership of each individual in each of three genetic clusters.

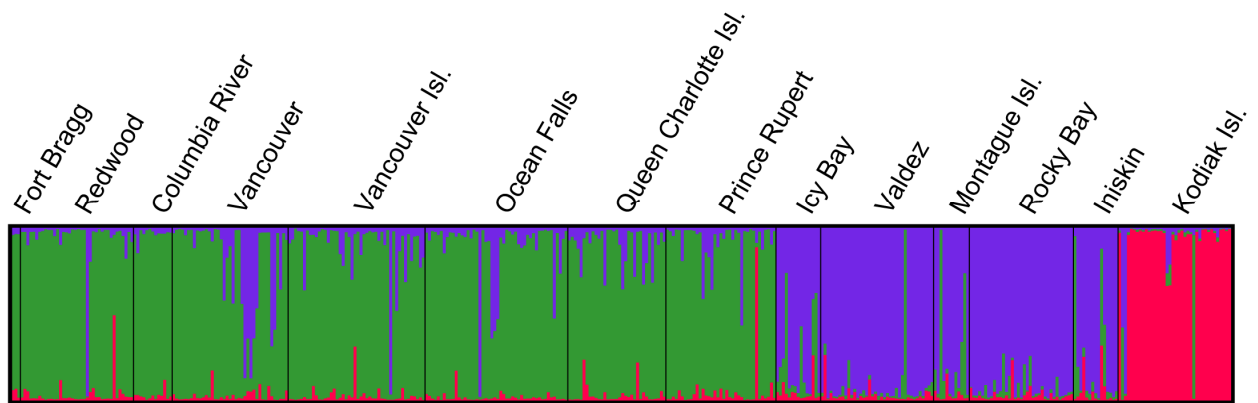


Figure 4.4. Distribution of pairwise kinship estimates for 410 Sitka spruce genotypes represented in this study.

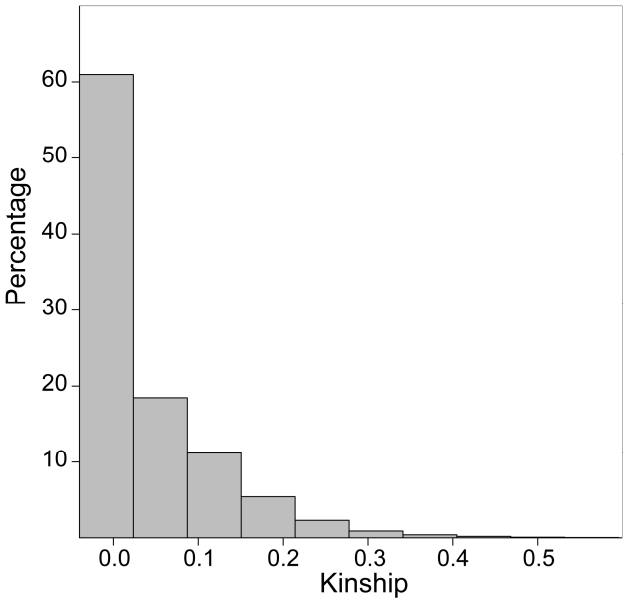


Figure 4.5. Venn diagram of significant associations ($Q < 0.1$) for locally adaptive traits.

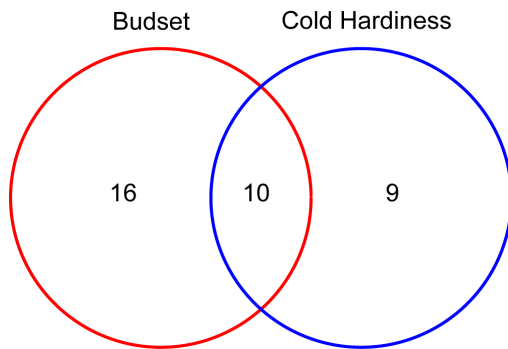


Figure 4.6. Clinal variation in nine SNPs that were associated with either budset, cold hardiness or both (see Table 4.2 for associations). Best fit lines and associated R^2 values given for regression of allele frequency on distance from southern limit of the species range at Fort Bragg, CA (associated P-values all < 0.001). SNP position given for genes with multiple associated SNPs.

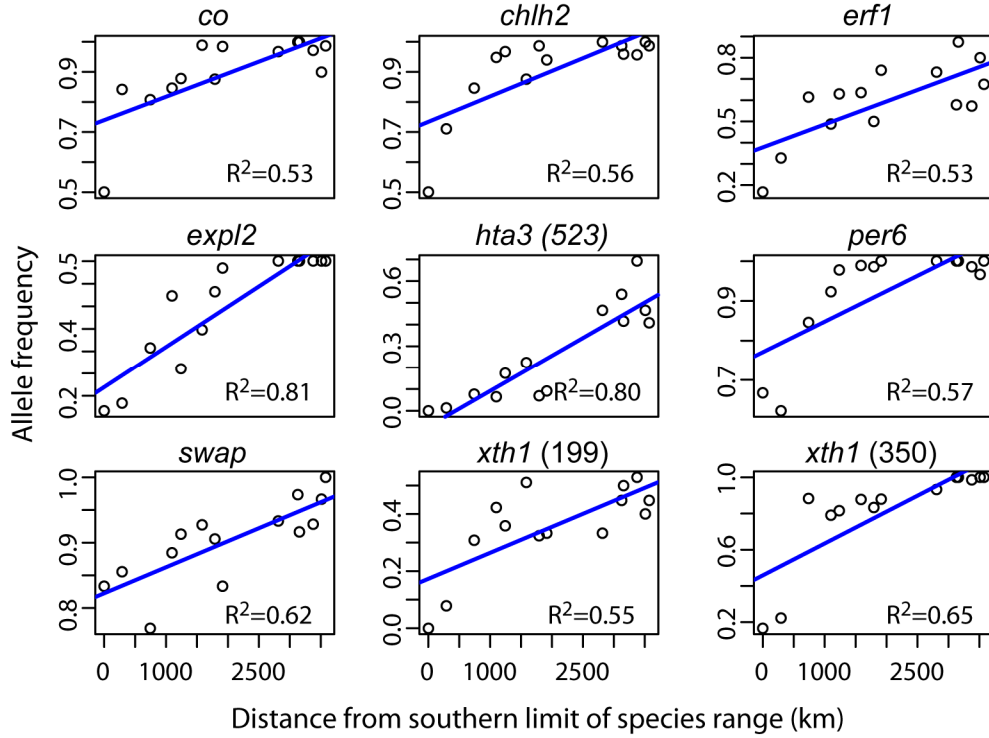
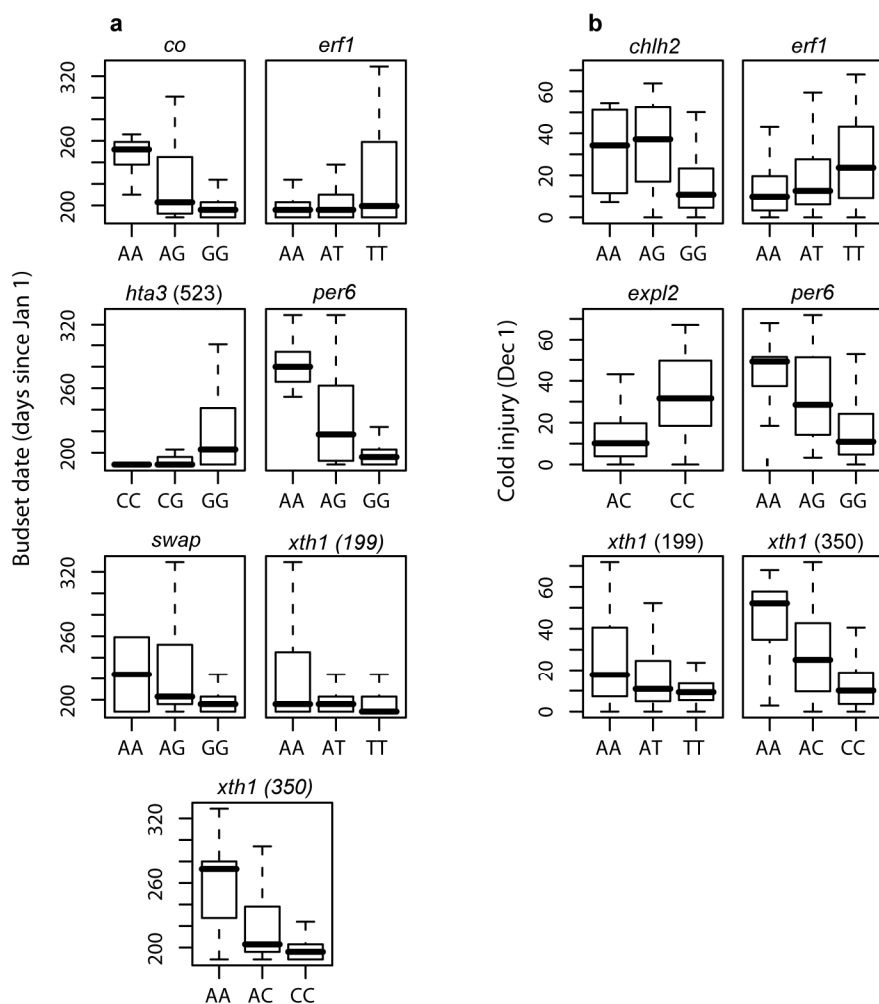


Figure 4.7. Genotypic effects of SNPs associated with either budset (a) or December cold hardiness (b) and that exhibited clinal variation in allele frequency. Boxplots only given for traits that were associated with a given SNP. Position given for genes with multiple associated SNPs.



4.6 Literature cited

- Aitken, S. N., S. Yeaman, J. A. Holliday, T. Wang, and S. Curtis-McLane. 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications* **1**:95-111.
- Altshuler, D., M. J. Daly, and E. S. Lander. 2008. Genetic Mapping in Human Disease. *Science* **322**:881-888.
- Aranzana, M. J., S. Kim, K. Y. Zhao et al. 2005. Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *PLOS Genetics* **1**:531-539.
- Beaumont, M. A., and D. J. Balding. 2004. Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13**:969-980.
- Bohlenius, H., T. Huang, L. Charbonnel-Campaa, A. M. Brunner, S. Jansson, S. H. Strauss, and O. Nilsson. 2006. CO/FT regulatory module controls timing of flowering and seasonal growth cessation in trees. *Science* **312**:1040-1043.
- Dongsu Choi, H.-T. C. Y. L. 2006. Expansins: expanding importance in plant growth and development. *Physiologia Plantarum* **126**:511-518.
- Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**:2611-2620.
- Fan, J. B., A. Oliphant, R. Shen et al. 2003. Highly parallel SNP genotyping. Cold Spring Harbor Symposia on Quantitative Biology **68**:69-78.
- Gapare, W. J., and S. N. Aitken. 2005. Strong spatial genetic structure in peripheral but not core populations of Sitka spruce *Picea sitchensis* (Bong.) Carr. *Molecular Ecology* **14**:2659-2667.
- Gapare, W. J., S. N. Aitken, and C. E. Ritland. 2005. Genetic diversity of core and peripheral Sitka spruce (*Picea sitchensis* (Bong.) Carr) populations: implications for conservation of widespread species. *Biological Conservation* **123**:113-123.
- Gonzalez-Martinez, S. C., D. Huber, E. Ersoz, J. M. Davis, and D. B. Neale. 2008. Association genetics in *Pinus taeda* L. II. Carbon isotope discrimination. *Heredity* **101**:19-26.
- Gonzalez-Martinez, S. C., N. C. Wheeler, E. Ersoz, C. D. Nelson, and D. B. Neale. 2007. Association genetics in *Pinus taeda* L. I. Wood property traits. *Genetics* **175**:399-409.
- Hamann, A., and T. L. Wang. 2006. Potential effects of climate change on ecosystem and tree species distribution in British Columbia. *Ecology* **87**:2773-2786.
- Hannerz, M., S. N. Aitken, J. N. King, and S. Budge. 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.
- Hardy, O. J., and X. Vekemans. 2002. SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2**:618-620.
- Holliday, J. A., S. G. Ralph, R. White, J. Bohlmann, and S. N. Aitken. 2008. Global monitoring of autumn gene expression within and among phenotypically divergent populations of Sitka spruce (*Picea sitchensis*). *New Phytologist* **178**:103-122.
- Howe, G. T., G. Gardner, W. P. Hackett, and G. R. Furnier. 1996. Phytochrome control of short-day-induced bud set in black cottonwood. *Physiologia Plantarum* **97**:95-103.
- Hrmova, M., V. Farkas, J. Lahnstein, and G. B. Fincher. 2007. A Barley Xyloglucan Xyloglucosyl Transferase Covalently Links Xyloglucan, Cellulosic Substrates, and (1,3;1,4)-beta-D-Glucans. *J. Biol. Chem.* **282**:12951-12962.
- Ingvarsson, P. K., M. V. Garcia, V. Luquez, D. Hall, and S. Jansson. 2008. Nucleotide polymorphism and phenotypic associations within and around the phytochrome B2 locus in European aspen (*Populus tremula*, Salicaceae). *Genetics* **178**:2217-2226.
- Jermstad, K. D., D. L. Bassoni, K. S. Jech, N. C. Wheeler, and D. B. Neale. 2001a. Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. I. Timing of vegetative bud flush. *Theoretical and Applied Genetics* **102**:1142-1151.

- Jermstad, K. D., D. L. Bassoni, N. C. Wheeler, T. S. Anekonda, S. N. Aitken, W. T. Adams, and D. B. Neale. 2001b. Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. II. Spring and fall cold-hardiness. *Theoretical and Applied Genetics* **102**:1152-1158.
- Jones, L., and S. McQueen-Mason. 2004. A role for expansins in dehydration and rehydration of the resurrection plant *Craterostigma plantagineum*. *FEBS Letters* **559**:61-65.
- King, J. N., R. I. Alfaro, and C. Cartwright. 2004. Genetic resistance of Sitka spruce (*Picea sitchensis*) populations to the white pine weevil (*Pissodes strobi*): distribution of resistance. *Forestry* **77**:269-278.
- Lander, E. S., and N. J. Schork. 1994. Genetic dissection of complex traits. *Science* **265**:2037-2048.
- Li, Q. H., and H. Q. Yang. 2007. Cryptochrome signaling in plants. *Photochemistry and Photobiology* **83**:94-101.
- Lynch, M., and K. Ritland. 1999. Estimation of pairwise relatedness with molecular markers. *Genetics* **152**:1753-1766.
- Mimura, M., and S. N. Aitken. 2007a. Adaptive gradients and isolation-by-distance with postglacial migration in *Picea sitchensis*. *Heredity* **99**:224-232.
- Mimura, M., and S. N. Aitken. 2007b. Increased selfing and decreased effective pollen donor number in peripheral relative to central populations in *Picea sitchensis* (Pinaceae). *American Journal of Botany* **94**:991-998.
- Namroud, M. C., J. Beaulieu, N. Juge, J. Laroche, and J. Bousquet. 2008. Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Molecular Ecology* **17**:3599-3613.
- Neale, D. B., and O. Savolainen. 2004. Association genetics of complex traits in conifers. *Trends in Plant Science* **9**:325-330.
- Olsen, J. E., O. Junttila, J. Nilsen, M. E. Eriksson, I. Martinussen, O. Olsson, G. Sandberg, and T. Moritz. 1997. Ectopic expression of oat phytochrome A in hybrid aspen changes critical daylength for growth and prevents cold acclimatization. *Plant Journal* **12**:1339-1350.
- Phillips, P. C. 2008. Epistasis - the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics* **9**:855-867.
- Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**:945-959.
- Ralph, S. G., H. J. E. Chun, D. Cooper et al. 2008a. Analysis of 4,664 high-quality sequence-finished poplar full-length cDNA clones and their utility for the discovery of genes responding to insect feeding. *BMC Genomics* **9**.
- Ralph, S. G., H. J. E. Chun, N. Kolosova et al. 2008b. A conifer genomics resource of 200,000 spruce (*Picea* spp.) ESTs and 6,464 high-quality, sequence-finished full-length cDNAs for Sitka spruce (*Picea sitchensis*). *BMC Genomics* **9**.
- Ralph, S. G., H. Yueh, M. Friedmann et al. 2006. Conifer defence against insects: microarray gene expression profiling of Sitka spruce (*Picea sitchensis*) induced by mechanical wounding or feeding by spruce budworms (*Choristoneura occidentalis*) or white pine weevils (*Pissodes strobi*) reveals large-scale changes of the host transcriptome. *Plant Cell and Environment* **29**:1545-1570.
- Reed, D. H., and R. Frankham. 2001. How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution* **55**:1095-1103.
- Shen, R., J. B. Fan, D. Campbell et al. 2005. High-throughput SNP genotyping on universal bead arrays. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis* **573**:70-82.
- Storey, J. D., and R. Tibshirani. 2003. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America* **100**:9440-9445.
- Tuskan, G. A.S. DiFazioS. Jansson et al. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**:1596-1604.

- Wang, T., A. Hamann, D. L. Spittlehouse, and S. N. Aitken. 2006. Development of scale-free climate data for western Canada for use in resource management. *International Journal of Climatology* **26**:383-397.
- Weiser, C. J. 1970. Cold resistance and injury in woody plants. *Science* **169**:1269-1278.
- Wheeler, N. C., K. D. Jermstad, K. Krutovsky, S. N. Aitken, G. T. Howe, J. Krakowski, and D. B. Neale. 2005. Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. IV. Cold-hardiness QTL verification and candidate gene mapping. *Mol. Breed.* **15**:145-156.
- Yu, J. M., G. Pressoir, W. H. Briggs et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* **38**:203-208.

5. Thesis Conclusions

5.1 Introduction

The spruce genus harbors some of the most economically important conifers in the world. These include species that are also ecologically dominant in many areas, including the sub-boreal and boreal regions of Scandinavia and Russia (Norway spruce), the sub-boreal and boreal region of Canada (white spruce), the southern interior of British Columbia (Engelmann, white, and their hybrids), and the west coast of North America (Sitka spruce). In addition, numerous species with more limited ranges that are ecologically important but economically unimportant are distributed across North America and Asia. Forest geneticists have traditionally focused on species in the former category, and on traits relevant to industrial forestry, such as wood volume and quality. Dormancy related traits are often only of tangential interest to the extent that they are negatively correlated with growth and yield. However, anthropogenic climate change necessitates an increased emphasis be placed on traits related to local adaptation, which do not increase the value of a stand directly, but most certainly do so indirectly by facilitating perennial growth habits that are appropriate to the local climatic conditions.

The primary goal of this thesis was to identify a set of ecologically relevant genetic markers for traits related to local adaptation in spruce. To achieve this, I first identified a suite of candidate genes with putative roles in budset, dormancy induction, or cold acclimation by monitoring gene expression during the fall using a 21,840 element cDNA microarray. Genes that were strongly upregulated and that had annotations suggestive of homology to cold hardiness genes in model plants were taken forward to a SNP discovery panel and ultimately to association tests. Although candidate-gene based association mapping is inevitably incomplete, as all variants cannot be tested for association, the large panel of genes tested here provides a substantial first step in the characterization of the genomic determinants of adaptation to local climate in spruce.

5.2 Wholesale remodeling of the transcriptome during the fall

I observed a massive reprogramming of the foliar transcriptome during the period of dormancy induction and cold acclimation in Sitka spruce. Out of 21,840 array elements, 1257 were upregulated and 967 downregulated. This result is in agreement with recently published studies in aspen that monitored similar periods, and found similar proportions of array elements up- and downregulated (Druart et al. 2007; Ruttink et al. 2007). Among the transcripts most strongly upregulated were canonical cold stress tolerance genes such as dehydrins, putative antifreeze genes,

and many related to oxidative stress. Numerous array elements that were upregulated had annotations suggestive of involvement in carbohydrate or lipid biosynthesis, which corroborates the importance of remodeling of both the soluble carbohydrate content and lipid membranes under cold stress. In addition, genes involved in cellular transport in general and metabolite transport in particular were upregulated, which is a feature of the cold hardiness transcriptome that may be underappreciated. Finally, many genes putatively involved in signal transduction and transcriptional regulation were induced, including putative calcium signaling genes, which is one of the primary second messengers in *Arabidopsis* cold acclimation, as well as genes putatively involved in ABA, ethylene, and auxin signaling. The latter result is somewhat surprising. Given the insensitivity of the cambium to exogenous auxin during dormancy, one might expect associated transcripts to be downregulated. This was not the case for the dormant cambial meristem in aspen, and further study is needed in this area.

The functional classes of genes induced aligned quite well with studies of the transition to dormancy in aspen, and in many ways with studies of the *Arabidopsis* cold stress response, suggesting that the cold hardiness transcriptome is reasonably well conserved across widely divergent plant taxa. However, the important difference between the perennial and annual cold hardiness transcriptional programmes is that the former appears to be regulated by nightlength in both angiosperms and gymnosperms, with no exogenous temperature cue required to initiate the wholesale changes I observed. I did not find a coordinated response to freezing temperatures the plants experienced in late November, though studies of aspen did report a second wave of transcription late in the fall related to adaptation to cold, possible related to the classic phase II of cold acclimation. More studies are required in this area under controlled conditions to increase the resolution of the late autumn transcriptome. However, it is possible that phase II of cold acclimation in Sitka spruce does not correspond to a significant transcriptional response, and may involve, for example, changes at the level of protein expression.

5.3 Diversity and selection in local adaptation-related candidate genes

Nucleotide diversity and linkage disequilibrium in the candidate genes I characterized in Sitka spruce were comparable to those published for other widespread gymnosperm and angiosperm tree species. In spite of being predominantly outcrossing with large effective population sizes, nucleotide diversity estimates are more similar to those found in the selfing annual *Arabidopsis* than to, for example, the outcrosser maize (*Zea mays*) (Savolainen and Pyhajarvi 2007). Lower than expected diversity in conifers may be the result of low mutation rates, as per-generation mutation rates, approximately 1.4×10^{-8} for *Pinus*, assuming a generation time of 20yrs (Willyard et al. 2007),

appear to also be closer to that of *Arabidopsis* (1.5×10^{-8}) (Koch et al. 2000) than maize (between 2.9×10^{-8} and 3.3×10^{-8}) (Clark et al. 2005). Indeed, nucleotide diversity in maize appears to be about twice that of *Arabidopsis*, and perhaps closer to three times that of the diversity in conifers, although this varies quite a bit among species, likely because so few genes have thus far been assayed (Savolainen and Pyhajarvi 2007).

I observed very rapid decay of linkage disequilibrium (LD), which is similar to most other conifer species studied thus far, though more rapid than the best-characterized conifer, loblolly pine (Brown et al. 2004). Though modern forestry has lead to a more widespread current distribution for loblolly, the natural range was substantially more restricted than many other gymnosperms studied so far (e.g., Scots pine, Norway spruce, Douglas-fir). The restricted natural range of loblolly is expected to increase LD due to increased local inbreeding, which is indeed the case. Nevertheless, all of these species share sufficiently low LD to enable efficient mapping of associations at within-gene resolutions, and the rapid decay in Sitka spruce in particular suggests that detected associations may more often be the result of having genotyped the actual functional variant. This situation not only provides highly transferable genetic markers for marker-aided selection and conservation genetics, but also facilitates future *in vivo* complementation studies to further characterize the physiological consequences of mutations that are associated with phenotypes.

Although the ultimate goal of this thesis was to conduct association tests, and re-sequencing was largely carried out to advance this goal, re-sequencing data provides opportunities to assess the effects of natural selection on coding sequences in ways that targeted genotyping of individual SNPs does not. Among the original and still among the most widely applied selective neutrality tests are the d_N/d_S ratio, Tajima's D, and the HKA test. The former is not particularly sensitive, as proteins are made up of domains, and genes subject to strong positive selection on a particular domain may appear to be under purifying selection when the gene sequence is taken as a whole. Nevertheless, I did observe one significantly positive deviation suggestive of positive selection to alter protein function. Tajima's D gave several significant results but interpretations were confounded by apparent non-equilibrium conditions in Sitka spruce, with negative values for both the mean and median for D. Modern frequency-spectrum based tests such as Fay and Wu's H that take into account outgroup sequences may more reliably separate true signals of positive selection from the noise of demographic history. Indeed, the median value of Fay and Wu's H was approximately zero, with several strong signals suggesting selective sweeps. The HKA test, which seeks outliers from the proportional relationship between diversity and divergence, is also less susceptible to demographic factors, and I identified several candidate genes with reduced diversity suggestive of positive selection. One of the most powerful approaches to the search for adaptive evolution are outlier detection methods that

compare within and between population diversity. Although I had relatively low power to detect outliers due to small sample sizes in the SNP discovery panel, a number of significant deviations were observed suggestive of the actions of diversifying selection. Notable among these were five genes involved in oxidative stress, one several primary stressors associated with freezing temperatures.

5.4 Significant associations between genotype and phenotype

Although the neutrality tests described above provide tantalizing clues as to the genetic targets of natural selection in Sitka spruce, they lack a connection to ecologically relevant phenotypes and as such no strong inferences can be made as to the reasons for their apparent non-neutral evolution. Association genetics provides a solution to this problem, and is generally more sensitive to the effects of selection as individual nucleotides can be readily assayed and analyzed for effects of genotypes on phenotypic traits of interest. By applying association mapping to three quantitative traits, namely, budset timing and cold hardiness on two dates, measured in 410 individuals, I was able to detect significant effects of 45 SNPs within 28 candidate genes on one or more of these traits. Surprisingly, no significant results were found for early autumn cold hardiness, which may reflect the fact that only the northern populations had begun hardening and in controlling for population of origin I lost power to detect significant effects at this timepoint. The significant associations that were found spanned a diverse array of putative gene function, and often were significant for both budset timing and December cold hardiness, which likely reflects the genetic correlation between these traits (Howe et al. 2003). One of the functional categories of genes that stood out among the significant associations was the putative light signaling genes, which included two cryptochrome-like genes, a phytochrome-like gene, and two genes similar to *CONSTANS* and *GIGANTEA*. These results are quite striking given that nightlength is the most important cue regulating the transition to dormancy (Howe et al. 2003). This corroborates evidence in aspen that *PHYTOCHROME B2* is associated with budset timing and also harbors SNPs with clinal variation across the north-south species range (Ingvarsson et al. 2006; Ingvarsson et al. 2008).

5.5 Limitations of the current study

For all of their potential benefits, genomic studies carried out in non-model systems have some significant drawbacks, chief among these is uncertain orthology of candidate genes with those in model systems. Although some genes are more likely to represent true functional orthologs, particularly those that appear to be single copy or members of compact gene families, care must be taken in making functional inferences that rely on orthology. A related problem is the large number of genes with no significant homology to any gene in the relevant model species. This issue is

pronounced in conifers, wherein 30% of putative unique transcripts return no hits with a bit score greater than 50 when searched against the non-redundant database of GenBank (Ralph et al. 2008). Although this suggests tantalizing avenues for future study, the difficulty involved in studying such genes means that few are likely to be targeted.

Although the number of candidate genes in the current study is substantial, such a survey is inherently incomplete as I have only assayed a small fraction of the likely 40-60 thousand genes in spruce. Nevertheless, the associations uncovered cumulatively explain a proportion of the variation in phenotypic traits that approaches the heritability, which is remarkable. I biased candidate gene selection toward those with gene expression patterns or annotations suggesting relevance to dormancy-related traits, but it is unlikely that this criteria is sufficient to have captured such a substantial fraction of the dormancy-related genetic variation present in the spruce genome. A more likely scenario is that I have somewhat overestimated the percentage of variance explained by each marker, which is common among both QTL and association studies. Known as the ‘Beavis effect’ or ‘winner’s curse’, this results from estimating effect sizes in the same population in which the association was established, and is a particular problem in cases of borderline significance (Xu 2003; Zöllner and Pritchard 2007). The magnitude of the overestimation depends on the power of the initial test for association. Simulation studies have shown that overestimates are greatest when on the order of 100 individuals are assayed, but only slightly overestimated when 500 individuals are assayed, and very close to the true value with 1000 individuals (Xu 2003). An important consequence of the Beavis effect is irreproducibility of associations due to inadequate sample sizes in validation populations, the size of which is often informed by the presumptive power of the initial population as evidenced by the associations detected. Although my sample sizes should be sufficient to limit this effect, I plan to validate the associations presented here in a larger independent mapping population of approximately 1000 individuals.

5.6 Future research

This study represents a significant forward step in our understanding of the genomic basis of local adaptation to climate in conifers, and, perhaps more importantly, provides practical molecular tools to advance marker-aided selection and conservation genetics. A substantial amount of work remains, however. Although difficult to carry out in a conifer, functional characterization of a subset of the candidate genes described here is needed to enhance our understanding of the molecular basis for the dormancy transition in conifers. This goal could be advanced through more nuanced real-time PCR gene expression studies of developmental transitions under controlled conditions, following cloning of full-length sequences for dormancy-related gene families. For particularly interesting

candidate genes, heterologous overexpression or knockdown (RNAi) studies should be conducted in poplar. Care must be taken in interpreting such studies, however, due to the high degree of evolutionary divergence between spruce and poplar. Although relatively recalcitrant to genetic manipulation, improvements in the efficiency of transformation also makes such work feasible in spruce (Pena and Seguin 2001), which has the advantage of maintaining the ecological and evolutionary context of the candidate gene.

An important aspect of the molecular biology of dormancy-related traits that has yet to be studied is the role of protein-protein interactions. Yeast-2-hybrid (Y2H) screens are an attractive approach to dissecting the regulatory networks that govern a particular trait, and are experimentally tractable on a large scale for any species with substantial sequence resources. Such studies would not only provide a functional context for the associations described here, but would also enhance our understanding of why some candidate genes do not contribute to adaptation. For example, although a new SNP variant may arise that is adaptive for one trait, negative consequences for a separate trait (i.e., pleiotropy) for which the gene is also required prevents its spread. In the context of a Y2H screen, we would expect genes with many interacting partners, so-called ‘hubs’, to affect multiple traits and therefore be constrained from fulfilling their ‘adaptive potential’ for any given trait. By contrast, genes with few connections are expected to more commonly be involved in positive or diversifying selection. Theory suggests this effect of pleiotropy is common (Otto 2004), but empirical examples have so far been limited to single genes (e.g., Scarcelli et al. 2007).

High throughput genotyping technologies that are already available (e.g., Gunderson et al. 2006), and sequencing technologies that are under development (e.g., Eid et al. 2009), will undoubtedly make association genetics feasible on a much larger scale. The application of, for example, Illumina’s Infinium assay (Gunderson et al. 2006) is now feasible for non-model species such as conifers (D. Neale, pers. comm.), which holds important advantages over the GoldenGate assay I employed in this thesis. Association studies tend to focus on common alleles as a way of improving statistical power and targeting genotyping efforts. The Infinium assay has the advantage of eliminating the need for parsing the available SNPs (adjacent SNPs can be genotyped), and may uncover previously undiscovered effects of minor alleles. Of course, such effects may only be found with a concomitant increase in the size of mapping populations. To realize the full potential of association genetics in conifers, however, efforts must be undertaken to sequence a full conifer genome. Given their size (~20-30 Gb for *Picea* and *Pinus* spp., or 7-10 times that of the human genome) (Ahuja and Neale 2005), this will require substantial resources. Nevertheless, a genome sequence would enable not only the targeting of any full-length gene sequence of interest, but also allow characterization of upstream regulatory regions. These regulatory regions are known to

contribute to trait variation, as evidenced by expression QTL studies demonstrating associations between segregating gene expression variation and phenotypic trait variation (Gilad et al. 2008).

In addition to increasing the number and types of target loci, the number of target species should also be increased. Closely related species that freely hybridize (e.g., white and Sitka spruce; white and Engelmann spruce) are likely to share polymorphisms, and there is even evidence that more distantly related species pairs share SNPs (Bouille and Bousquet 2005). It is therefore possible that polymorphisms segregating in multiple species contribute to adaptive gradients in each of those species. The associations I have uncovered would clearly be more useful if they were transferable among species, and even knowing whether the same genes, if not the same nucleotides, drive adaptation in related species, would improve the efficiency of enumerating adaptive genetic variation genus-wide. These issues bear directly on the more general question of whether adaptation is the result of new mutations or standing variation. Although there are examples of adaptation from standing genetic variation (Colosimo et al. 2005; Hoekstra et al. 2006), the generality of this phenomenon is currently not known (Orr 2005; Barrett and Schluter 2008). A better understanding of the source of adaptive mutations is crucial to predicting the potential for adaptation of conifer populations to climate change. If adaptation to climate is due in whole or in part to changes in frequencies of alleles already segregating, rather than new mutations, the speed of adaptation would be increased dramatically. New mutations, whether adaptive or not, have a high probability of being eliminated through genetic drift, and take longer to reach high frequency if they are not lost to drift (Barrett and Schluter 2008). The repeated advance and retreat of the ice sheets in North America during the Late Pleistocene, as well as closely related species which form hybrids, provides a basis for investigating this question in spruce. If adaptive alleles spread from standing genetic variation, they should be present in the basal population(s), namely, in glacial refugia. Incorporating closely related species (e.g., white spruce and Sitka spruce) provides the added advantage of corroborating evidence from refugia by determining if the same mutations contribute to adaptation in both species.

The pace of genomics research today is stunning. Driven largely by the human biomedical community, rapid advances in sequencing and genotyping technologies have put goals within reach that were unimaginable only a few years ago. Acquiring a large amount of expressed sequence data, which previously required a coordinated and well-funded approach, can now be achieved for non-model tree species with a modest input of research funds. With these resources, we are now on the verge of a coherent description of the genomic complexity of adaptation of woody perennials to life in a cold climate.

5.7 Literature cited

- Ahuja, M. R., and D. B. Neale. 2005. Evolution of genome size in conifers. *Silvae Genetica* **54**:126-137.
- Barrett, R. D. H., and D. Schluter. 2008. Adaptation from standing genetic variation. *Trends in Ecology & Evolution* **23**:38-44.
- Bouille, M., and J. Bousquet. 2005. Trans-species shared polymorphisms at orthologous nuclear gene loci among distant species in the conifer *Picea* (Pinaceae): Implications for the long-term maintenance of genetic diversity in trees. *American Journal of Botany* **92**:63-73.
- Brown, G. R., G. P. Gill, R. J. Kuntz, C. H. Langley, and D. B. Neale. 2004. Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proceedings of the National Academy of Sciences of the United States of America* **101**:15255-15260.
- Clark, R. M., S. Tavaré, and J. Doebley. 2005. Estimating a nucleotide substitution rate for maize from polymorphism at a major domestication locus. *Molecular Biology and Evolution* **22**:2304-2312.
- Colosimo, P. F., K. E. Hosemann, S. Balabhadra et al. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* **307**:1928-1933.
- Druart, N., A. Johansson, K. Baba et al. 2007. Environmental and hormonal regulation of the activity-dormancy cycle in the cambial meristem involves stage-specific modulation of transcriptional and metabolic networks. *Plant Journal* **50**:557 - 573.
- Eid, J., A. Fehr, J. Gray et al. 2009. Real-Time DNA sequencing from single polymerase molecules. *Science* **323**:133-138.
- Gilad, Y., S. A. Rifkin, and J. K. Pritchard. 2008. Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends in Genetics* **24**:408-415.
- Gunderson, K. L., F. J. Steemers, H. G. Ren et al. 2006. Whole-genome genotyping. Pp. 359-376. *Methods in Enzymology*.
- Hoekstra, H. E., R. J. Hirschmann, R. A. Bunday, P. A. Insel, and J. P. Crossland. 2006. A Single Amino Acid Mutation Contributes to Adaptive Beach Mouse Color Pattern. *Science* **313**:101-104.
- Ingvarsson, P. K., M. V. Garcia, D. Hall, V. Luquez, and S. Jansson. 2006. Clinal variation in phyB2, a candidate gene for day-length-induced growth cessation and bud set, across a latitudinal gradient in European aspen (*Populus tremula*). *Genetics* **172**:1845-1853.
- Ingvarsson, P. K., M. V. Garcia, V. Luquez, D. Hall, and S. Jansson. 2008. Nucleotide polymorphism and phenotypic associations within and around the phytochrome B2 locus in European aspen (*Populus tremula*, Salicaceae). *Genetics* **178**:2217-2226.
- Koch, M. A., B. Haubold, and T. Mitchell-Olds. 2000. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis* and related genera (Brassicaceae). *Molecular Biology and Evolution* **17**:1483-1498.
- Orr, H. A. 2005. The genetic theory of adaptation: A brief history. *Nature Reviews Genetics* **6**:119-127.
- Otto, S. P. 2004. Two steps forward, one step back: the pleiotropic effects of favoured alleles. *Proceedings of the Royal Society of London Series B-Biological Sciences* **271**:705-714.
- Pena, L., and A. Seguin. 2001. Recent advances in the genetic transformation of trees. *Trends in Biotechnology* **19**:500-506.
- Ralph, S. G., H. J. E. Chun, N. Kolosova et al. 2008. A conifer genomics resource of 200,000 spruce (*Picea* spp.) ESTs and 6,464 high-quality, sequence-finished full-length cDNAs for Sitka spruce (*Picea sitchensis*). *BMC Genomics* **9**.
- Ruttink, T., M. Arend, K. Morreel, V. Storme, S. Rombauts, J. Fromm, R. P. Bhalerao, W. Boerjan, and A. Rohde. 2007. A molecular timetable for apical bud formation and dormancy induction in poplar. *Plant Cell* **19**:2370 - 2390.

- Savolainen, O., and T. Pyhajarvi. 2007. Genomic diversity in forest trees. *Current Opinion in Plant Biology* **10**:162-167.
- Scarcelli, N., J. M. Cheverud, B. A. Schaal, and P. X. Kover. 2007. Antagonistic pleiotropic effects reduce the potential adaptive value of the FRIGIDA locus. *Proceedings of the National Academy of Sciences of the United States of America* **104**:16986-16991.
- Willyard, A., J. Syring, D. S. Gernandt, A. Liston, and R. Cronn. 2007. Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for pinus (vol 24, pg 90, 2007). *Molecular Biology and Evolution* **24**:620-620.
- Xu, S. 2003. Theoretical basis of the Beavis effect. *Genetics* **165**:2259-2268.
- Zöllner, S., and J. K. Pritchard. 2007. Overcoming the winner's curse: estimating penetrance parameters from case-control data. *The American Journal of Human Genetics* **80**:605-615.