**Genomic and phenotypic architecture of a spruce hybrid zone
(*Picea sitchensis* x *P. glauca*)**

by

Jill A. Hamilton

B.Sc.H. University of Winnipeg, 2002

M.Sc. Queen's University, 2005

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)

August 2012

© Jill A. Hamilton 2012

Abstract

Natural hybrid zones may be viewed as important biological systems for examining the role of selection in creating and maintaining species differences. Where ecological differences exist between hybridizing species, these zones may provide useful insight into the genetic architecture of important traits involved in adaptation. I have evaluated the genomic and phenotypic architecture of the economically and ecologically important *Picea sitchensis* (Sitka spruce) $\times$ *P. glauca* (white spruce) hybrid zone along the Nass and Skeena river valleys in northwestern British Columbia using chloroplast and mitochondrial markers, twelve microsatellite loci (SSRs), and 268 single nucleotide polymorphisms (SNPs), in combination with morphological variation, and phenotypic data from a common garden. Maternally- and paternally-inherited organelle markers, in combination with bi-parentally inherited nuclear markers, were used to estimate both the historic and contemporary direction and extent of gene flow within the hybrid zone. Sitka spruce mitotype ‘capture’ throughout the introgression zone point towards asymmetric gene flow, congruent with microsatellites and SNPs, indicating extensive long-term introgression and widespread recombination with more Sitka spruce than white spruce ancestry in hybrid populations. Significant clinal variation was observed for marker-based hybrid indices and morphological traits associated with climate and geography, while growth and cold hardiness traits evaluated in a common garden exhibited weak to non-significant clines. These results indicate extrinsic selection appears to play a strong role in the distribution and structure of this hybrid zone, which fits expectations for the environmentally-determined bounded hybrid superiority model of hybrid zone maintenance. However, intrinsic mechanisms of hybrid zone maintenance could not be ruled out. Finally, broad-scale patterns of variation, combined with fine-scale analysis of candidate SNP-specific patterns of introgression revealed a suite of candidate loci that may be targets of extrinsic or intrinsic selection. These loci may be involved in either adaptation to climate across the zone, particularly precipitation gradients, or involved in the maintenance of species barriers. These results have important implications for genetic conservation of adaptive variation, selection of seed sources for current reforestation within this ecologically transitional area, and appropriate scale and direction of seed transfer relating current genotype-climate associations to future climate predictions for this region.

Preface

This thesis was written as a series of manuscripts with the intent of publication in peer-reviewed journals. I took the lead in developing ideas, collecting and analyzing data, writing the draft manuscripts, along with figures and tables for all chapters. This research would not have been possible, however, without the contributions of my supervisor, Sally Aitken. She provided assistance with proposal development, experimental design, monetary support, editing and guidance throughout the execution of all chapters. For her efforts I have included her as a co-author on all chapters. Christian Lexer hosted me over three months at the University of Fribourg, providing analytical advice, expertise in hybrid zone analysis and thoughtful feedback with respect to manuscripts. For his efforts I have included him as a co-author on Chapter 3 and 4.

A version of chapter 3 has been accepted for publication at *Molecular Ecology* as:
“Hamilton, J. A., Lexer, C., and S. N. Aitken. Genomic and phenotypic architecture of a spruce hybrid zone (*Picea sitchensis* x *P. glauca*).”

A version of chapter 4 is currently in revision for resubmission to *New Phytologist* as:
“Hamilton, J. A., Lexer, C., and S. N. Aitken. Differential introgression reveals candidate genes for selection across a spruce (*Picea sitchensis* x *P. glauca*) hybrid zone.”

A version of chapters 2 will be submitted as “Hamilton, J.A. and S. N. Aitken. Genetics and morphology of a spruce (*Picea sitchensis* (Bong) Carr. x *P. glauca* Moench Voss) hybrid zone along a climatic gradient.”

Table of contents

Abstract.....	ii
Preface	iii
Table of contents	iv
List of tables	vi
List of figures	vii
Acknowledgements	viii
1. Literature review and research objectives	1
1.1 Introduction	1
1.2 Hybridization.....	2
1.3 Gene flow and hybridization	3
1.4 Natural selection and models of hybrid zone maintenance	5
1.5 Genomic resources in <i>Picea</i>	7
1.6 Background on <i>Picea sitchensis</i> x <i>P. glauca</i> zone of introgression	8
1.7 Climate and biogeoclimatic zones across the zone of introgression	9
1.8 Ecological and economic implications	10
1.9 Research objectives.....	12
2. Genetics and morphology of a spruce (<i>Picea sitchensis</i> x <i>P. glauca</i>) hybrid zone along a climatic gradient.....	15
2.1 Introduction	15
2.2 Materials and methods.....	18
2.2.1 Sampling methods.....	18
2.2.2 Molecular marker assays.....	19
2.2.3 Genetic structure analysis.....	20
2.2.4 Genomic composition	21
2.2.5 Morphological analysis	21
2.2.6 Statistical analysis	21
2.3 Results.....	22
2.3.1 Chloroplast and mitochondrial diversity and differentiation	22
2.3.2 Genetic diversity and differentiation	23
2.3.3 Genomewide estimate of admixture	24
2.3.4 Morphological data.....	25
2.3.5 Climatic and distance analysis	25
2.4 Discussion	25
2.4.1 Historic patterns of introgression.....	26
2.4.2 Nuclear genetic structure.....	27
2.4.3 Climate and distance relationships with ancestry and morphology.....	28
2.4.4 Replicate genetic and morphological structure	30
3. Genomic and phenotypic architecture of a spruce hybrid zone (<i>Picea sitchensis</i> x <i>P. glauca</i>).....	40
3.1 Introduction	40
3.2 Methods	43
3.2.1 Plant material and common garden experiment.....	43
3.2.2 Artificial freeze tests	43
3.2.3 SNP selection and genotyping.....	44
3.2.4 Genomic admixture analysis and hybrid identification	45

3.2.5 Quantitative genetic data analysis	47
3.2.6 Statistical analysis	47
3.3 Results.....	48
3.4 Discussion	50
3.4.1 Historic introgression.....	51
3.4.2 Pre- and post-zygotic incompatibilities.....	51
3.4.3 Genetic and phenotypic relationships with climate and distance.....	52
3.4.4 Genotypic classification	53
3.4.5 Fitness of genotypic classes	54
4. Differential introgression reveals candidates for selection across a spruce (<i>Picea</i> <i>sitchensis</i> x <i>P. glauca</i>) hybrid zone	65
4.1 Introduction	65
4.2 Materials and methods.....	68
4.2.1 Sampling	68
4.2.2 Genetic data analysis.....	68
4.2.3 Descriptive statistics	69
4.2.4 Identifying outlier loci potentially affected by selection.....	69
4.2.5 Genomic cline analysis	70
4.2.6 Geographic cline analysis	70
4.3 Results.....	72
4.3.1 Detection of loci under selection	73
4.3.2 Geographic clines.....	73
4.3.3 Genomic cline analyses	73
4.4 Discussion	74
4.4.1 Interspecific genetic differentiation reveals targets of selection	75
4.4.2 Geographic clines.....	77
4.4.3 Genomic clines	78
5. Thesis conclusions.....	87
5.1 Introduction	87
5.2 Neutral genetic structure of the Sitka-white spruce hybrid zone.....	89
5.3 Broad-scale genomic and phenotypic architecture of the spruce hybrid zone.....	90
5.4 Fine-scale patterns of differential introgression	91
5.5 Limitations of the current study	93
5.6 Future research	96
5.6.1 Test for asymmetrical barriers to genetic exchange	97
5.6.2 Gene expression and association analysis with drought tolerance	97
5.6.3 Predicting phenotypes using genomic ancestry	99
5.6.4 Genomewide resequencing and clinal variation	99
5.6.5 Conclusions	100
References	101
Appendix: supplementary tables and figures.....	114

List of tables

Table 2.1 Geographic origin, climatic variables and cone morphological characteristics of spruce provenances for neutral genetic structure study	32
Table 2.2 Genetic variability across 12 microsatellite loci for Sitka spruce, white spruce and their hybrids	33
Table 2.3 Relationship between hybrid index and morphological characteristics within the Sitka spruce and white spruce zone of introgression with climate and distance variables.....	34
Table 2.4 Analysis of covariance comparing geographic transects along the Nass and Skeena rivers for associations between hybrid index and morphology characteristics within the Sitka spruce and white spruce zone of introgression with climate and distance variables.....	35
Table 3.1 Geographic origin, climatic and geographic variables of provenances planted within a common garden experiment within the Sitka-white spruce hybrid zone and genotyped for single nucleotide polymorphisms (SNPs)	57
Table 3.2 Regressions of hybrid index on geographic and climatic variables within the Sitka-white spruce zone of introgression	59
Table 3.3 Analysis of variance results for phenotypic traits among parent and hybrid genotypic classes within the Sitka-white spruce zone of introgression.	60
Table 4.1 Summary statistics for genomic clines, geographic clines and genetic differentiation across all SNP loci within the Sitka-white spruce zone of introgression	82
Table 4.2 Analysis of molecular variance of 268 candidate gene SNPs between Sitka spruce and white spruce	83
Table S1 Principal component analysis summary of morphological variables of Sitka and white spruce cone scales and bractst	114
Table S2 Summary of locus-specific patterns across the Sitka-white spruce zone of introgression, including genetic data analysis, geographic and genomic cline analysis for 268 candidate gene SNPs.	115

List of figures

Figure 1.1 Rangewide distribution of Sitka spruce (<i>P. sitchensis</i>) and white spruce (<i>P. glauca</i>) within North America.....	14
Figure 2.1 Map of origin of provenances for neutral genetic structure study, species ranges of Sitka spruce and white spruce, and chloroplast haplotype frequencies within the Sitka-white spruce zone of introgression	36
Figure 2.2 Morphological measurements of spruce cone scales and bracts	37
Figure 2.3 Relationship between pairwise genetic and geographic distances for separate Nass (N) river, Skeena (S) river, and parent (P) reference populations	38
Figure 2.4 Relationship between microsatellite-based hybrid index and geographic and climatic variables of individuals sampled within the Sitka-white spruce zone of introgression.....	39
Figure 3.1 Map of origin of provenances sampled and common garden experiment location for SNP study, with rangewide distribution of Sitka spruce and white spruce	61
Figure 3.2 Triangle plot comparing interspecific heterozygosity vs. hybrid index of individuals sampled for SNP study within the Sitka-white spruce zone of introgression	62
Figure 3.3 Relationship between hybrid index and geographic and climatic variables of individuals sampled for SNP study within the Sitka-white spruce zone of introgression.....	63
Figure 3.4 Boxplots comparing phenotypic traits among hybrid and reference Sitka spruce and white spruce genotypic classes from individuals sampled from the common garden experiment	83
Figure 4.1 <i>Bayescan</i> plot of Bayes factor vs. locus-specific F_{ST} of 268 SNP loci between Sitka spruce and white spruce.....	84
Figure 4.2 Geographic clines, genomic clines, and genotypic distribution across a precipitation gradient for SNPs with narrow cline width within the Sitka-white spruce zone of introgression	85
Figure S1 Relationships between morphological PC1 (Length), and morphological PC2 (Width) with geographic and climatic variables spanning the Sitka-white spruce zone of introgression, including drainage distance (km), mean annual temperature (°C), mean annual precipitation (mm), and degree days below 0 °C.....	124
Figure S2 Regression analysis of hybrid index (Q-index), morphology PC1, and morphology PC2 of Sitka x white spruce (<i>Picea sitchensis</i> x <i>P. glauca</i>) individuals in separate river transects, the Skeena and Nass respectively with drainage distance (km), mean annual temperature (°C), mean annual precipitation (mm), and degree days below 0°C	125
Figure S3 Locus-specific geographic clines for 268 loci indicating the relationship between minor Sitka spruce allele frequency and drainage distance.....	126
Figure S4 Genomic clines for 268 loci indicating locus-specific patterns of introgression using the genomewide estimate of admixture	156

Acknowledgements

There are a number of people I'd like to acknowledge who have supported me throughout my PhD. First, to my supervisor, Sally Aitken, for her continued patience, support and guidance throughout the tenure of my degree. I am sincerely grateful for the opportunities she has provided and the breadth of knowledge she has shared. Thank-you to John King, and the BC Ministry of Forests, Lands, and Natural Resources Operations who provided data and allowed use of the incredible provenance trial resource within the Nass-Skeena region. To my committee members, Loren Rieseberg and Kermit Ritland, I am extremely grateful for the thoughtful advice and support provided that have improved this work along the way. I would also like to thank my 'sabbatical' host at the University of Fribourg, Christian Lexer, whose collaboration and approach to scientific query continues to instill enthusiasm. For technical assistance and guidance in laboratory components I am extremely grateful to Carol Ritland, Hesther Yueh, and Allyson Miscampbell, and volunteers Janine Welton and Kate English. I especially would like to acknowledge my partners in the field: Christine Chourmouzis, Lisa Erdle, Nina Lobo, and Jon Sweetman, without whom this project would not be possible. I am extremely grateful to Lisa Erdle and Jordan Bemmels for their efforts measuring cone scales. Special thanks to those who opened their homes to us, offering a warm bed and friendship, as we traveled throughout northwestern British Columbia: Lou Allison and Jeremiah Randall, Bill and Betty Geier, and Donna and John Geier.

I would also like to acknowledge members of the Aitken and Lexer labs. In particular, Jason Holliday, Nina Lobo, Pia Smets, Tongli Wang, Dorothea Lindtke, Kai Stölting, Celine Caseys and Maria Joao Amaral for many thoughtful conversations.

Finally, to my friends and family who have supported me throughout the term of my PhD, especially my husband, Jon Sweetman, who has been and continues to be my rock.

This project was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Canada Graduate Scholarship, University of British Columbia Graduate Fellowship, and UBC Forestry Strategic Recruitment Fellowship to J. Hamilton, Genome British Columbia funding to S. Aitken and K. Ritland and a NSERC Discovery grant to S. Aitken.

1. Literature review and research objectives

1.1 Introduction

Local adaptation results from the dynamic interactions between natural selection, gene flow, genetic drift and mutation (Bridle and Vines, 2006; Pertoldi *et al.*, 2007). Populations are considered locally adapted if they have the appropriate genetic composition to survive and reproduce under current and local conditions (Pertoldi *et al.*, 2007). But environmental conditions do not remain constant, thus the appropriate genetic composition will change through time. Increased genetic variation may lead to a reduced risk of extinction in the face of environmental stochasticity as there will be a greater likelihood of having the appropriate genetic composition present in the population to respond to changing environmental conditions (Hellmann and Pinedakrch, 2007). Increased genetic diversity due to interspecific hybridization may result in a further capacity to adapt to changing environments. Evidence suggests adaptation to novel environments is facilitated by interspecific hybridization, where hybridization results in increased genetic variation leading to a range of variation in functionally adaptive traits (Seehausen, 2004). In this study I use the genomic and phenotypic architecture of the Sitka-white spruce (*Picea sitchensis* x *P. glauca*) hybrid zone to make inferences regarding the evolutionary mechanisms contributing to hybrid zone maintenance, included gene flow and natural selection dynamics. The broad goal is to identify and quantify the genetic consequences of introgression, and the degree to which local adaptation in traits of economic, ecological value or those involved in adaptation to climate are exhibited across the hybrid zone.

The traditional goal of forest genetics is to produce forest reproductive material that is both genetically diverse and locally adapted (White *et al.*, 2007). This end-goal, which includes both generation of appropriate genetic material for breeding and measurement of proxies for harvest rotation-length fitness, is time-consuming in such long-lived organisms (Lexer *et al.*, 2004). Commercial tree breeding is, compared to other agricultural crops, still in its infancy, where most commercial species are only one to three generations from their wild progenitors (Dungey, 2001).

While substantial genetic gains in growth and quality have been achieved using traditional breeding methods, natural hybrid zones can circumvent the time required to develop and test experimental crosses, and thus offer a valuable tool for exploring genetic variation. The fitness of natural hybrids can be assessed directly in the field under natural conditions, allowing estimates of the role of natural selection within the native environment, which is particularly

important in these long-lived species (Lexer and Fay, 2005). Furthermore, the maintenance of hybrid zones has practical implications for forest genetics and conservation. Both silvicultural choices, specifically seed source selection for reforestation, seedling nursery and silvicultural regimes and establishment, along with genetic conservation strategies will be influenced by interspecific hybridization (Sutton *et al.*, 1994). A study of the factors that drive natural selection and local adaptation in tree hybrid zones will, therefore, have broad ecological, genetic conservation, and economic implications as both forest genetics and silviculture depend on understanding and managing genetic diversity in tree populations. Finally, this research hopes to build upon the wealth of knowledge developed from intra- and inter-specific provenance trials and genomic studies within conifer to help foresters identify seed sources for future climates. With a move towards molecular approaches in forest breeding, the goal will be to reduce the risk of planting maladapted seedlings by being able to predict local adaptation using information from molecular markers.

1.2 Hybridization

Hybridization is broadly defined as the successful mating between individuals from two populations or groups of populations, where populations are differentiated on the basis of one or more heritable traits (Harrison, 1990; Arnold, 1997). The term ‘hybrid’ is applied to both those offspring of an initial cross between parental lineages (F1s), as well as subsequent crosses within populations where varying degrees of admixture result from backcrosses and further hybridization (F2s, etc.). Both artificial hybridization beyond the F1-generation, a major component of plant breeding programs, and natural hybridization result in progeny of varying degrees of admixture. The progeny of natural hybridization, however, are the products of mating between wild populations in sympatry, allopatry or parapatry and comprise ‘hybrid zones’ of mixed ancestry (Harrison, 1993). Hybridization is viewed as an important source of new variation in which hybrids are the raw material of evolution and a source of functional novelty (Rieseberg and Wendel, 1993; Arnold, 1997). Although the evolutionary role of hybridization in animals remains contentious (Mayr, 1963; Seehausen, 2004), evolutionary botanists consider hybridization to be an important and frequent component of the life history variation in plant species (Harrison, 1993). Moreover, hybrid zones offer a natural laboratory for examining the mechanisms that underlie evolution in natural environments and are important sources of genetic variation on which selection may act (Hewitt, 1988).

Introgression is the permanent infiltration of the genes of one species into the genome of another through repeated backcrossing (Stebbins, 1959; Wheeler and Guries, 1987). Natural introgression may act to extend a species' gene pool where new adaptations are more likely to be developed by modifying existing adaptations than being produced *de novo*, producing a wide array of natural variation on which natural selection may act (Stebbins, 1959; Rweyongeza *et al.*, 2007). Hybrid zone theory predicts where parental species have different adaptive norms, hybrids may be intermediate to the parental species, although heterosis, the presentation of higher fitness in hybrids relative to parent species, is often observed in the early-generation, especially F1, recombinants (Rieseberg *et al.*, 1999a). Natural hybrids (particularly advanced generation), however, exhibit a range of fitness characteristics, displaying lesser, equal or greater fitness in adaptive characters relative to either parental species (Rieseberg and Ellstrand, 1993; Rieseberg, 1995). Moreover, some hybrids may exhibit adaptive characters outside the natural parental range, termed transgressive segregation, resulting in novel adaptations (Welch and Rieseberg, 2002). Transgressive segregation is distinguished from heterosis, which results in extreme phenotypes solely in a positive direction typically in early-generation recombinant classes, with phenotypes exceeding parents in either a positive or negative direction in a range of recombinant classes (Rieseberg *et al.*, 1999a). The mechanism attributed to transgressive segregation, complementary gene action following hybridization, may contribute substantially to adaptive evolution (Rieseberg *et al.*, 2003). Transgressive phenotypes are heritable, hence transgressive segregation may permit rapid niche or habitat divergence among hybrid lineages (Rieseberg *et al.*, 1999a). This may have direct implications in terms of conserving species' evolutionary potential and their ability to adapt in the face of climate change and increasing environmental stochasticity (Rweyongeza *et al.*, 2007).

1.3 Gene flow and hybridization

Determining how historic and contemporary processes influence and maintain patterns of genetic variation and differentiation within and among populations, and consequently affect their evolutionary potential, remains an important challenge in evolutionary biology. Patterns of variation in maternally, paternally, and bi-parentally inherited molecular markers provide a means to infer degree and direction of gene flow. The advantage of combining bi-parentally inherited nuclear markers (e.g. microsatellites, single nucleotide polymorphisms (SNPs)) with uni-parentally inherited organelle markers (mitochondrial and chloroplast DNA) is that, we can both estimate contemporary admixture among hybridizing populations and trace the long-term

effects of historical hybridization in the paternal and maternal genotypic distributions that are preserved in the organelle genomes (Whittemore and Schaal, 1991). Where haplotype structure is based on taxonomic identity, the distribution of mtDNA and cpDNA may indicate historical effects regarding the relative age of the introgression zone corresponding to the timing of secondary contact (Jaramillo-Correa and Bousquet, 2005; Anderson *et al.*, 2006; Ran *et al.*, 2006; Pyhäjärvi *et al.*, 2007). Furthermore, clinal variation in cytoplasmic and nuclear markers may vary, where steep clines across an environmental gradient reveals the historic structure of the cytoplasm, and nuclear genes exhibit more gradual clines of an advancing contact zone (Wu and Campbell, 2005). An *Ipomopsis* hybrid zone presents an excellent example of contrasting clinal variation across an environmental gradient using a combination of cpDNA and nuclear amplified fragment length polymorphisms (AFLP) markers, identifying a strong influence of environment influencing hybrid zone genetic structure in concert with historic and contemporary influence of gene flow (Wu and Campbell, 2005).

Unlike angiosperms, which typically exhibit maternal inheritance of both organelle genomes, the Pinaceae exhibit both a maternally inherited mitochondrial genome and paternally inherited chloroplast genome, which offers the unique opportunity to compare bi-parental lineages (Neale and Sederoff, 1989; Sutton *et al.*, 1991). This permits independent estimation of the contribution of each maternal and paternal species to the genetic structure of the hybrids (Neale and Sederoff, 1989). However, because there are numerous cases of ‘chloroplast capture’ described in the literature due to introgression, patterns of genetic variation between nuclear and organelle markers may not be congruent (Rieseberg and Soltis, 1991; Rieseberg and Welch, 2002).

Phylogeographic studies of the structure of sympatric *Quercus* species and *Eucalyptus* species based on organellar DNA indicate haplotype differentiation is not necessarily based on taxonomic identity, but rather geographic location (Whittemore and Schaal, 1991; McKinnon *et al.*, 2001; Petit *et al.*, 2002). Dumolin-Lapegue *et al.* (1997) noted in *Quercus* that maternally inherited chloroplast was transferred between species through repeated backcrossing. Similarly in *Eucalyptus*, putative ‘capture’ of cpDNA from *E. cordata* by *E. globulus* has been documented on several occasions and attributed to geographic proximity (McKinnon *et al.*, 2004; McKinnon *et al.*, 2010). In the case of both *Quercus* and *Eucalyptus* haplotype structure is linked to the recolonization process, identifying an important role for the direction of gene flow and the influence of pre- and post-mating barriers on the genetic structure of populations during and following range expansion (Petit *et al.*, 1997; McKinnon *et al.*, 2004).

The rate of introgression within plastid genomes may be greater than that of the nuclear genome due to linkage within the nuclear genome to loci under selection constraining introgression (Rieseberg and Soltis, 1991; Whittemore and Schaal, 1991). Assuming selective neutrality, the haploid nature and uniparental inheritance of the smaller cytoplasmic genomes suggest these genes are more likely to become fixed in the receiving population than nuclear genes (Martinsen *et al.*, 2001). Selection, however, may influence plastid genome introgression. The chloroplast and mitochondrial genomes may be under strong selection, resulting in geographic structuring of haplotypes based on local selective pressures (Sambatti *et al.*, 2008). While we often assume a neutral model of gene flow for both nuclear microsatellite and cytoplasmic markers, departures from neutrality are possible.

1.4 Natural selection and models of hybrid zone maintenance

Natural selection within a hybrid zone is broadly classified into two categories: (i) endogenous selection (environment-independent) and (ii) exogenous selection (environment-dependent; (Carney *et al.*, 2000). The tension zone model, described in detail by Barton and Hewitt (1985) describes a hybrid zone shaped by intrinsic genetic factors independent of the environment. A tension zone is maintained by the combination of gene flow into the hybrid zone and selection against hybrid formation, resulting in steep clines across contact zones (Barton and Hewitt, 1981). For example, recombination causing chromosomal structural differences or the disruption of coadapted gene complexes may lead to lower fitness and reduced viability of some hybrid genotypes, a form of endogenous selection (Arnold, 1997; Kawakami *et al.*, 2009). Where tension zones have been described, clinal variation coincident with physical barriers, steep parallel clines indicating abrupt changes in allele frequency, or significant linkage disequilibria contributing to stepped clines with shallow introgression tails have been identified (Szymura and Barton, 1986; Barton and Gale, 1993; Kawakami *et al.*, 2009). However, these patterns are not mutually exclusive to environmentally-independent clines, but bear consideration alongside hybrid zones maintained by environmentally-dependent selection.

Where there are no major barriers to hybridization, environment-dependent selection is implicated in two conceptual hybrid zone models: the mosaic model and the bounded hybrid superiority model (Arnold, 1997). The mosaic model is differentiated from the bounded hybrid superiority model based on spatial heterogeneity of the landscape. The mosaic model of hybrid zone maintenance has been ascribed to the *Populus alba*-*P. tremula* hybrid zone, where hybridization has been tightly linked to a habitat ‘mosaic’ of disturbance along a forest

floodplain (Lexer *et al.*, 2005). The habitat requirements for *P. alba* and *P. tremula* differ substantially, where the former favours floodplains areas adjacent to or within actively flooded areas and the latter is considered an upland species, characteristic of submontane or montane zones (Lexer *et al.*, 2005). The ‘mosaic’ habitat of the floodplain forest has resulted in hybridization between these two species and establishment of hybrid backcrosses associated with environmentally-mediated habitat-selection (Lexer *et al.*, 2005; Lexer *et al.*, 2007).

While the mosaic model of hybrid zone maintenance assumes a heterogeneous landscape or ‘mosaic’, the bounded hybrid superiority model assumes an ecological gradient between two parental habitats (Arnold, 1997). Furthermore, the bounded hybrid superiority model assumes that hybrids exhibit increased fitness relative to parental genotypes due to genotype-by-environment interactions within the ecologically transitional habitat (Moore, 1977). An explicit test for the bounded hybrid superiority model involves establishment of reciprocal transplants between parent and hybrid genotype groups across the native ecological condition of all groups comparing relative lifetime fitness or proxies thereof between environments (Carney *et al.*, 2000).

Several studies have evaluated the performance of hybrid and parent genotypes in a range of environments, including contact zones between *Artemisia* (Wang *et al.*, 1997; Miglia *et al.*, 2005), *Penstemon* (Kimball *et al.*, 2008), *Ipomopsis* (Campbell and Waser, 2001), *Prunella* (Fritsche and Kaltz, 2000), and *Iris* species (Emms and Arnold, 1997). Interestingly, in the case of *Artemisia*, early estimates of fitness within reciprocal gardens pointed toward increased fitness of hybrid genotypes relative to parents within the native hybrid zone common garden, supporting a bounded hybrid superiority model of hybrid zone maintenance (Wang *et al.*, 1997). However, evaluation of the same experiment nine years post-establishment identified a change in fitness ranking of parent and hybrid genotypes within the hybrid zone (Miglia *et al.*, 2005). While the results of this subsequent study still partly support the bounded hybrid superiority model and identify a significant role for environmental selection in the maintenance of the *Artemisia* hybrid zone, the role of intrinsic factors could not be entirely discounted (Miglia *et al.*, 2005).

Although models of hybrid zone maintenance clearly differ in their assumptions with respect to selection, the presence of either endogenous or exogenous selection does not preclude the other. On average, expectations follow that hybrids will be less fit, on average, than parents in a range of environments as recombinant genotypes have not been tested by natural selection (Barton, 2001). However, frequent observations of a range of fitness characteristics where

hybrids exhibit less, equal or greater fitness relative to parental genotypes support the combined influences of extrinsic and intrinsic selection in hybrid zone maintenance (Rieseberg and Ellstrand, 1993; Rieseberg, 1995). Consequently, one of the main goals of this study will be to tease apart putative environmental and non-environmental influences contributing to hybrid zone maintenance to better understand the role of environment and genetic factors in hybrid zone persistence.

1.5 Genomic resources in *Picea*

The large number of molecular tools available for *Picea* facilitates using the *P. glauca* x *P. sitchensis* zone of introgression for addressing questions regarding the genetic mechanisms of hybridization and local adaptation. The development of putatively selectively neutral microsatellites from non-coding regions (simple sequence repeats; SSRs) and expressed sequence tag SSRs (Hodgetts *et al.*, 2001; Rajora *et al.*, 2001; A'Hara and Cottrell, 2004; Rungis *et al.*, 2004; A'Hara and Cottrell, 2007) offers the opportunity to address questions regarding putatively neutral genetic structure, gene flow and selection within the spruce hybrid zone. As well, numerous studies have had success in characterizing haplotypic variation in *Picea* using both maternally inherited mtDNA (Ran *et al.*, 2006) and paternally inherited cpDNA markers (Vendramin *et al.*, 1996; Anderson *et al.*, 2006; Ran *et al.*, 2006).

The advantage to using single nucleotide polymorphisms (SNPs) within candidate genes, or in gene regions closely linked to them, is the potential for association with ecologically important adaptive variation (Mitchell-Olds and Schmitt, 2006). Recent studies have identified putative candidate genes, and SNPs within them that are associated with adaptive genes related to cold hardiness and bud set in Sitka spruce (Holliday *et al.*, 2008; Holliday *et al.*, 2010a), resistance to the white pine shoot tip weevil (*Pissodes strobi* (Peck)) in Sitka and white spruce (K. Ritland *et al.* unpublished data), as well as adaptive variation in growth and survival candidate genes found within white spruce in Quebec (Namroud *et al.*, 2008; Pavy *et al.*, 2008; Pelgas *et al.*, 2011). These genomic resources greatly facilitate the study of the genetic basis of adaptive variation in hybrid populations, correlating functionally important genetic polymorphisms with intraspecific variation and adaptive phenotypes. Where previous investigations within the spruce hybrid zone have characterized neutral genetic structure, these new genomic resources offer the opportunity to associate genetic polymorphisms with related function to environmental and phenotypic variation important to adaptation within the hybrid zone.

1.6 Background on *Picea sitchensis* x *P. glauca* zone of introgression

The introgression zone between Sitka spruce (*Picea sitchensis*) and white spruce (*P. glauca*) along the Nass and Skeena river valleys in northern British Columbia offers an excellent system for addressing questions regarding the genetic consequences of hybridization and the evolutionary mechanisms involved in adaptation across a zone of introgression. Sitka spruce has a long, narrow range along the Pacific coast from California to Alaska, whereas white spruce spans much of the northern latitudes of North America from coast to coast as a large component of the boreal forest (Figure 1.1). Introgression is also extensive in Alaska, where hybrids are called Lutz spruce (*Picea x lutzii* Little, (Little, 1953), however this study will focus on the Nass and Skeena watersheds of British Columbia.

Previous studies have characterized the introgression zone using morphology (Daubenmire, 1967; Roche, 1969), phenology (Roche, 1969), physiology (Grossnickle *et al.*, 1996; Fan *et al.*, 1997; Silim *et al.*, 2001), and molecular genetic markers (Yeh and Arnott, 1986; Bennuah *et al.*, 2004). Morphologically, individuals sampled from both the lower and upper Skeena river regions are intermediate in morphological traits characteristic of Sitka and white spruce, such as cone scales (Daubenmire, 1967; Roche, 1969) and cone bracts (Sutton *et al.*, 1994). Moreover, Daubenmire (1967) suggested that those hybrid individuals sampled from the lower Skeena appeared more Sitka-like and those from the upper Skeena more white spruce-like, suggesting an increased rate of backcrossing and introgression with the geographically more proximal parent.

In a review of the genecology of *Picea*, Roche (1969) observed that provenances within the Nass/Skeena region entered dormancy at a later date than those provenances east of the region, suggesting that an increasingly continental climate to the east has imposed selection pressure on phenological variation among hybrids. The wide degree of variation observed in physiological parameters, including gas exchange, drought and freezing tolerance measured from individuals across the introgression zone has been attributed to variation in parental contribution to the genome and environment-mediated selection (Grossnickle *et al.*, 1996; Fan *et al.*, 1997; Silim *et al.*, 2001).

Multiple studies have assessed the degree of introgression using molecular markers, including ribosomal RNA probes (Sutton *et al.*, 1994), allozymes (Yeh and Arnott, 1986) and sequence tagged-sites (STS;(Bennuah *et al.*, 2004). Both Bennuah *et al.* (2004) and O'Neill *et al.* (2002) supported Daubenmire's (1967) observation of greater introgression with the

geographically proximal parent, describing a cline in hybrid index values across the region that correlated with distance along the Skeena and Nass rivers to the nearest ocean inlet. Strong clinal variation in hybrid index coincident with the transition from maritime to continental climates observed within the Bennuah *et al.* (2004) study supported the bounded hybrid superiority model of hybrid zone maintenance, providing evidence for environmentally-mediated selection. To date there is no evidence of endogenous selection or hybrid inferiority (although the fitness of controlled F2 crosses has not been explicitly tested). The presence of a stabilized introgression zone between Sitka and white spruce is evident in all studies. The strong evidence for clinal patterns in selection gradients from a maritime to continental climate have resulted in the large degree of variation observed in morphological, phenological, physiological characters, and molecular characters across the hybrid zone.

1.7 Climate and biogeoclimatic zones across the zone of introgression

Sitka spruce is a significant component of the Coastal Western Hemlock (CWH) Biogeoclimatic Ecological Zone (BEC zone) that characterizes low to middle elevations along the coast of British Columbia and penetrates to some extent eastward into the subarctic region along major river valleys, including the Nass and Skeena river valleys (Pojar *et al.*, 1991). The wet maritime climate of the CWH makes one of the most productive forest regions in Canada, where species thrive at a mean annual temperature of approximately 8°C and mean annual precipitation exceeding two metres, with relatively moderate variation in temperature from winter to summer. The CWH shares many ecological features with the second most productive forest zone in Canada, the Interior Cedar-Hemlock Zone (Ketcheson *et al.*, 1991; Pojar *et al.*, 1991). The ICH includes, but is not dominated by, white spruce. It is characterized by a continental climate that is strongly influenced by easterly moving air masses from the Pacific Ocean. The ICH zone is distinguished by cold, wet winters and warm, dry summers, although periodic extremes in temperature occur due to dry, high-pressure continental air masses (Ketcheson *et al.*, 1991). The introgression zone between Sitka and white spruce occurs across the ecotone between the CWH and adjacent ICH zone in northwestern British Columbia. Both the ICH and CWH exhibit moist, cool climates throughout the area where introgression is prevalent.

1.8 Ecological and economic implications

This study has both ecological and economic implications with regards to silvicultural practices, selective breeding, climate change and genetic conservation in forestry. Sitka spruce, the third tallest conifer in the world, is an economically important tree with high value wood, garnering one of the highest log prices on the west coast (O'Neill *et al.*, 2002). Following harvest, however, only 2% of seedlings planted in areas where the species occurs are Sitka spruce due to the high susceptibility of planted seedlings to white pine shoot tip weevil (O'Neill *et al.*, 2002). White spruce exhibits greater resistance than Sitka spruce to this damaging insect. Moreover, hybrid genotypes with weevil-resistant characteristics have been identified within the zone of introgression (O'Neill *et al.*, 2002). Geographic and climatic variables correlated with weevil-resistant characters show similar clines to those based on hybrid index estimates using genetic markers (O'Neill *et al.*, 2002; Benuah *et al.*, 2004).

The expression of genes involved in terpenoid formation and phenolic secondary metabolism (both associated with resin duct formation) was previously investigated in the Treenomix project at UBC in both Sitka and white spruce, and gene expression changes markedly follows induced weevil damage (Ralph *et al.*, 2006; Keeling *et al.*, 2011). To estimate the contribution of Sitka spruce or white spruce to resistance at the gene-level, variation in select candidate genes associated with the constitutive and inducible defense systems, including terpene synthase genes, dirigent proteins, and a number of chitinases and glucanases will be assessed (Ralph *et al.*, 2006; Friedmann *et al.*, 2007). Hence, identification of those candidate genes associated with weevil resistance, along with their degree of introgression, may allow identification of candidate hybrid genotypes for replanting in areas where pure Sitka spruce regeneration has been decimated by this insect.

Over the past few decades, global warming has had significant impacts on many biological systems and predictions for the future imply massive biological shifts (Parmesan, 2006). Climate modeling scenarios predict extensive changes in population reproductive biology, phenology, and geographic ranges, as well as extensive community and ecosystem-level changes in composition and functioning (Marty, 2001; Hamann and Wang, 2006; Parmesan, 2006). Predictions of the magnitude and rate of change of climate over the next century under various scenarios suggest some tree populations may not have enough genetic variation for sufficiently rapid adaptation in response to intense directional climate-based selection (Jump and Penuelas, 2005; Aitken *et al.*, 2008).

Within this context, hybrids, and the consequent admixture of two distinct gene pools, may offer greater adaptive evolutionary potential than intraspecific variation alone (Aitken *et al.*, 2008). This includes an increased capacity to respond adaptively to a changing climate, increased capacity for range expansion, and in extreme cases where loss of parental species may be imminent, the persistence of locally adapted alleles at risk of extinction (Lewontin and Birch, 1966; Allendorf *et al.*, 2001). Healthier populations are also more likely to be able to migrate as they will likely have higher fecundity and the capacity for greater dispersal, thus adaptation and migration are not independent (Davis and Shaw, 2001). However, introgression from source populations with alleles maladapted to local conditions within recipient populations may also hinder adaptation (Aitken *et al.*, 2008; Holliday *et al.*, 2011). Consequently, identification of those candidate genes associated with local adaptation to climate, and degree of introgression of genes between heterospecifics across the hybrid zone, may have important implications with regards to future forest health, productivity, and ecosystem services within the introgression zone. Identifying those genotypes suitable for replanting in areas where climatic conditions have or are predicted to change will influence both the future economic value and conservation status of the species.

Current seed transfer standards in British Columbia are based on elevation and latitude, resulting in seed generally moved in a northerly and upward direction, although transfer is still fundamentally based on using local seed (O'Neill *et al.*, 2008a; Snetsinger, 2010). Sitka spruce readily hybridizes with white spruce, however Sitka and white spruce and their hybrids have all exhibited different optimal nursery cultural regimes based on photoperiod (Woods, 1988). Characterization of genotypes will have implications for rearing seedlings for replanting as well as for seed transfer decisions for reforestation (Woods, 1988; Sutton *et al.*, 1994). Seed transfer within the Nass-Skeena transitional area is problematic due to uncertainty regarding the varying adaptive requirements for optimal growth of hybrid genotypes that result in greater tree fitness and, ultimately, increased economic value. Seed transfer standards for the sub-maritime spruce (Sxs) hybrid zone are fairly stringent, favouring local seed in the seed planning zone of origin, restricting latitudinal movement to 2° north and 1° south and an elevational transfer within ± 200 metres from seed origin (Snetsinger, 2010). Interestingly, there is currently no limit to east-west transfer from the mean longitude of origin (Snetsinger, 2010), which may present additional uncertainty to seed transfer within this east-west transitional area (O'Neill *et al.*, 2008b). New seed transfer strategies linking climate (mean annual temperature and mean annual precipitation) with elevational transfer distance are currently being explored in British Columbia, however

eastward transfer and inter-regional transfer has yet to be considered (O'Neill *et al.*, 2008b). Conservation of genetic resources within this ecologically transitional area may become increasingly important with current climate change predictions, consequently strategies have identified populations spanning elevational gradients and peripheral populations at the hybrid zone margin as areas for *ex situ* collection (Krakowski *et al.*, 2009).

As a member of the CWH and ICH biogeoclimatic zones, Sitka spruce currently retains adequate *in situ* representation within protected areas within the CWH, but both Sitka spruce and white spruce occur with low frequency in protected areas of the ICH (Chourmouzis *et al.*, 2009). As these species are currently considered transitional within the ICH and are not predicted to increase their range within the ICH given current climate warming projections, they have not been identified as species of special concern in terms of genetic conservation (Chourmouzis *et al.*, 2009). However, as climate prediction models are refined (Mbogga *et al.*, 2010; Wang *et al.*, 2012b), and where local-interest for hybrid genotypes may persist or range expansion implications are informed, conservation of these ICH populations, along with adjoining CWH populations could prove increasingly valuable in the future (Chourmouzis *et al.*, 2009). The results of my study will provide genetic and genomic information with which to assess the adequacy of *in situ* targets and current genetic conservation within this zone.

1.9 Research objectives

In the following chapters I evaluate the genomic and phenotypic architecture of a conifer hybrid zone between Sitka spruce and white spruce by: (i) characterizing the neutral genetic and morphological structure of the hybrid zone spanning a climatic gradient; (ii) examining the broad-scale direction and extent of introgression across the hybrid zone, inferring the respective roles of genetics, climate and their interactions; and (iii) testing for fine-scale signatures of locus-specific selection inferred from differential patterns of introgression.

In Chapter 2, I evaluate gene flow across the introgression zone. I combine data from putatively neutral bi-parentally inherited nuclear microsatellite markers and uni-parentally inherited chloroplast and mitochondrial markers with morphological variation for samples from replicate transects along the Nass and Skeena river valleys. I predict if genetic markers are under selection within the introgression zone, there will be a strong and detectable correspondence between haplotype occurrence or allele frequency with geographic or climatic location. While I expect maternal and paternal haplotypes will be geographically structured, chloroplast haplotypes will likely exhibit geographic structure at greater distances due to long-distance

pollen dispersal, while mitochondrial haplotypes will exhibit significant geographic structure at shorter distances due to more limited seed dispersal. Finally, I predict that species-diagnostic haplotypes for organelle markers will reflect the independent history of Sitka and white spruce on a geographic and evolutionary scale.

In Chapter 3, I describe the genomic architecture of the hybrid zone and examine the role of extrinsic and intrinsic selection in hybrid zone maintenance. I analyze genotypes of individuals sampled across the hybrid zone and planted within a common garden experiment for 268 candidate gene SNPs. I predict hybrids will exhibit fitness equal to or greater than either parental species within the ecological conditions found within the hybrid zone, supporting the bounded hybrid superiority model of hybrid zone maintenance.

In Chapter 4, I describe differential patterns of locus-specific introgression across the hybrid zone using outlier tests, geographic clines, and genomic clines to identify candidate gene SNPs that may be targets of or linked to targets of natural selection. I predict that candidates for differential selection between species will exhibit more differentiation than expected under neutral conditions as estimated from the genome-wide average. In addition, I expect the rate and extent of fine-scale introgression will vary widely depending upon the magnitude of selection across the landscape and within the genome.

Finally, in Chapter 5 I discuss my most significant findings and identify future research directions that will build upon the previous and current body of research on this spruce hybrid zone, and on plant hybrid zones more broadly.

Figure 1.1 Rangewide distribution of Sitka spruce (*P. sitchensis*; dark grey) and white spruce (*P. glauca*; light grey) within North America. The red box highlights the general area of hybridization in British Columbia.

2. Genetics and morphology of a spruce (*Picea sitchensis* x *P. glauca*) hybrid zone along a climatic gradient

2.1 Introduction

The genetic structure of natural populations is determined by the complex interactions between natural history and evolutionary forces, including gene flow, drift and natural selection. Natural hybrid zones provide an additional layer of structural complexity within this context. Horizontal transmission of genetic material between species may result in the gradual exchange or capture of genetic material leaving genetic signatures of historical influences (Rieseberg and Welch, 2002). Furthermore, these products of mixed ancestry often contain a wide array of novel interspecific variation on which natural selection may act (Stebbins, 1959; Rweyongeza *et al.*, 2007). Fitness variation among these natural genetic recombinants influences the maintenance and stability of a hybrid zone, and ultimately its contemporary evolutionary trajectory (Culumber *et al.*, 2011). Consequently, the genetic structure of natural genetic recombinants offers the unique opportunity to examine factors influencing hybrid zone maintenance across various time scales, both preceding and following secondary contact.

The influence of the Pleistocene epoch on the genetic structure of plant populations throughout the Pacific Northwest has been well described using combinations of molecular markers relating the degree and direction of gene flow and colonization during range expansion (Soltis *et al.*, 1997; Shafer *et al.*, 2010). Long-term effects of historical hybridization may be preserved in organelle genomes (Whittemore and Schaal, 1991), while contemporary dynamics maintaining the introgression zone may be revealed by the association of nuclear molecular markers or morphological traits with geography and climate (Culumber *et al.*, 2011).

Unlike angiosperms, which typically exhibit maternal inheritance of both organelle genomes, the Pinaceae exhibit both a maternally inherited mitochondrial genome and paternally inherited chloroplast genome (Neale and Sederoff, 1989; Sutton *et al.*, 1991). Due to these contrasting modes of inheritance, the combination of chloroplast (cpDNA) markers (reflecting pollen-mediated gene flow), mitochondrial (mtDNA) markers (reflecting seed-mediated gene flow), and bi-parentally inherited nuclear markers can further elucidate both historic and contemporary patterns of gene flow and differentiation across a zone of introgression. Previous investigations of haplotype structure in both chloroplast and mitochondrial genomes of parental and closely related species have provided distinct patterns of variation resulting from

independent histories (Jaramillo-Correa and Bousquet, 2005; Anderson *et al.*, 2006; Ran *et al.*, 2006).

Both contemporary evolutionary forces and historical influences contribute to the distribution of genetic and phenotypic variation within an introgression zone, along with historic influences. Models of hybrid zone maintenance implicate either extrinsic (environmental) or intrinsic (non-environmental) selection where these contact zones persist, and empirical studies have found that while some zones are maintained by one or the other, others involve both (Moore, 1977; Barton and Hewitt, 1985). Spatial variation in selective pressures across heterogeneous environments is often associated with genetic and phenotypic differentiation, supporting a strong role for local adaptation linked to climate (Mimura and Aitken, 2007; Savolainen *et al.*, 2007; Holliday *et al.*, 2010a). However, pre- and post-zygotic mechanisms of selection within hybrid zones may further contribute to patterns generated, influencing the direction and extent of introgression (Rieseberg and Blackman, 2010; Field *et al.*, 2011). Therefore, where environmental selection gradients influence the genetic and morphological composition of populations within the hybrid zone we expect similar gradients to produce replicate transects (Culumber *et al.*, 2011).

By combining mtDNA, cpDNA and nuclear markers with morphological information, we aim to dissect the influences of history, geography, climate and their interactions within the Sitka spruce (*Picea sitchensis* (Bong) Carr.) and white spruce (*P. glauca* Moench Voss) zone of introgression. The introgression zone between Sitka spruce and white spruce provides an excellent opportunity to evaluate how the historic and contemporary influences of gene flow and selection facilitate the spread of genetic variation. Sitka spruce exhibits a narrow latitudinal distribution along the Pacific coast from California to Alaska. The tallest of spruce species, Sitka spruce is found in the Coastal Western Hemlock ecosystem in British Columbia, characterized by mean annual temperatures of approximately 8°C and annual precipitation values exceeding two metres (Pojar *et al.*, 1991). White spruce is distributed from Alaska to Newfoundland and is a major component of the North American boreal forest. Within British Columbia white spruce is found within the Interior Cedar Hemlock ecosystem which is characterized by warm, dry summers and cold, moist winters (Ketcheson *et al.*, 1991). The introgression zone between the two species spans a maritime to continental climatic gradient along the Nass and Skeena river valleys in northwestern British Columbia and into Alaska. The introgression zone along the Nass and Skeena rivers characterizes the transition between coastal and inland ecosystems. Previous research has described genetic and morphological clines

associated with both climatic gradients and geographic distance variables spanning the contact zone (Roche, 1969; Bennuah *et al.*, 2004).

Phylogeographic evidence support a glacial refugia of Sitka spruce along the Pacific coast south of the Cordilleran ice sheet, while white spruce likely refuged in both Alaska and east of the Rocky Mountains south of the Laurentide ice sheet (Soltis *et al.*, 1997; Anderson *et al.*, 2006; de Lafontaine *et al.*, 2010). Recolonization history of the Nass-Skeena indicates a stepping stone model of colonization along the Pacific coast for Sitka spruce, while white spruce likely moved both south from Alaska and west through the Rockies (Anderson *et al.*, 2006; Holliday *et al.*, 2011). The geographic history experienced by each species during the last glacial maxima suggests that prior to contact they took independent routes to the zone of introgression (Daubenmire, 1967; Anderson *et al.*, 2006; Mimura and Aitken, 2007; de Lafontaine *et al.*, 2010).

Previous work within the introgression zone describes Sitka spruce mitotypes (mtDNA haplotypes) further into the eastern end of the introgression zone than expected, while similar patterns have not been observed in chlorotype (chloroplast haplotype) structure (Sutton *et al.*, 1991; Sutton *et al.*, 1994). These results, however, must be interpreted with caution as Sutton *et al.* (1991; 1994) did not use white spruce as the reference parent, but rather used interior spruce, which includes mostly hybrids resulting from white spruce and Engelmann spruce (*P. engelmannii*), a third spruce species found at higher elevations in eastern British Columbia. Multiple studies have associated ancestry with cone and needle morphology, physiological traits, insect resistance and neutral genetic structure (Roche, 1969; Silim *et al.*, 2001; O'Neill *et al.*, 2002; Bennuah *et al.*, 2004). Steep climatic gradients have consistently been associated with patterns of genetic variation observed, indicating a strong role for extrinsic selection in the maintenance of the hybrid zone.

Current seed transfer standards in British Columbia are based on elevation and latitude, resulting in seed generally moved upward and northward, presenting uncertainty to seed transfer within this east-west ecological transition zone (O'Neill *et al.*, 2008b). Specifically, within the maritime and submaritime regions spanning the Nass and Skeena rivers, while there is preference for local seed, movement is restricted to 2°N latitude and 1°S latitude, $\pm$ 200 metres elevation, although there are few restrictions spanning the longitudinal east-west axis (Snetsinger, 2010). Furthermore, reforestation efforts within the region remain minimal, where currently only 2% of reforestation sites are planted with seedlings (O'Neill *et al.*, 2002). Seed transfer within the Nass-Skeena transitional area has remained problematic due to uncertainty

regarding the varying adaptive requirements for optimal growth of hybrid genotypes.

Association of admixed populations with particular environments across the introgression zone will have implications both for reforestation and direction of seed transfer given present and future climate scenarios. This study aims to elaborate on previous work within this hybrid zone describing historic and contemporary patterns of putatively neutral genetic and morphological variation to infer direction and extent of introgression over time.

In this study we examine the genetic composition of the Sitka-white spruce hybrid zone; exploring variation in chloroplast and mitochondrial sequence data, along with nuclear microsatellite data to determine how patterns of gene flow vary between nuclear and organelle genomes. We combine these data with morphological measures of spruce seed cones and bracts sampled along replicate transects across the hybrid zone spanning from maritime to continental climates. Specifically, we ask the following questions: (1) What is the geographic structure of maternal and paternal haplotypes, reflecting pollen and seed dispersal, and can it reveal the independent history of Sitka spruce and white spruce on a geographic and evolutionary scale? (2) Is there congruence between genetic clines for putatively neutral genetic markers and morphological clines across the climatic gradients that characterize the zone of introgression? (3) Do patterns of variation along replicate transects reflect similar patterns of introgression associated with geographic and climatic selection gradients? This research will inform forest management strategies regarding direction and extent of contemporary introgression and historic patterns of gene flow, and will assist with the selection of seed sources and the direction of seed transfer along environmental gradients within this ecologically transitional area.

2.2 Materials and methods

2.2.1 Sampling methods

Newly flushed lateral buds were sampled with a pole pruner from 582 trees across 21 populations across the introgression zone between Sitka and white spruce along the Nass and Skeena rivers in May 2009 (Table 2.1, Figure 2.1). In addition, two reference pure parental species populations from locations geographically well-separated from the contact zone were sampled: Sitka spruce from Haida Gwaii, BC and white spruce from Fort Nelson, BC. A minimum of 20 trees were sampled per population. Seed cones from the previous year still present on individual trees were sampled where present for morphological analysis (Table 2.1). Buds were frozen in liquid nitrogen and stored at -80°C for subsequent DNA extraction. All 582

individuals were genotyped for microsatellite markers, while a subset of individuals were randomly chosen for chloroplast and mitochondrial sequencing (Table 2.1).

2.2.2 Molecular marker assays

Genomic DNA was extracted from 50 mg of ground frozen needle tissue using a modified CTAB protocol (Doyle and Doyle, 1990). An initial screen of 34 microsatellite markers developed for Sitka spruce, white spruce or both, were amplified using a polymerase chain reaction (PCR) protocol using previously published PCR programs and annealing temperatures (Hodgetts *et al.*, 2001; A'Hara and Cottrell, 2004; Rungis *et al.*, 2004). Based on this initial screen, 12 microsatellite markers were selected for amplification in all individuals. PCR reactions were performed using an MJ Research PTC-100 thermal cycler in 10 µL reactions of 20 ng of genomic DNA, 1.0 µL of 2.0 mM dNTP, 0.05 µL 10x *Paq5000* DNA polymerase (Stratagene), 0.5 pmol of M13 IR Label Primer (LiCor) and 1pmol each of M13-labelled primers. PCR products were resolved in 5% polyacrylamide gels on a Li-Cor 4200 automated DNA sequencer. Fragments were scored and sized using Gene ImagIR with 50 to 350 bp ladders (LiCor) size standards for reference.

Two plastid DNA fragments, one cpDNA and one mtDNA, were examined for polymorphisms across individuals from throughout the hybrid zone. The mtDNA marker used in this study is a fragment of the *nad5a* gene (Jaramillo-Correa and Bousquet, 2003), while the cpDNA marker is a fragment of the non-coding *trnTF* region (Taberlet *et al.*, 1991). Markers were chosen based on a test pre-screen for haplotype variation from reference parent species populations indicating unique Sitka and white haplotypes for both plastid markers. MtDNA primers were re-designed from sequences obtained to target those regions that exhibited differentiation between parent species (*nad5a.1*-Forward: 5'-CGCATATGGGTAGCAAGAGGGC-3' and *nad5a.1*-Reverse: 5'-GAGGTTTCCCATCACACGGCTCACC-3'). Each PCR mixture contained 40 ng of DNA, 3 x PCR buffer (Stratagene), 2 µL of 2.0 mM dNTP, 1 mM each of each primer, 0.05 µL of 10x *Paq5000* DNA polymerase (Stratagene), and sterile water to a volume of 30 µL. PCR protocol from Ran *et al.* (2006) was used for *trnTF* and *nad5*. Five µL of each sample was run on a 1% agarose gel to verify amplification. The remaining PCR products were sent to Macrogen Inc. (Korea) for sequencing with the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems). We used Bioedit (v 7.1.3) to edit and combine DNA sequences.

2.2.3 Genetic structure analysis

Diversity and differentiation measures were first estimated in *Arlequin 3.0* and *GenoDive* ver2 (Meirmans and Tienderen, 2004; Excoffier *et al.*, 2005). Here, we combine F -statistics genetic differentiation estimate, F_{ST} (Wright, 1951) with genetic differentiation estimate R_{ST} developed specifically for microsatellites based on the stepwise mutation model, as opposed to the infinite alleles model assumed using F -statistics (Wright, 1951; Slatkin, 1995).

We calculated observed and expected heterozygosities to test for deviations from Hardy-Weinberg equilibrium and the possible influence of null alleles in both reference parent and hybrid populations using exact tests in *Genepop 4.0.10* (Raymond and Rousset, 1995). Null alleles are a potential problem for all microsatellite-based studies, and may be particularly problematic during cross-species amplifications (Lexer *et al.*, 2005). The reliability of the genotypic data was assessed for the presence of stuttering, large allele dropouts or null alleles using the program *Microchecker* (Van Oosterhout *et al.*, 2004). While stuttering and large allele dropouts were not detected, some evidence of null alleles was observed in half the loci. To adjust for the presence of nulls in our estimates of differentiation we used the program *FreeNA* (Chapuis and Estoup, 2007). This program estimates F_{ST} using methods described in Weir (1996), along with values adjusted for the presence of null alleles (F_{ST}^{ENA}) (Chapuis and Estoup, 2007). Both F_{ST} and F_{ST}^{ENA} were estimated globally and for individual loci. Bootstrap resampling averaged across all loci provided a 95% confidence interval for the evaluation of F_{ST} adjusted for nulls, with those unadjusted for the presence of null alleles (Table 2.2). Unfortunately, to date there has been no program developed to estimate the influence of nulls in R_{ST} .

Pairwise genetic distance was estimated between all pairs of populations. The correlation between genetic distance and geographic distance was used to test for patterns of isolation by distance using the *IBDWS* program (v.3.21, (Jensen *et al.*, 2005). Geographic distance was calculated using the great circle distance between populations. A Mantel's test with 1000 permutations was used to compare matrices of linearized estimates of genetic distance ($F_{ST}/1-F_{ST}$) with geographic distance. In addition, we highlighted pairwise comparisons between populations in the Nass (N) and Skeena (S) watersheds, and with parent populations (P) respectively.

2.2.4 Genomic composition

We estimated genetic ancestry among individuals using the Bayesian clustering approach applied in *Structure* (Version 2, (Pritchard *et al.*, 2000)). Multilocus genotypes were used to describe the posterior probability (*q-index*) of individual genotype assignment to k genotypic clusters. Five replicate runs of the models using k from one to 23 confirmed $k = 2$ had the highest log-likelihood probability (results not shown). From this we assume that $k = 2$ represents two species, Sitka spruce and white spruce. The *q-index*, or hybrid index, provides an estimate of the individual parent species' proportional contribution to the admixed gene pool. Hybrid index is likely impacted by the presence of null alleles; however, the dominant allele model option in *Structure* accounts for the presence of null alleles by designating a 'dummy' recessive allele (Falush *et al.*, 2007). *Structure* runs were carried out using the recessive alleles model, with a burn-in step of 50 000, followed by 100 000 Markov chain Monte Carlo iterations. The frequency of the additional recessive null allele is accounted for when calculating the posterior probability of ancestry (*q-index*) during assignment to k genotypic categories.

2.2.5 Morphological analysis

We measured six morphological components of spruce cone scales and bracts; including five width and length measurements estimated based on Roche (1969), along with an additional width measurement (W3) of the bract at its widest point (Figure 2.2). Morphological gradation in cone scales provided the first evidence for extensive introgression across the Sitka-white spruce complex in northwestern British Columbia (Roche, 1969).

We measured a minimum of six different cones per location, and three scales and bracts per cone. Cone scales and bracts were sampled from the middle of the cone and a digital photograph of each was taken. Measurements were taken using the measurement tool in Adobe Photoshop Elements (6.0) using a conversion factor from "screen" to "true-life" measurements for each photo by determining the number of screen centimeters corresponding to one centimeter ruler estimates with each set of scales and bracts. In order to summarize the variation in morphology we used principal components analysis.

2.2.6 Statistical analysis

The relationships between both hybrid index (*q-index*) and cone morphology (first two principal components) and environment were examined using regressions of both climate and distance variables. The geographic origin of individuals, described by latitude, longitude and

elevation were input into *ClimateBC v.3.2* (Wang *et al.*, 2006a), which estimates seasonal and annual climate variables based on geographic coordinates. Non-linear physical drainage distance up river valleys from the Pacific Ocean was estimated using *ArcGIS* (Version 10), replicating methods from Bennuah *et al.* (2004). Distance from the ocean in this region is strongly associated with climatic gradients spanning from maritime to continental climates. Univariate regression was performed on individual climate variables, identifying those climate variables that exhibited strong clinal variation that may be important in terms of local adaptation. As non-linear clines are more common across hybrid zones (Arnold, 1997) the use of non-linear regression may be more appropriate for the present analysis, therefore linear and non-linear regressions were estimated and compared using Akaike's Information Criterion (AIC). Using ANCOVA, we compared regressions within each transect individually, comparing environmental associations between the Nass and Skeena rivers. Using the river as a group variable we evaluated the slope and intercept of the regressions within the shared transect distance for differences between rivers. For the purpose of ANCOVA all inverse fit ($y=1/x$) regressions were transformed taking the inverse of the dependent variable to perform a linear comparison between the rivers.

2.3 Results

2.3.1 Chloroplast and mitochondrial diversity and differentiation

Two chloroplast haplotypes were observed within the ~500 bp region in the 255 individuals sequenced, one fixed in the Haida Gwaii reference Sitka spruce population, and one in the Fort Nelson reference white spruce population. These haplotypes were distinguished by single nucleotide base pair mutations at five sites across the sequenced region (at 87, 188, 193, 268, and 285 bp) and a single base pair deletion 339 bp from the 5' end of the sequenced region. The Sitka spruce haplotype occurred at a frequency of 77% across all individuals sequenced outside of the reference parent individuals, although both haplotypes were observed in 6 out of 23 populations (Figure 2.1). Of the 17 populations that exhibited fixed chloroplast haplotypes, only three were fixed for the white spruce haplotype: the Bulkley and Smithers populations on the eastern end of the hybrid zone, and the reference Fort Nelson white spruce population (Figure 2.1). Patterns along the two transects were largely concordant. Westernmost polymorphic populations along the Skeena and Nass had 90% and 92% frequency of the Sitka chlorotype respectively, while easternmost polymorphic populations exhibited 40% and 16% of the Sitka spruce haplotype for the Skeena and Nass. The spatial association of Sitka spruce

haplotypes declining with increasing drainage distance up river valleys indicates that geographic location is associated with chloroplast haplotype observed.

Only two mitochondrial haplotypes were observed within the ~500 bp region sequenced. The mitochondrial haplotypes distinguishing white spruce mitotypes from Sitka spruce mitotypes were identified as having a four nucleotide tandem base repeat (CTTGACTTG) at 276 base pairs from the 5' end of the sequenced *nad5a* region. This mitotype was found solely in the 11 individuals within the Fort Nelson reference white spruce population. The alternative haplotype was fixed in all other populations, including hybrid populations and the reference Haida Gwaii Sitka spruce species reference population.

2.3.2 Genetic diversity and differentiation

The twelve microsatellites were highly polymorphic with up to 20 alleles segregating for individual loci among populations. Expected (H_E) and observed heterozygosities (H_O) as high as 0.933 (range from 0 to 0.933) and up to 1.00 (range from 0 to 1.00), respectively (Table 2.2). Average expected heterozygosity was greatest in hybrid populations ($H_E=0.575$), although average observed heterozygosity was lower ($H_O=0.441$). Deviations from Hardy-Weinberg equilibrium were significant for two loci within the reference Sitka spruce population, and four loci within white spruce. Population subdivision may contribute to these significant deviations, but the presence of null alleles within reference parental species' populations is a substantial contributing factor. A greater number of loci (seven) exhibited significant deviations from HWE within hybrid populations. This is consistent with potential departures from random mating typical of hybrid zones, a potential factor where gene flow is asymmetric (Lexer *et al.*, 2005).

Only the *SPAGG3* locus exhibited deviations from HWE within both reference parent species and hybrid populations, indicating that in the majority of cases null alleles are likely present in one parent and sometimes hybrid populations, but not both reference populations. Because null alleles are likely present for *SPAGG3* in both parent species and hybrid populations, it was excluded from further analysis. The presence of null alleles likely biases estimates of the inbreeding coefficient (F_{IS}) upwards across all loci that exhibited deviation from HWE due to heterozygote deficiency. As a result, we accounted for the presence of null alleles in all subsequent estimates of genetic differentiation and hybrid indices.

Overall, we observed low locus-specific values of genetic differentiation as estimated by F_{ST} , R_{ST} and F_{ST}^{ENA} . Global estimates of genetic differentiation ($F_{ST}=0.05$) are within the range

of that estimated from intraspecific studies (Jaramillo-Correa *et al.*, 2001; Gapare and Aitken, 2005; Mimura and Aitken, 2007) indicating that gene flow is likely widespread across the hybrid zone and there are few barriers to reproduction. Interspecific differentiation between allopatric reference species populations is much greater (0.23) than the overall average across all populations, providing further evidence for high levels of gene flow within the hybrid zone. There is little difference between F_{ST} and F_{ST}^{ENA} , which corrects for the presence of null alleles, and values for R_{ST} are mostly within ± 0.03 of F_{ST} . Large discrepancies were observed in the *WS0092.A19* locus suggesting that these estimates may be influenced by the presence of null alleles or allele size, reflected in the mutation rate of the stepwise mutation model.

Isolation by distance was estimated using pairwise comparisons between genetic ($F_{ST}/(1-F_{ST})$) and geographic distance for Skeena River (S), Nass River (N), and parent species (P) reference populations respectively. Using a Mantel's test with 1000 permutations the correlation between genetic distance and geographic distance for all populations indicate a significant correlation between distance measures ($Z=5969.20$, $r=0.59$). The bootstrapped linear model with all populations accounts for 20% of the variation across 253 permuted comparisons. Pairwise comparisons between Nass (NN), Skeena (SS), Nass-Skeena (NS, Figure 2.3A) along with parent (PP), Nass-Parent (NP), Skeena-Parent (SP) populations indicate that pairwise differences are geographically structured (Figure 2.3B). Pairwise comparisons within the Nass (NN), the Skeena (SS), and between (NS) overlap considerably, suggesting that the majority of differentiation observed results from genetic and geographic distance from pure parent populations.

2.3.3 Genomewide estimate of admixture

Admixture was estimated based on microsatellite markers using *Structure*, where a q -index of one indicates pure Sitka spruce, and zero indicates pure white spruce. A high proportion of individuals within the hybrid zone exhibit mixed ancestry. Of the 582 individuals analyzed, *Structure* identified 177 individuals with a q -index greater than or equal to 0.9, 99 equal to or less than 0.1 and 306 individuals ranging between. The average admixture estimate of 0.60 suggests a slight excess of Sitka spruce ancestry within individuals sampled. This is consistent with a previous study using an independent sample of populations combined with genotypes for single nucleotide polymorphisms for candidate genes for local adaptation (Hamilton *et al.* submitted).

2.3.4 Morphological data

Over half of the variation in cone scale and bract morphology was explained by the first principal component (50.8%). All scale and bract length measurements had positive loadings on the first principal component (PC1) corresponded to length, while width estimates had positive loadings on PC2, accounting for an additional 26.9% of the variation (Table S1). Both L1 and L3 appeared to contribute the most to length variation contributing to PC1, while W1 and W2 contributed the most respectively to the width principal component (PC2, Figure 2.2).

2.3.5 Climatic and distance analysis

Regressions of the nuclear microsatellite-based *q-index* and the first two morphological principal components with climate and geographic variables revealed significant clines (Table 2.3, Figure 2.3, Figure S1). Clines associated with mean annual precipitation (MAP) explained most variation, while drainage distance, mean annual temperature (MAT), and degree days less than 0°C were identified as strongly associated morphology and ancestry. Linear relationships produced the best fit model for both drainage distance and degree days, while an inverse transformation of MAT and MAP produced the best model fit. Drainage distance up rivers from the coast exhibited significant linear relationships with respect to *q-index* ($R^2=0.62$, $p<0.001$, Figure 2.3). Likewise, morphological variables exhibited associations with climate and distance variables, although to a lesser degree (PC1, $R^2=0.14$, $p<0.001$; PC2, $R^2=0.29$, $p<0.001$; Table 2.3, Figure S1). Associations were greatest with MAP, which exhibited steep inflection indicating a possible threshold at approximately 1,000 mm of annual precipitation.

A comparison between the two transects along the Nass and Skeena rivers, using ANCOVA, indicated that for the most part there was little difference between slope or intercept of the two rivers using regression (Table 2.4). While the ANCOVA assumes a linear relationship, we found that an inverse function explained more variation in select relationships, similar to results in Table 2.3, therefore we provide the river-specific slope and intercept of those relationships in Table 2.3 along with the supplementary Figure S2. The clines for ancestry (*q-index*) differed between rivers for drainage distance, MAT, and DD<0°C, although no difference was observed for MAP, nor for clines associated with morphological variables.

2.4 Discussion

The genetic and morphological architecture of the Sitka-white spruce zone of introgression reflects colonization history, direction of gene flow, and contemporary climatic

gradients. Genetic patterns preserved in the organelle genome are consistent with Sitka spruce colonization prior to white spruce. This appears to have resulted in unidirectional introgression reflected in ‘mitochondrial capture’ of the maternally inherited Sitka spruce mitotype across the hybrid zone. In contrast, chloroplast haplotypes, reflecting pollen donor, indicate spatial patterning across the hybrid zone that may reflect pre-zygotic barriers to reproduction.

Low global estimates of genetic differentiation among populations within the hybrid zone based on nuclear microsatellite markers suggests that gene flow is widespread and interspecific barriers to reproduction are weak. A modest excess of Sitka spruce ancestry and corresponding deficiency of white spruce ancestry is consistent with asymmetric introgression observed in an independent investigation using SNP markers (Hamilton *et al.* submitted). Genetic clines along climatic and geographic gradients for both ancestry (*q-index*) and cone morphology suggest that extrinsic selection plays a strong role in maintaining the hybrid zone. Indeed, clines along distinct geographic transects along the Nass and Skeena river valleys reveal parallel patterns, particularly for drainage distance along rivers from the ocean and annual precipitation gradients.

2.4.1 Historic patterns of introgression

The phylogeographic history of Sitka spruce and white spruce indicates independent routes were likely taken to the Nass-Skeena region during post-glacial recolonization following the Pleistocene epoch (Anderson *et al.*, 2006; de Lafontaine *et al.*, 2010; Shafer *et al.*, 2010). Recent molecular work and palaeoclimatic envelope modeling confirm separate refugia and recolonization routes during climate warming (Mimura and Aitken, 2007; Roberts and Hamann, 2012). Pollen records describe a dominance of *Picea* pollen in eastern British Columbia, putatively attributed to white spruce, approximately 9,000 YBP (Pisaric *et al.*, 2003), whereas paleoclimatic evidence indicates Sitka spruce rapidly expanded its range along the Pacific coast around 15,000 YBP (Pielou, 1991; Roberts and Hamann, 2012). These patterns support the colonization of Sitka spruce within the Nass-Skeena region in advance of white spruce (Pisaric *et al.*, 2003; Mimura and Aitken, 2007). Sitka spruce likely employed a stepwise recolonization pattern along the Pacific Coast in response to glacial retreat (Soltis *et al.*, 1997; Holliday *et al.*, 2011), whereas white spruce populations within the Nass-Skeena may have colonized via an Alaska refugium or expansion through the Rocky Mountains (Anderson *et al.*, 2006; de Lafontaine *et al.*, 2010).

The sequence of post-glacial recolonization of the Nass-Skeena region may be preserved in the fixation of the Sitka spruce mitotype throughout the introgression zone. Consistent with

previous work, these findings support uni-directional introgression from the local (Sitka spruce) and later-colonizing (white spruce) species contributing to the distribution of mitochondrial and chloroplast haplotypes.

Our results suggest that established Sitka spruce individuals were the likely seed parents of early hybrid generations, receptive to pollen flow from later-colonizing white spruce. However, fixation of the Sitka spruce mitotype throughout the introgression zone suggests that gene flow may not have been bi-directional. Sutton *et al.* (1991; 1994) attributed haplotype structure to eastern movement of seed and pollen from prevailing winds over multiple generations. However, numerous cases of ‘chloroplast capture’, where the chloroplast is maternally inherited, have been described in the literature, in which the native cytoplasm is retained at the cost of the invasive cytoplasm (Rieseberg and Soltis, 1991; Freeman *et al.*, 2001). Sitka mitotype retention may result from a ‘capture’ of the local mitotype rather than the invading (white spruce) mitotype. This, in addition to asynchrony in reproductive phenology may contribute to patterns observed.

Prevailing wind directions (Sutton *et al.*, 1994) suggest easterly movement of seed and pollen will be more frequent than westerly dispersal the introgression zone. Given this, we expected a greater frequency of Sitka spruce cpDNA, due to greater pollen movement, towards the eastern end of the introgression zone than white spruce cpDNA at the western end. However, chloroplast haplotypes were geographically structured, suggesting that pollen may not be exchanged widely across the entire range examined. Asynchrony in reproductive phenology may provide a barrier to reproduction between the two species. Reproductive phenology is a highly heritable trait in a common garden determined by environmental cues, including heat sum requirements, in forest trees (Hanninen and Tanino, 2011). Therefore, heritable genetic variation may play a role in reproductive barriers across the hybrid zone. However, environmental variation contributing to heat sum accumulation between maritime and continental climates, reflected in degree days over 5°C (Table 2.1), may also result in interspecific variance in reproductive phenology, contributing to unidirectional patterns of gene flow preserved in organelle genomes.

2.4.2 Nuclear genetic structure

Contemporary genetic variation assessed using putatively neutral microsatellite markers reveals a range of values in both parent species and hybrid populations. Our results are consistent with previous studies based on expressed sequence tag markers and SNP

polymorphisms within candidate genes (Bennuah *et al.*, 2004). Our global estimates of genetic differentiation ($F_{ST}=0.05$) indicate that gene flow is likely widespread between the two species. Interspecific gene flow where barriers between species are porous has been implicated in the range of heterozygosity observed between parent and hybrid populations in *Populus* (Lexer *et al.*, 2005). Consistent with porous species barriers, heterozygosity estimates between Sitka spruce, white spruce and hybrid populations indicate that hybrids on average have higher heterozygosity than Sitka spruce parent species populations, although not higher than white spruce (Table 2.2). While there may be negative consequences of hybridization resulting from increased heterozygosity in hybrids (i.e. Dobzhansky-Muller incompatibilities), there may be advantages, including an increased capacity to adapt to changing conditions (Rieseberg and Wendel, 1993; Aitken *et al.*, 2008; Moehring, 2011). Combined with historical recolonization estimates of Sitka spruce and white spruce into the Nass-Skeena region, these data suggest that introgressants are likely the products of long-term introgression, leading to the formation of a stable hybrid zone spanning the ecotone between maritime and continental climates. Finally, excess Sitka spruce ancestry is congruent with unidirectional gene flow patterns preserved in both organelle genomes and previous observation of asymmetric introgression in this zone (Hamilton *et al.* submitted).

2.4.3 Climate and distance relationships with ancestry and morphology

Strong clinal variation in microsatellite-based hybrid index and morphology indicates that the climatic gradient spanning from maritime to continental climates across the introgression zone strongly influences genetic structure (Figure 2.4). Climatic and spatial variation in the genetic architecture of the hybrid zone is consistent with independent studies using a small number of neutral genetic markers (Bennuah *et al.*, 2004) and putatively adaptive SNPs from candidate genes (Hamilton *et al.* submitted).

The transition from maritime to continental climates, corresponding to increasing distance along the valleys of the Nass and Skeena rivers from the ocean, appears to strongly influence the genetic architecture of the hybrid zone. With increasing drainage distance up the valleys of the Nass and Skeena rivers the transition from maritime to continental climate appears to be a pervasive force influencing the genetic architecture of the hybrid zone. Furthermore, the strong clinal gradients in temperature, precipitation, and degree days less than 0°C indicating freezing period, accounting for much of the climatic variation observed across the hybrid zone, are all correlated with increasing drainage distance. While mean annual precipitation explains

much of the variation in ancestry observed across all individuals ($R^2=0.66$), drainage distance explains a near equal amount of variation ($R^2=0.62$) reflective of the cumulative transition from maritime to continental climate. Significant isolation-by-distance indicates that much of the differentiation between types of populations comes from differentiation of hybrid populations from reference parent populations reflected in increasing geographic distances (NP, SP, Figure 2.3B).

A steep inflection in hybrid index is observed at approximately 1000 mm of mean annual precipitation (Figure 2.4). This steep cline may indicate that water availability is a strong selective agent within the hybrid zone, reflecting an extrinsic barrier to gene flow. Steep clines in ancestry associated with mean annual precipitation were also observed for candidate gene SNPs by Hamilton *et al.* (submitted). We posited that there may be an abrupt change in selection pressures at a threshold around 1,000 mm of annual precipitation, preventing much introgression of Sitka spruce alleles further inland, and that white spruce alleles may have a selective advantage in drier climates. Indeed, variation in water use efficiency between parent and hybrids classes indicate that white spruce may exhibit greater fitness under low moisture conditions, while hybrids are intermediate (Silim *et al.*, 2001). These data support the bounded hybrid superiority model of hybrid zone maintenance across the ecological transition, identifying a strong role for extrinsic selection in which parent species are locally adapted to their native ecological condition, and hybrids persist within the transitional climate. However, current climatic patterns for this region and climate predictions into the future indicate this region is likely to become both warmer and wetter (Mbogga *et al.*, 2010). Consequently, precipitation may become less of a limiting factor, and the movement of Sitka spruce further inland may result in a geographic shift of the hybrid zone in response to changing climates (Buggs, 2007).

Morphological variation across the introgression zone was similarly associated with climate and distance variables, although to a lesser extent compared to genetic ancestry (Table 2.3, Figure S1). Consistent with previous research, differentiation was observed in the width and length of both scales and bracts, where Sitka spruce cone scales and bracts are longer on average, and white spruce cones and bracts are wider (Roche, 1969). We predicted that the size of cone scales and bracts were likely neutral traits, and would exhibit similar clines to neutral genetic markers. Although we observe similar clinal associations with climate and distance variables, cone morphology may not be a neutral trait. Particularly with steeper clines at 1000 mm of annual precipitation, these patterns suggest that extrinsic selection may influence morphological structure, where morphology is adaptive, or that genes controlling cone morphology are linked to

loci under selection (Figure S1). Morphological traits related to cone shape may be an adaptive trait contributing to species-specific pollination (Niklas and Tha Paw U, 1982; Niklas, 1984). Wind tunnel experiments in *Pinus spp.* reveal that pollen has a higher probability of reaching species-specific ovulate cones due to cone-bract architecture (Niklas and Tha Paw U, 1982). This questions the selective neutrality of these traits, and their comparison with genetic ancestry based on putatively neutral microsatellite markers.

2.4.4 Replicate genetic and morphological structure

Natural hybrid zones are commonly observed along climatic gradients, across which hybrids may exhibit increased fitness relative to parents, supporting a ‘bounded hybrid superiority’ model of hybrid zone maintenance (Moore, 1977; Benuah *et al.*, 2004; Yanchukov *et al.*, 2006; Culumber *et al.*, 2011). While reciprocal transplants may provide a more direct approach to examining local adaptation (Miglia *et al.*, 2005), examining replicate transects may identify common extrinsic selective pressures influencing the architecture of the introgression zone. A comparison of independent transects along the Nass and Skeena rivers where sampling overlapped (Table 2.4, Figure S2) revealed similar associations with both climate and distance variables. Replicated genetic and morphological clines across these independent transects provide further evidence of the role of environmental selection along these river valleys in hybrid zone maintenance. Ancestry estimates exhibit steep clines associated with precipitation and these clines appear concordant along independent transects. Significant differences between river valleys were only observed comparing ancestry across drainage distance, mean annual temperature, and degree days less than 0°C (Table 2.4), which may result from slightly steeper gradients within the Nass transect for select climatic variables.

We cannot exclude possible influence of intrinsic selection resulting in genetic incompatibilities or decreased hybrid fitness within this introgression zone. However, low global estimates of genetic differentiation, combined with strong climate associations for putatively neutral genetic and morphological characters support a strong role for extrinsic selection in the maintenance of the hybrid zone. Furthermore, anecdotal evidence from experimental crosses suggests that hybrids are readily formed (J. King pers. comm.). Advanced generation hybrid classes may exhibit hybrid breakdown, but this remains untested experimentally to our knowledge.

This study implicates an important role for extrinsic selection in hybrid zone maintenance. Historic patterns of haplotype structure suggest the sequence of arrival of parental

species. Combined with evidence for predominantly unidirectional gene flow that likely results from geographic variation in degree days for chilling and heat sum accumulation resulting in asynchronous reproductive phenology. Furthermore, contemporary patterns of nuclear and morphological variation provide consistent associations with climate and distance variables, replicated across independent transects. Current seed transfer guidelines within British Columbia are based on climatic associations observed from provenance trials or observed climatic variation and its influence on ecosystems, with movement primarily confined to latitudinal or elevational transfer. The results from this study identify an important role for climatic associations with respect to natural variation, however, they include an east-west longitudinal climatic gradient up the Nass and Skeena river valleys, emphasizing precipitation as a strong selective agent within this ecological transition zone. These findings suggest that identifying seed sources with appropriate climatic associations and directing seed transfer in the future should include longitudinal east-west climatic gradients. In addition, this study points toward the importance of water availability within the Nass-Skeena region. Precipitation, a strong selective agent influencing both genetic and morphological structure within the hybrid zone, may need to be incorporated into prescriptions regarding seed transfer within this transitional area into the future.

Table 2.1 Origins of Sitka x white spruce (*Picea sitchensis* x *P. glauca*) study populations along the Nass and Skeena rivers, sample s individuals analyzed for chloroplast (cpDNA) and mitochondrial (mtDNA) variation, associated geographic and climatic variable number of cones (N) sampled per population for morphological analysis and mean principal component values per population descri

Location	River	N (cpDNA/mtDNA)	Latitude	Longitude	Elevation (m)	Drainage Distance (km)	MAT (°C)	MAP (mm)	DD < 0 °C	DD5 ₁₀₀	Q-index
Port Edward	Skeena	26 (6/10)	54.18	130.18	27	0	8.3	3061	1	116	0.85
Digby Island	Skeena	26 (-/15)	54.28	130.38	11	0	7	4917	38	123	0.85
Tuck Inlet	Skeena	25 (-/10)	54.41	130.28	11	12	8	3155	16	120	0.92
Tyee Cannery	Skeena	25 (7/9)	54.20	129.97	7	15	7.7	3252	52	121	0.90
Kwinitsa	Skeena	25 (-/10)	54.23	129.53	10	15	6.4	3215	280	124	0.85
Exchamskis	Skeena	25 (8/10)	54.34	129.27	28	69	5.4	2495	448	134	0.86
Shames	Skeena	25 (-/10)	54.41	128.89	50	97	5.7	1398	422	126	0.80
Copper River	Skeena	25 (10/10)	54.49	128.33	138	136	5.1	1778	501	138	0.78
Pitman	Skeena	25 (-/10)	54.71	128.31	112	156	5.1	1333	504	132	0.75
Flint Creek	Skeena	25 (9/10)	54.96	128.39	83	182	5.8	837	446	124	0.45
Kitwanga	Skeena	25 (10/10)	55.10	128.07	178	212	5.4	762	586	123	0.21
Skeena Crossir	Skeena	25 (10/10)	55.11	127.79	210	232	5	605	663	125	0.26
Bulkley	Skeena	26 (10/10)	55.10	127.37	441	285	3.8	566	812	135	0.10
Smithers	Skeena	29 (29/15)	54.85	127.22	511	326	3.6	1247	871	136	0.10
Gingolx	Nass	25 (-/12)	55.00	129.98	3	51	6.9	2772	150	125	0.80
Greenville	Nass	29 (-/11)	55.06	129.54	24	83	6.1	1869	317	126	0.77
Lava	Nass	25 (-/10)	55.19	129.22	27	83	5.9	1078	391	125	0.77
Nass	Nass	25 (4/13)	55.41	129.06	133	144	5.7	1073	439	127	0.60
Cranberry	Nass	25 (-/10)	55.53	128.76	167	169	5.6	1129	474	126	0.49
Meziadin	Nass	25 (-/14)	56.09	129.29	271	235	5	909	566	131	0.27
VanDyke	Nass	25 (-/12)	55.82	128.93	218	196	5.2	700	565	127	0.19
Haida Gwaii ^a	-	25 (9/12)	53.25	132.11	5	0	7.9	2450	28	127	1.00
Fort Nelson ^b	-	21 (11/12)	58.32	122.50	350	1034	0.9	548	2046	138	0.00

^a reference Sitka spruce parent population

^b reference white spruce parent population

MAT, mean annual temperature; MAP, mean annual precipitation; DD, degree days; DD5₁₀₀, Julian date of budburst for most plants

Table 2.2 Genetic variability at 12 microsatellite loci in *Picea sitchensis*, *P. glauca* and their hybrids, including locus name, annealing temperature (T_m) repeat type, average F_{IS} , F_{ST} and R_{ST} across all individuals, F_{ST}^{ENA} adjusted for presence of null alleles, number of alleles (A), variance in allele size (Var), expected heterozygosity (H_E), and observed heterozygosity (H_O) for each locus.

Locus	T_m	Repeat Type	F_{IS}	F_{ST}/R_{ST}	F_{ST}^{ENA}	<i>P. sitchensis</i>				<i>P. glauca</i>				<i>P. sitchensis</i> x <i>P. glauca</i>			
						A	Var	H_E	H_O	A	Var	H_E	H_O	A	Var	H_E	H_O
WS0011.P12	53	4	0.00	0.11 / 0.07	0.11	3	6.9	0.286	0.227	2	3.76	0.470	0.429	4	3.9	0.313	0.317
WS0061.K02	53	2	0.06	0.02 / 0.02	0.03	2	0.9	0.481	0.250	2	0.3	0.136	0.143	4	1.8	0.478	0.459
PAAC17	53	2	0.48	0.10 / 0.13	0.09	4	8.2	0.319	0.235	5	4.1	0.662	0.385*	8	8.5	0.564	0.297**
SPAGG3 ^a	53	2	0.23	0.03 / 0.14	0.02	8	71.3	0.802	0.389**	11	35.7	0.872	0.700*	20	66.9	0.878	0.692**
WS0092.M15	53	3	-0.11	0.03 / 0.02	0.03	3	2.11	0.332	0.389	1	0	0.000	0.000	4	2.1	0.309	0.343
WS0092.A19	53	2	0.18	0.05 / 0.18	0.05	5	5.8	0.747	0.579	5	5.3	0.717	0.714	8	6	0.727	0.592**
WS0046.M11	53	3	0.00	0.14 / 0.11	0.14	1	0	0.000	0.000	3	2.7	0.528	0.714	4	2.7	0.296	0.287
UAPgAG150.1	54	2	-0.05	0.06 / 0.04	0.06	1	0	0.000	0.000	2	3.7	0.203	0.111	2	2.8	0.157	0.168
UAPgAG150.2	54	2	0.33	0.04 / 0.04	0.04	3	1.2	0.521	0.429	5	5.1	0.636	0.765*	10	3.3	0.607	0.390**
WS00716.F13	53	2	0.23	0.01 / 0.05	0.01	14	57.3	0.932	0.692**	10	48.6	0.821	0.722	19	52.9	0.887	0.681**
WS0022.B15	53	2	0.25	0.01 / 0.00	0.01	7	9.9	0.853	0.800	13	118	0.933	0.600**	15	38.8	0.882	0.669**
UAPgCT3	60	2	0.46	0.05 / 0.04	0.05	7	13.8	0.784	0.647	5	70.3	0.822	1.000	14	24	0.801	0.401**

* $P < 0.05$, ** $P < 0.005$, exact tests for Hardy-Weinberg equilibrium across all populations

^a Excluded from further analysis

Table 2.3 - Summary of clinal variation across the Sitka spruce and white spruce hybrid zone for hybrid index (*q-index*) and first two principal components accounting for morphological variation using all individuals. Type of regression, sign of slope, R^2 , and statistical significance are provided for each regression.

Geoclimatic variable	Q-index				Morphological PC1 (Length)				Morphological PC2 (Width)			
	Regression	Sign	R^2	P	Regression	Sign	R^2	P	Regression	Sign	R^2	P
Drainage Distance (km)	Linear	-	0.62	<0.001	Linear	-	0.14	<0.001	Linear	+	0.29	<0.001
MAT (°C)	Inverse	-	0.42	<0.001	Linear	-	0.07	<0.001	Inverse	+	0.21	<0.001
MAP (mm)	Inverse	-	0.66	<0.001	Inverse	-	0.09	<0.001	Inverse	+	0.30	<0.001
DD < 0°C	Linear	-	0.51	<0.001	Linear	-	0.11	<0.001	Linear	+	0.21	<0.001

MAT, mean annual temperature; MAP, mean annual precipitation; DD, degree days

Table 2.4 - Summary of clinal comparison between Skeena river and Nass river populations for hybrid index (Q-index) and first two principal components accounting for morphological variation. Slope and intercept of regressions following regression analysis used in Table 3, along with F-value and Pr(<F) for an ANCOVA testing differences between the two rivers with predictor climate variables and dependent variables.

Geoclimatic variable	Q-index				Predictor:River	
	Skeena		Nass			
	Slope	Intercept	Slope	Intercept	F-value	Pr (>F)
Drainage Distance (km)	-0.004	1.14	-0.004	1.06	8.58	<0.01
MAT (°C)	-6.501	1.78	-12.240	2.700	28.06	<0.01
MAP (mm)	-653.940	1.176	-600.459	1.082	0.22	0.64
DD < 0°C	-0.001	1.357	-0.001	1.172	15.53	<0.01
Geoclimatic variable	Morphological PC1				Predictor:River	
	Skeena		Nass			
	Slope	Intercept	Slope	Intercept	F-value	Pr (>F)
Drainage Distance (km)	-0.001	0.293	-0.001	0.109	0.05	0.82
MAT (°C)	0.084	-0.421	0.117	-0.697	1.26	0.26
MAP (mm)	-87.876	0.139	-269.954	0.222	0.90	0.34
DD < 0°C	0.000	0.404	0.000	0.217	1.00	0.32
Geoclimatic variable	Morphological PC2				Predictor:River	
	Skeena		Nass			
	Slope	Intercept	Slope	Intercept	F-value	Pr (>F)
Drainage Distance (km)	0.001	-0.218	0.001	-0.208	0.68	0.41
MAT (°C)	2.31	-0.444	4.113	-0.75	1.66	0.20
MAP (mm)	206.312	-0.189	169.538	-0.175	1.760	0.19
DD < 0°C	0.001	-0.032	0.000	-0.207	0.460	0.50

MAT, mean annual temperature; MAP, mean annual precipitation; DD, degree days

Figure 2.1 Map of individuals sampled across the introgression zone between Sitka and white spruce in northwestern British Columbia along the Nass and Skeena rivers, indicating provenance of origin of samples (black circle) collected and frequency of chloroplast haplotypes (blue, Sitka haplotype; green, white haplotype) in each population. Inset map provides North American range of *P. sitchensis* (dark grey) and *P. glauca* (light grey).

Figure 2.2 Diagrammatic representation of spruce scale and bract showing six basic morphological measurements adapted from Roche (1969).

Figure 2.3 Relationship between pairwise estimates of genetic distance (based on $F_{ST}/1 - F_{ST}$) with pairwise geographic distance based on the great circle distance between populations. Pairwise distance comparisons between populations on the Nass river (NN), Skeena river (SS), between Nass and Skeena (NS) river populations were estimated (A), and Nass and parent (NP), Skeena and parent populations (SP), and between parent populations (PP) were compared separately (B).

Figure 2.4 - Relationship between microsatellite-based hybrid index (0=white spruce, 1=Sitka spruce) and geographic and climatic variables across 582 individuals spanning the Sitka-white introgression zone, including drainage distance (km), mean annual temperature ($^{\circ}\text{C}$), mean annual precipitation (mm), and degree days below 0°C .

3. Genomic and phenotypic architecture of a spruce hybrid zone (*Picea sitchensis* x *P. glauca*)

3.1 Introduction

Natural hybrid zones are important biological systems for examining the roles of natural selection and gene flow in the maintenance of reproductive isolation and species differences. Interspecific hybridization is an important source of novel genetic variation in which new genetic recombinants may gain a further capacity for adaptive evolution (Rieseberg and Wendel, 1993; Fritz *et al.*, 2006). This is particularly important in a changing climate, where there is growing evidence that some tree species may exhibit pronounced adaptational lag (Aitken *et al.*, 2008). Introgression, where allelic variants are exchanged between species permanently and repeatedly, may preserve a species' evolutionary potential and ability to adapt in the face of climate change and increasing environmental stochasticity (Rweyongeza *et al.*, 2007; Sapir *et al.*, 2007). Although introgression between congeneric forest tree species is widely documented (Lexer *et al.*, 2004), hybrid zones have traditionally remained a 'black box' in tree breeding programs, where the goal is to produce seed for reforestation optimally adapted to relatively local environments. Increased understanding of the mechanisms underlying local adaptation in forest tree hybrid zones may produce additional tools to mitigate the influence of climate change in managed forests.

Hybrid zone maintenance has a rich theoretical background that can be distilled into two types of models differing fundamentally in the mode of selection implicated: endogenous (non-environmental), or exogenous (environmental). The tension zone model described in detail by Barton and Hewitt (1985) describes a hybrid zone shaped by endogenous genetic factors independent of the environment. In contrast, the bounded hybrid superiority model (Moore, 1977; Arnold, 1997) applies to hybrid zones in which exogenous selection shapes variation within the contact zone, and the fitness of hybrid and parental genotypes differ among varying environments (Wilhelm and Hilbish, 1998). Furthermore, this model assumes that hybrids exhibit greater fitness relative to parental genotypes due to genotype-by-environment interactions within the ecologically transitional habitat (Moore, 1977). Although these models have clearly differentiated assumptions, clinal variation resulting from geographic variation in environmental selection gradients (Endler, 1977; Moore, 1977) or intrinsic selection against hybrids is often indistinguishable (Barton and Hewitt, 1985; Kruuk *et al.*, 1999). Teasing apart

the mechanism(s) that contribute to hybrid zone maintenance may help us to understand the role of environment and genetic factors in hybrid zone persistence.

While exogenous selection may play a significant role in hybrid zone maintenance, explanations based on one mode of selection may be an oversimplification. For example in *Artemisia tridentata*, Wang et al. (1997) observed highest fitness of hybrids within transitional habitat during the first two years of growth, supporting a bounded hybrid superiority model of hybrid zone persistence. However, in re-assessing the experiment nine years later, Miglia et al. (2005) found that early-generation patterns broke down, and although there were clear genotype-by-environment interactions, differential success of parent and hybrids also supported a genetic basis for hybrid fitness, implicating both endogenous and exogenous selection. Both first and advanced-generation hybrids exhibit a range of fitness values, and comparison among genotypic classes may help tease apart the influence of heterosis in early generation hybrids from hybrid breakdown or transgression in later-generation hybrids where environmental variation plays a less significant role (Rieseberg and Ellstrand, 1993; Rieseberg, 1995).

Several studies have used molecular markers to evaluate the frequency of hybrid formation and the spread of alleles in a natural setting, capitalizing on a growing wealth of genetic and genomic resources (Lexer *et al.*, 2005; Milne and Abbott, 2008; Field *et al.*, 2010; Gompert *et al.*, 2010). Evaluation of interspecific gene flow and the distribution of hybrid genotypes offers an understanding of the mechanisms, both extrinsic and intrinsic, that maintain distinct species (Field *et al.*, 2011). Genomewide scans of mapped loci provide increased resolution of potential locus-specific effects for maintaining species differences, while genomic clines of hybrid indices and ecological variables may quantify relationships maintaining the hybrid zone (Nolte *et al.*, 2009; Teeter *et al.*, 2010; Gompert and Buerkle, 2011). Finally, combining genetic and phenotypic data within a common garden allows for inference of both the extent and direction of gene flow, and ultimately aids in teasing apart genotype-by-environment interactions. This application will have broad ecological, conservation, and economic implications in forest tree hybrid zones, where identification of seed source populations for reforestation within transitional habitat remains problematic. Moreover, where admixture results in an increased capacity to respond to a changing climate (Aitken *et al.*, 2008), associating fitness proxies in long-lived tree species such as growth and cold hardiness (Howe *et al.*, 2003), with climatic variation may aid in identifying those genotypes suitable for replanting in areas where climatic conditions have or are predicted to change.

The introgression zone between Sitka spruce (*Picea sitchensis* (Bong) Carr.) and white spruce (*P. glauca* Moench Voss) spans maritime to continental climates along the Nass and Skeena river valleys in northern British Columbia. In British Columbia, Sitka spruce is a significant component of the Coastal Western Hemlock (CWH) ecosystem. This largest of spruce species thrives at a mean annual temperature of approximately 8°C and mean annual precipitation exceeding two metres (Pojar *et al.*, 1991). White spruce, a component of the Interior Cedar Hemlock (ICH) ecosystem characterized by cold, wet winters and warm, dry summers along with periodic extremes, comes into contact with Sitka spruce along river valleys where moist, cool climates offer transitional habitat between these coastal and inland ecosystems (Ketcheson *et al.*, 1991). Previous studies have characterized the zone of contact using morphology (Daubenmire, 1967; Roche, 1969), phenology (Roche, 1969, Benuah and Aitken, unpublished data), physiology (Grossnickle *et al.*, 1996; Fan *et al.*, 1997; Silim *et al.*, 2001), insect resistance (O'Neill *et al.*, 2002) and molecular genetic markers (Yeh and Arnott, 1986; Sutton *et al.*, 1994; Benuah *et al.*, 2004). Steep climatic gradients, from maritime to continental climates, have been implicated in the large degree of variation observed.

Although there is strong evidence for the role of exogenous selection maintaining this hybrid zone, fitness of parent and hybrid individuals have not been assessed. Benuah *et al.* (2004) described wide geographic clines in hybrid index across environmental gradients in the zone using a small number of expressed sequence tag markers, suggesting environmental selection strongly influences genetic composition. However, clinal variation may also arise from endogenous selection (Kruuk *et al.*, 1999) which was not examined within the Benuah *et al.* (2004) study. By capitalizing on genomic resources developed for both Sitka and white spruce, and a ten-year-old common garden experiment previously established within the hybrid zone, we aim to better understand the mechanisms underlying the maintenance of this hybrid zone.

In the current study, we have used a genomic survey of single nucleotide polymorphisms (SNPs) to examine the composition of individuals collected from across the contact zone as well as reference parental species, all planted in a common garden experiment located within the hybrid zone. Using genomic composition estimated by hybrid index, and interspecific recombination evaluated using heterozygosity, we first asked whether the direction and extent of introgression across the contact zone reflects a bounded hybrid superiority or tension zone model of hybrid zone maintenance. We then compared associations between hybrid index and both climate variables, indicative of exogenous selection, and distance variables, indicating a pattern of isolation-by-distance, to infer potential mechanisms influencing genetic structure across this

zone. Finally, we tested for differences in phenotypic traits across genotypic classes to gain further support for the presence or absence of hybrid superiority. This research will provide insight into the role of intrinsic and extrinsic factors in the maintenance of this hybrid zone, and inform seed transfer guidelines for current and future climates for forest trees in ecologically transitional areas.

3.2 Methods

3.2.1 *Plant material and common garden experiment*

Open-pollinated seeds were collected across 29 locations throughout the introgression zone, along the Nass and Skeena River drainage basins, in 1997 by the British Columbia Ministry of Forests, Range and Natural Resource Operations. Two additional allopatric populations, Sitka spruce on Haida Gwaii, British Columbia (HG) and white spruce from the Ottawa valley region, Ontario (ENA, Table 3.1 and Figure 3.1) were also included in the study as reference parental populations. Cones were collected from the upper canopy of three to seven seed parents per location by helicopter. Open-pollinated progeny from each seed parent were grown for one year in a nursery, and ten seedlings per family were subsequently transplanted into a common garden near Kitwanga, British Columbia (55°17'N, 128°10'W, Figure 3.1). A split-plot block design was used for the common garden with 10 reps, where the main plots were regional sets and the subplots were sampling locations within regions. Subplots were noncontiguous and populations were randomized throughout the main plot. On average, six progeny were sampled per open-pollinated family, for a total of 721 trees from sampling locations spanning the contact zone, with on average six progeny per open-pollinated family. Trees were measured for ten-year height, tested for cold hardiness, and sampled for DNA extraction and genotyping.

3.2.2 *Artificial freeze tests*

Branches were harvested in late August 2010 from the common garden trial in Kitwanga, BC for artificial freeze testing using methods described in detail by Hannerz *et al.* (1999). Five ~ 5mm needle segments per individual branch were cut and placed in a tube with 0.2 ml of distilled water with a tiny amount of silver iodide to facilitate ice nucleation. Samples were frozen at three temperatures, -8, -18 and -28°C in order to obtain intermediate injury levels for at least one temperature. Chamber temperature was lowered at a rate of 4°C/h and then held for

one hour at each test temperature before removal. Control samples were kept at 4°C for the duration. After freezing, 3.5mL of distilled water was added to each sample and they were refrigerated for 24 hours. Samples were then shaken for 60 minutes prior to initial measurement of electrolytic conductivity, followed by a heat-kill at 95°C for 1.5 hours, after which they were refrigerated for 24 hours, shaken, and conductivity re-measured. Flint's Index of Injury I_t (Flint *et al.*, 1967) was calculated as a percentage of injury as follows:

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

where $R_t = L_t/L_k$, $R_o = L_o/L_d$, estimated from a ratio of conductivity between frozen samples preceding (L_t) and following heat kill (L_k) with unfrozen control samples preceding (L_o) and following heat kill (L_d). Damage levels at -28°C were too high to be informative, therefore this temperature was dropped from further analyses.

3.2.3 SNP selection and genotyping

Newly-flushed lateral shoots were collected in June 2009, flash frozen in the field and transported in liquid nitrogen for genetic analysis. Fifty mg of tissue was ground to a fine powder using a Mixer Mill MM 400 (Retsch) and DNA was extracted using a modified CTAB protocol (Doyle and Doyle, 1990). Pure Sitka (N=1088), white (N=40), and Engelmann spruce (*Picea engelmannii*, N=40) were genotyped for 1536 SNPs identified as putative candidates for roles in budset timing or cold-hardiness development (Holliday *et al.*, 2008; Holliday *et al.*, 2010a), white pine shoot tip weevil (*Pissodes strobii*) resistance (K. Ritland *et al.* unpublished data), and growth (J. Bousquet *et al.* unpublished data). A subset of 384 SNPs was selected from this study based on the degree of interspecific genetic differentiation ($F_{ST} \geq 0.20$) among the three spruce species. These spruces have a relatively recent common ancestor, therefore the likelihood of finding diagnostic alleles to develop hybrid indices given phylogeny (Ran *et al.*, 2006) and history (Anderson *et al.*, 2006; Mimura and Aitken, 2007) was unlikely. Consequently, genetic differentiation (F_{ST}) was used as a tool for selecting SNPs for development of hybrid indices where high allele frequency differentials between species contain more information about ancestry (Buerkle, 2005). The selected SNPs are likely subject to some ascertainment bias in terms of underrepresentation of rare alleles because of the way in which the SNPs were selected (Namroud *et al.*, 2008).

Selected SNPs were genotyped using the Illumina bead array platform in conjunction with the GoldenGate allele-specific assay in a 96-well, 384-SNP format (Shen *et al.*, 2005; Fan

et al., 2006). This highly multiplexed assay hybridizes two oligos specific to the alleles at the SNP site along with a locus-specific third oligo to genomic DNA downstream from the SNP. The primers used in PCR product generation contain fluorescent moieties, which differentiate homozygotes and heterozygotes based on fluorescent signals. Genotyping quality was examined using the Genomestudio Genotyping Module (v1.0) for successfully genotyped SNPs (N=338). Quality was assessed using a number of criteria, including call rates greater than 90%, a minimum 'GenTrain' score of 0.40, visual assessment of differentiation of homozygote and heterozygote clusters, and minimum intensity threshold for fluorescent signals from replicate beads (>1000). SNPs were ranked based on five categories (1) failed; (2) not polymorphic; (3) polymorphic, but low genotyping quality; (4) clear genotyping, but heterozygote excess, heterozygote deficiency, or lack of clustering of homozygous genotypes and (5) high-quality genotyping. Of the initial 384 SNPs, 268 met the high quality standard of (5) for subsequent analysis.

3.2.4 Genomic admixture analysis and hybrid identification

To estimate admixture among individuals, we carried out analyses using two different approaches, a maximum likelihood estimate of hybrid index calculated using the program *hindex* (Buerkle, 2005), and a Bayesian clustering approach implemented in *Structure* version 2 (Pritchard *et al.*, 2000). *Hindex* requires *a priori* knowledge of pure parental individuals, here identified as allopatric populations – HG (Sitka spruce) and ENA (white spruce) to estimate the proportion of ancestry for either parent, where a hybrid index of zero corresponds to pure white spruce and one corresponds to pure Sitka spruce. *Structure*, however, estimates a posterior probability (q) of assignment to k genotypic cluster categories rather than *a priori* assignments of pure species. Multilocus genotypes describing the proportion of individual genotype assignment to k clusters, with the assumption of $k = 2$ representing two species, were used to estimate the proportion of individual parent species' contribution to the admixed gene pool. *Structure* runs were carried out using the admixture model assuming independent allele frequencies with a burn-in step of 50 000, followed by 100 000 Markov chain Monte Carlo iterations. Five replicate runs of models using genotypic clusters (k) from one to 31 confirmed $k = 2$ had the highest log-likelihood probability (results not shown).

Using *hindex* and *Structure* provides additional confidence in assignment of genotypic classes. Pearson correlations between hybrid index and q -value was performed to demonstrate that *a priori* assignment of pure parental species did not prejudice the calculation of ancestry.

ENA individuals used to represent pure white spruce are from populations thousands of kilometres from the hybrid zone that have no possibility of introgression with *P. engelmannii*, which hybridizes extensively with white spruce in mountainous areas of southeastern British Columbia and southwestern Alberta. To confirm that ENA samples were representative of white spruce found in British Columbia, a subset of 120 loci genotyped both in ENA and in white spruce individuals from Fort Nelson, British Columbia (58° 49' 59", 123° 27' 0"), were further compared using *Structure*. Those results indicated a high posterior probability (q , where 0 is white spruce) that individual ENA and Fort Nelson genotypes originated from the same genotypic cluster (average ENA = 0.01 ± 0.005 , average Fort Nelson = 0.006 ± 0.001).

Interspecific heterozygosity was estimated in the R-package *Introgress* (Gompert and Buerkle, 2010) for each individual and compared with hybrid index to provide a means to categorize individuals into genotypic classes. Expectations follow Milne and Abbott (2008), where F1 individuals have both 100% interspecific heterozygosity and a hybrid index of 0.5 in which each species contributes equally to the genomewide estimate of admixture. Furthermore, due to the complex and overlapping nature of hybrid classes, e.g. an F2 may resemble an F3 (Milne and Abbott, 2008), classes are broadly defined as backcrosses to either parent (BCS, BCW) if they exhibited less than 100% proportion, but greater than nearly 75% ancestry from one parent, or as advanced-generation hybrids (FN) if ancestry proportions remained relatively equal between species, but interspecific heterozygosity decreased below 85%. In effect, classification was achieved based on genomic discontinuities visible in two-dimensional hybrid index-heterozygosity space for all SNPs (Figure 3.2A). This analysis is similar to that implemented in *NewHybrids* (Anderson and Thompson, 2002) but is more robust distinguishing pure parentals from advanced-generation introgressants, and is prone to fewer assumptions than the Bayesian approach of *NewHybrids*. Although there is always still some uncertainty in class assignment with advanced generations, the use of broad categories for each class allows for assignment of those more complex derivatives with reasonable confidence (Milne and Abbott, 2008).

A lack of diagnostic species-specific markers present between spruce species results in higher estimates of interspecific heterozygosity for pure parent species than expected. We compared interspecific heterozygosity further, evaluating a subset of loci with an allele frequency differential greater than 0.90 ($n=31$, Figure 3.2B). Comparison of interspecific heterozygosity values for these strongly differentiated loci with all 268 SNPs is useful to

determine whether allele segregation contributing to interspecific heterozygosity values are due to ancestral polymorphisms or are more consistent with long-term interspecific recombination.

3.2.5 Quantitative genetic data analysis

To better understand phenotypic variation in height and cold injury across the hybrid zone, an analysis of variance was conducted using the linear model:

$$y_{ijkl} = u + r_i + s_j + r_i \times s_j + p_k(s_j) + r_i \times p_j(s_j) + f(p(s_j)_{kl})$$

where r is the effect of rep i , s is the effect of set j , p_k is the effect of population k and $f(p)_{kl}$ is the effect of family l within population k . This full model was reduced to:

$$y_{ijk} = u + r_i + p_j + r_i \times p_j + f(p)_{jk}$$

where r is the effect of rep i , p_j is the effect of population j and $f(p)_{jk}$ is the effect of family k within population j , as all terms involving regional set were not significant and had no effect on the model. Additive genetic variance and narrow sense heritability was estimated for each trait. Quantitative genetic differentiation among populations (Q_{ST}) was estimated using the relationship described in Spitze (1993) using PROC VARCOMP in SAS with the restricted maximum likelihood option (METHOD=REML).

3.2.6 Statistical analysis

Spatial structure of relationships between introgression and the environment were examined using regressions of both climate and distance variables for provenances on genotypes (hybrid index) and phenotypes. The geographic origin of individuals, described by latitude, longitude and elevation were input into *ClimateBC* v.3.2 (Wang *et al.*, 2006a) which estimates seasonal and annual climate variables based on geographic coordinates. Non-linear physical drainage distance up river systems from the Pacific Ocean was estimated using *ArcGIS* (Version 10), replicating methods from Bennuah *et al.* (2004). Distance from the ocean in this region is strongly associated with climatic gradients from maritime to continental climates. Both linear and non-linear regressions were estimated. As non-linear clines are common across hybrid zones (Arnold, 1997) the use of non-linear regression may be more appropriate for the current analysis. In addition, multiple regressions were performed to evaluate the relationship between geographic and climatic variables with genomewide estimate of ancestry and putatively adaptive phenotypic traits.

Height and cold injury phenotypes were analyzed as components of fitness, as conventional lifetime fitness measures are not possible in such long-lived organisms. Response

variables were analyzed using ANOVA to compare the means among genotypic classes, followed by a post-hoc analysis using Tukey's HSD (honest significant difference) test for multiple comparisons. These tests were accompanied by a non-parametric Kruskal-Wallis test, which does not assume a normal distribution.

3.3 Results

Analysis of admixture using both *hindex* and *Structure* indicate a high proportion of hybrid individuals and an asymmetry towards Sitka spruce. Of the 721 individuals analyzed, *Structure* identified 345 individuals with a *q*-value greater than or equal to 0.9, 35 less than or equal to 0.1, and 346 between 0.1 and 0.9, where a *q*-value equal to one indicates pure Sitka spruce. Hybrid index values estimated from *hindex* also showed an excess of ancestry toward Sitka spruce. The high correlation between ancestry estimates from *Structure* and *hindex* ($r = 0.997$, $p < 0.001$), even though *hindex* requires *a priori* identification of pure parents, indicates that hybrid index estimates do not depend on pure parental samples being identified *a priori*.

Low overall interspecific heterozygosity across all individuals (0.06 – 0.49) indicates that this hybrid zone is largely comprised of advanced generation recombinants, where weak or no barriers to reproduction between species exist. Individuals were assigned to genotypic classes by comparing hybrid index and interspecific heterozygosity jointly (Figure 3.2A). The complete absence of an F1 generation (with 100% heterozygosity and intermediate hybrid index) is consistent with an advanced-generation hybrid zone and a lack of reproductive barriers between species. Although low interspecific heterozygosity suggests widespread recombination, the distribution of the majority of hybrid index estimates over 0.5 suggests that introgression is skewed towards Sitka spruce (Figure 3.2A). This asymmetry is reflected in the distribution of genotypic classes, where only 15 individuals were classified as backcrosses towards white spruce (BCW), compared to 139 individuals classified as advanced-generation hybrid class (FN) and 501 individuals classified as backcrosses to Sitka spruce (BCS) based on *hindex*. Evaluation of a subset of 31 loci with an allele frequency differential greater than 0.90 produced lower values of interspecific heterozygosity for white spruce (zero to 0.35, mean=0.05) and Sitka spruce (zero to 0.13, mean=0.03, Figure 3.2B) when compared to results for 268 loci. Patterns were consistent with results using the full set of markers, indicating that estimates of interspecific heterozygosity for pure species, although higher than expected, were consistent

with widespread recombination and allele-sharing, rather than divergent ancestral polymorphisms.

Narrow-sense heritabilities across all traits were typical for these traits in conifers, estimated as 0.16 for cold injury at -8°C, 0.17 at -18°C and 0.17 for height. Quantitative genetic differentiation (Q_{ST}) among populations was moderate to low, at 0.35 for cold injury at -8°C, 0.38 at -18°C and 0.13 for height.

Regressions of both hybrid index and phenotypic traits on distance and climatic variables revealed strong geographic and climatic clines for hybrid index but weak clines for phenotypic traits (Table 3.2). Clines of hybrid index with independent variables other than mean annual temperature (MAT) were non-linear (Figure 3.3). A quadratic regression for hybrid index on both drainage distance and continentality, and an inverse transformation for mean annual precipitation (MAP) produced the best model fit. Drainage distance up rivers from the coast exhibited a strongly significant non-linear relationship with hybrid index with increasing distance from the coast to the interior (quadratic regression: $R^2 = 0.73$, $p < 0.001$, Figure 3.3). This relationship was mirrored in the cline for MAP, which appeared to have a threshold pattern with a steep inflection observed at 1000 mm, above which hybrid index estimates remained close to one (Sitka spruce) and below which a steep decline in hybrid index was associated with decreased precipitation.

Significant clines were also observed for MAT and continentality (difference between mean warmest month and mean coldest month temperatures). While clinal variation in continentality explains a large proportion of the variance in hybrid index (quadratic regression: $R^2 = 0.41$, $p < 0.001$), MAT is only weakly associated with hybrid index ($R^2 = 0.28$, $p < 0.001$). Although hybrid index exhibited significant clinal variation with respect to some climate variables, drainage distance overall explained more variation than any one climate variable. As drainage distance is highly correlated with several climatic variables (MAT $r = -0.80$, MAP $r = -0.80$, continentality $r = 0.89$, $p < 0.001$ all), both climate and drainage distance variables reflect the transition between maritime and continental climates characteristic of the introgression zone. Multiple regression of hybrid index with drainage and climate variables (MAT, MAP, and continentality) indicated all variables were significantly related to hybrid index ($R^2 = 0.68$, $p < 0.001$), although drainage distance explained the most variance (F-value=1344.06, $p < 0.001$).

Cold injury exhibited only shallow linear clines along climatic gradients and geographic variables, and somewhat surprisingly, there was no clinal variation detected for height (Table 3.2). Drainage distance was significantly associated with cold injury ($p < 0.001$), but explained

just five to seven percent of the variation in cold injury among individuals. For all phenotypic traits stepwise multiple regression with distance and climate variables did not improve the relationship observed over univariate regressions.

Comparison of phenotypic characters across all genotypic classes using ANOVA with post-hoc tests for multiple comparisons, along with a non-parametric equivalent, Kruskal-Wallis, indicated means differ significantly among genotypic classes (Figure 3.4, Table 3.3). However, pairwise comparisons between genotypic classes provided a clearer view of differentiation (Table 3.3). For cold injury at -8°C , all individuals with membership in a hybrid class appeared to have greater cold hardiness than either pure parental species (Figure 3.4). Multiple comparisons indicated that hybrid classes were not significantly different from each other, but were significantly different from both parental species classes, and Sitka and white spruce classes were not significantly different from each other (Table 3.3). However, this pattern was not observed at the colder test temperatures. Overall, at -18°C , pure Sitka exhibited the greatest cold injury, and as the contribution of Sitka spruce decreased and white spruce increased, there was increased cold hardiness (Figure 3.3). The BCW class appeared slightly more cold tolerant on average than white spruce at -18°C , although not significantly different from white spruce (Table 3.3). Surprisingly, although there were significant differences among genotypic classes in height (Table 3.3), there were no significant pairwise comparisons. This suggests that the differences observed are largely due to the lower height observed in the BCW class (Figure 3.3).

3.4 Discussion

The genomic architecture of the Sitka-white spruce zone of introgression revealed some surprising paradoxes. Low overall interspecific heterozygosity based on all SNPs, ranging from 0.06-0.49, indicates an advanced-generation hybrid zone in which recombination is widespread. However, asymmetric introgression toward Sitka spruce suggests pre- or post-zygotic barriers may influence structure within this zone. Variation in the proportion of ancestry across the zone appears to be strongly influenced by both geography and climate, indicating an important role for exogenous selection shaping the genomic architecture of the zone. Surprisingly, a large proportion of phenotypic variation within the common garden remained unexplained by provenance geography or climate. Phenotypic variability compared across genotypic classes indicates that hybrid classes are transgressive for cold hardiness at minimally damaging cold

temperatures, but at lower temperatures, lower cold hardiness in more maritime populations and greater hardiness in more continental populations fit predictions.

3.4.1 *Historic introgression*

Low levels of interspecific heterozygosity combined with a high frequency of hybrids suggest this hybrid zone is the product of both long-term contact and widespread recombination between the two species (Figure 3.2). Sitka and white spruce are closely related (Ran *et al.*, 2006) and their current distribution and structure reflect events during the Pliocene epoch and subsequent expansion and contraction of ice sheets during the Pleistocene (Morgenstern, 1996). Recolonization history suggests these species came into contact most recently following the Pleistocene, but may have had repeated, intermittent contact during glacial and interglacial cycles (Pisaric *et al.*, 2003). Pollen records from eastern British Columbia suggest a dominance of *Picea* pollen, putatively attributed to white spruce, only 9,000 YBP (Pisaric *et al.*, 2003), whereas Sitka spruce rapidly expanded its range along the Pacific coast approximately 15,000 YBP following the Quarternary period (Pielou, 1991; Mimura and Aitken, 2007; Shafer *et al.*, 2010). Molecular evidence and recent climate envelope modeling support rapid expansion of Sitka spruce following a stepping stone model of serial recolonization along the Pacific coast, placing it within the Nass-Skeena region well in advance of white spruce (Mimura and Aitken, 2007; Holliday *et al.*, 2010b). These findings suggest introgression may have first occurred between the putative local species (Sitka spruce) and later colonizing species (white spruce), contributing to asymmetrical patterns of introgression towards Sitka spruce (Currat *et al.*, 2008; Petit and Excoffier, 2009).

3.4.2 *Pre- and post-zygotic incompatibilities*

Although patterns resulting from colonization sequence may persist, pre- and post-zygotic incompatibilities can generate asymmetric introgression (Rieseberg and Soltis, 1991; Rieseberg and Blackman, 2010). Anecdotal evidence from artificial crosses between species indicates no intrinsic barriers to reproduction exist in the F1 generation (J. King, pers. comm.) although hybrid breakdown in the F2 generation cannot be ruled out. Additionally, disproportionate pollen movement via air masses moving predominantly west to east may contribute to asymmetry. Average wind direction from April to June within the hybrid zone suggests excess coastal pollen could flow towards the interior (Environment Canada), influencing hybrid composition, assuming reproductive phenology between species is

synchronized. Reproductive phenology, a likely pre-zygotic mechanism that may contribute to asymmetry, results from genetically determined responses to environmental cues in forest trees including highly heritable heat sum requirements (Hanninen and Tanino, 2011). If white spruce is receptive during Sitka spruce pollen dispersal, greater fertilization of white spruce by Sitka spruce pollen may occur. Variation in heat sum accumulations across environments may also contribute to interspecific differences in phenology. While the common garden site experiences an average of approximately 969 degree days over a threshold of 5°C (Cannell and Smith, 1983), there is a wide range exhibited among population origins sampled (848 degree days at Douglas Creek to 1571 degree days at Bella Bella). Therefore, the combination of environmental and genetic variation in phenology may contribute to patterns of asymmetric introgression.

Post-zygotic barriers to reproduction may result in reduced fitness of hybrids, influencing the distribution of hybrid genotypes (Rieseberg and Blackman, 2010). Given evidence for successful mating under artificial conditions producing F1s, and the presence of few backcrosses towards white spruce in natural populations, it is highly unlikely that hybrids are completely inviable under natural conditions. However, spruce hybrids may also exhibit reduced fitness as a result of the breakup of co-adapted gene complexes, Bateson-Dobzhansky-Muller (BDM) incompatibilities (Bomblies and Weigel, 2007), or cytoplasmic incompatibilities between paternally inherited chloroplast and maternally inherited mitochondrial DNA (Neale and Sederoff, 1989; Sambatti *et al.*, 2008).

3.4.3 Genetic and phenotypic relationships with climate and distance

Strong clinal variation in SNP-based hybrid index indicates that the transition between maritime and continental climates along the Nass and Skeena rivers strongly impacts the genetic structure of populations throughout the contact zone. We observed similar climatic and geographic patterns for hybrid index using hundreds of potentially adaptive SNP markers as those seen by Bennuah *et al.* (2004) for a small number of putatively neutral markers. Strong relationships between hybrid index and drainage distance may be attributed to isolation by distance or climate; however, concurrent and strong correlations with climatic variables suggest environmental selection may be a more pervasive force influencing spatial structure. While hybrid index exhibited significant clines with all climate variables tested, mean annual precipitation was associated with particularly steep clines at 1000 mm of precipitation (Figure 3.3). This may suggest introgressants do not tolerate high levels of precipitation, or are out-competed by pure Sitka spruce, but around the threshold of 1000 mm there may be strong

selection against Sitka spruce. Clinal variation in hybrid index is suggestive of exogenous selection, supporting a bounded hybrid superiority model of hybrid zone persistence, although endogenous incompatibilities implicated in the tension zone model have not been tested using advanced-generation (F2 and beyond) controlled crosses (Barton and Hewitt, 1985).

Clinal variation for phenotypic traits was surprisingly weak across the introgression zone. Cold injury phenotypes exhibited shallow clines while height exhibited no clinal pattern. The large degree of variation within populations observed in phenotypes is typical of hybrid zones (Rieseberg and Carney, 1998; Arnold *et al.*, 2008). Different gene combinations can result in similar hybrid indices, but produce a range of phenotypic values. Our expectations were that increased variation due to introgression would be counter-balanced by spatially varying environmental selection, producing phenotypic clines associated with climatic variance. In Sitka spruce, phenotypic variation in cold hardiness and bud set phenology is associated with at least 35 SNPs, of which 15 show temperature and latitudinal clines in frequency across the species range (Holliday *et al.*, 2010a). While a large proportion of phenotypic variation is explained by geographic and climatic variation within species (Holliday *et al.*, 2010a), it remains unexplained between species. If additional phenotypic traits were included, however, such as drought hardiness or pest resistance, clinal patterns may have been resolved. In addition, few individuals were sampled from the eastern end of the introgression zone, and this may contribute to the lack of significant clines among phenotypic traits. Given the relatively moderate environment of the common garden within the hybrid zone, extension of sites into allopatric zones, along with assessment of additional phenotypic traits, may provide a more robust test of the bounded hybrid superiority model.

3.4.4 Genotypic classification

Most genotypes were classified as backcrosses toward Sitka spruce (69%) or as advanced-generation hybrids (FN – 19%). There was no clear evidence of an F1 generation across the hybrid zone, the lack of which is consistent with an advanced generation hybrid zone characterized by widespread interspecific recombination (Lexer *et al.*, 2010; Teeter *et al.*, 2010). The lack of early-generation hybrid derivatives is in line with selection for segregating hybrids within ecologically transitional habitat, where hybrid genotypes may exhibit novel adaptations or combinations of traits (Rieseberg *et al.*, 1999a; Milne and Abbott, 2008). Other hybrid zones have similarly observed a dearth of early-generation hybrid classes where F1-hybrid genotypes either make up a minority of hybrids present or are absent (Milne and Abbott, 2008; Lexer *et al.*,

2010). Assignment of genotypic classes, however, comes with some expectations of error or mis-assignment. Using our approach to class assignment, which was relatively flexible and broad, genotypes with very similar profiles are unlikely to be grouped inaccurately (Milne and Abbott, 2008). The sampling scheme in the pre-existing experiment may have introduced a bias in the distribution of genotypic classes. Few individuals were sampled from the eastern (coldest) end of the introgression zone, which may have exacerbated the appearance of asymmetric patterns. The sampling design of the experiment, initially established to compare local adaptation and productivity among seed sources for reforestation within the hybrid zone, reflects that goal, focusing within the hybrid zone and to a lesser extent its periphery and beyond.

One consequence of the greater movement of Sitka spruce genome towards white spruce populations is that new alleles or allele combinations will expand the gene pool of Sitka spruce further into the hybrid zone over time, potentially constituting a selective advantage for Sitka genes within the hybrid zone (Keim *et al.*, 1989). Accumulation of genes through directional introgression may allow new Sitka-like genotypes to become better adapted to the colder, drier environments generally characteristic of white spruce. Climate modeling scenarios predict extensive changes in population reproductive biology, phenology, and geographic ranges, as well as extensive community and ecosystem-level changes in composition and functioning over the next century (Hamann and Wang, 2006; Parmesan, 2006; Wang *et al.*, 2006b; Aitken *et al.*, 2008). Introgression may provide novel allelic variation, resulting in an increased ability to adapt to new conditions (Aitken *et al.*, 2008). It is likely that a major limiting factor to the introgression of Sitka-like genotypes further inland is the development of cold and drought hardiness, although the latter was not directly phenotyped. The combined influence of climate warming and positive transgression evident in hybrids for cold injury suggests suitable habitat for hybrids may be increasing. Although steep clines in hybrid index related to mean annual precipitation indicate that precipitation may be a limiting factor to Sitka spruce, Sitka genotypes may survive further inland over time and are more likely to become productive across the introgression zone, particularly given this region has become both warmer and wetter over the past 25 years (Mbogga *et al.*, 2010).

3.4.5 *Fitness of genotypic classes*

The ability to grow competitively and develop sufficient cold hardiness in advance of first frost is critical for the health and survival of boreal tree species (Morgenstern, 1996; Howe *et al.*, 2003). Increased cold hardiness in hybrids relative to parental genotypes within the hybrid

zone provides support for the bounded hybrid superiority model. While we have not directly tested the model using reciprocal transplants, we have examined fitness proxies for both parental and hybrid genotypes within this zone. Interestingly, we saw that hybrid genotypes exhibited slightly greater cold hardiness than parental species on average when tested at -8°C, but that with colder test temperatures this pattern changed (Figure 3.4). Heterosis in early-generation hybrids usually results from dominant allelic effects, but is generally not retained in advanced-generation recombinants (Rieseberg and Carney, 1998). As cold injury patterns at -8°C are common for all hybrid classes, none of which are early-generation hybrids, heterosis is unlikely. Rather, this pattern may reflect transgressive segregation due to complementary gene action of alleles differentiated in parental populations combining to produce phenotypic extremes (Rieseberg *et al.*, 1999a; Latta *et al.*, 2006). Given the well-differentiated climatic norms exhibited between species, we predict that opposing allele combinations within the species may contribute to transgressive phenotypes for hybrid individuals. Examining both genomic and geographic clines of mapped SNP loci may allow us to identify regions of the genome that are well-differentiated between parent species, along with locus-specific effects of candidate genes that may play a role in cold adaptation.

Surprisingly, there was little variance in height across genotypic classes, and no significant pairwise differences after accounting for multiple comparisons. The only class to exhibit slight differences was the backcross to white spruce class, which also exhibited the greatest cold tolerance across genotypic classes at both test temperatures. While this class may exhibit the well-established trade-off between growth and development of cold hardiness, it also has the fewest members of any class (N=15). Intrinsic incompatibilities may contribute to lower overall fitness and lower numbers. Additionally, the test site had a relatively benign, intermediate environment (MAT=2.9°C and MAP=802 mm) relative to the parental habitats, and had relatively high survival and growth of both species as well as hybrids, contrary to predictions based on the bounded hybrid superiority model. Height data was analyzed for two additional common garden sites on which this experiment was planted within the hybrid zone that were not genotyped or phenotyped for cold hardiness that exhibited more extreme temperatures (more continental site, MAT=2.2°C, MAP=1057mm and more maritime site, MAT=5°C, MAP=1578 mm). These sites also revealed a similar lack of significant provenance or climatic variance in height across genotypic classes (unpublished data).

While environmental selection plays an important role in maintenance of the Sitka-white spruce hybrid zone, we caution against conclusions based exclusively on exogenous or

endogenous selection alone. Given current patterns of introgression along with predicted patterns of climate change for this region (Tongli Wang, UBC, pers. comm.), we expect that introgression of Sitka spruce genes further inland will have predictable impacts on the genetic architecture of recipient populations. However, selection of seed sources for reforestation across the Nass-Skeena transitional area remains problematic. Uncertainty in identifying seed optimally adapted to particular environments that will translate into tree health and productivity, and ultimately greater economic or ecological value, remains. A better understanding of the precise nature of interactions between dispersal, selection, and drift maintaining the spruce hybrid zone will require analysis of geographic and genomic clines in a genetic map-based context, along with testing for reproductive success of advanced-generation controlled crosses and fitness of the resultant progeny.

Table 3.1 Origins of Sitka x white spruce (*Picea sitchensis* x *P. glauca*) study populations, sample size (climatic and geographic distance variables).

Population	N	Latitude	Longitude	Elevation (m)	Drainage Distance (km)	MAT (°C)
Haida Gwaii (HG) *	24	53.65	-132.2	10	0	7.5
Mt. Kaien	18	54.3	-130.33	600	0	6.2
Seal Cove	18	54.32	-130.32	30	0	8
Bella Bella	16	52.17	-128.16	60	0	8.6
Ocean Falls	17	52.36	-127.69	60	24.825	7.8
Ramsden Point	30	55.02	-130.13	20	42.878	7
Kinkolith - Low	30	55	-129.93	30	51.011	6.6
Exstew - Low	18	54.42	-129.07	30	85.175	5.6
Eskers O/P - Low	24	54.41	-129.03	100	85.175	5.6
Eskers O/P - Mid	18	54.41	-129.02	450	85.175	3.9
Greenville - Low	24	55.09	-129.5	220	86.665	5.2
Kitimat - Low	30	54.07	-128.54	100	108.438	6.3
Hagensburg	18	52.39	-126.58	60	115.241	7
Kitimat - High	30	54.2	-128.55	250	119.074	5.4
Nass Camp - Low	30	55.3	-129.05	185	126.398	5.5
Nass	29	55.28	-129.02	100	127.246	5.5
Douglas Creek - High	30	54.83	-128.73	800	150.426	2.5
Douglas Creek - Mid	30	54.83	-128.69	450	152.721	4.6
Kiteen - Low	29	55.44	-128.83	310	155.786	4.4
Doreen - Low	24	54.83	-128.31	220	165.852	5
Talchako River - Low	18	52.15	-125.96	305	180.381	4.9

MAT, mean annual temperature; MAP, mean annual precipitation

* parental populations used to calculate hybrid index

Table 3.1 Cont'd

Population	N	Latitude	Longitude	Elevation (m)	Drainage Distance (km)	MAT (°C)	MAP (mm)	Continentality (°C)
Cedarvale - Low	18	55.02	-128.35	110	187.612	5.4	832	20.3
Talchako River - High	18	52.06	-125.92	488	190.646	3.6	985	19.8
Cedarvale - Mid	18	55.03	-128.42	230	192.253	4.5	894	20.1
Donaker Creek	18	57.45	-131.08	300	217.39	1.8	692	26.5
Kitwanga - Low	24	55.37	-128.14	450	219.64	3.4	640	21.8
Kitwanga Lake - Mid	24	55.36	-128.13	680	221.451	2.8	673	21.8
Skeena Crossing	18	55.1	-127.81	210	221.747	4.9	590	22.8
Meziadin	18	56.1	-129.3	265	233.807	4.7	908	21.1
Moricetown	18	55.03	-127.34	440	277.244	3.7	540	23.1
Eastern North America (ENA) *	42	45.38	-75.71	79.2	NA	6.3	914.2	NA

MAT, mean annual temperature; MAP, mean annual precipitation

* parental populations used to calculate hybrid index

Table 3.2 Results of linear regression of cold injury at two artificial freezing temperatures and tree height measured for 721 individuals sampled across the Sitka x white spruce hybrid zone and grown in a common garden, with geographic and climatic variables across all 721 individuals

	Cold Injury -8°C			Cold Injury -18°C			Height (mm)		
	R ² (%)	Slope	F-value (Pr>F)	R ² (%)	Slope	F-value (Pr>F)	R ² (%)	Slope	F-value (Pr>F)
Geographic variable									
Drainage Distance (km)	7.0	-0.09	52.57 (p<0.001)	7.0	-0.06	53.25 (p<0.001)	0.0	0.00	0.00 (0.98)
Climatic variables									
MAT (°C)	6.0	3.59	40.15 (p<0.001)	4.0	1.99	26.77 (p<0.001)	0.0	-0.38	0.04 (0.56)
MAP (mm)	6.0	0.00	40.87 (p<0.001)	3.0	0.00	20.29 (p<0.001)	0.0	0.00	0.56 (0.77)
Continentality (°C)	6.0	-1.64	41.87 (p<0.001)	4.0	-0.89	26.64 (p<0.001)	0.0	0.90	0.99 (0.32)

Table 3.3 ANOVA results for comparison of genotypic class across phenotypes putatively related to fitness in trees, along with non-parametric validation and significant pairwise comparisons based on Tukey's honest significant difference test for multiple comparisons ($\alpha=0.05$).

Source	SS	DF	MS	F (p-value)	χ^2 (p-value)
Cold injury -8C					
Genotype Class	31545	4	7886.3	14.874 (p<0.001)	44.2922 (p<0.001)
Error	367430	693	530.2		
Cold injury -18C					
Genotype Class	14817	4	3704.2	15.046 (p<0.001)	34.0644 (p<0.001)
Error	170114	691	246.2		
Height (mm)					
Genotype Class	80114	4	20029	3.1126 (p=0.01)	12.9968 (p=0.01)
Error	4562176	709	6434.7		

Figure 3.1 Map of individuals sampled across the introgression zone between Sitka and white spruce in northwestern British Columbia, indicating provenance of origin of samples (circle) collected and planting location of the common garden experiment (square). Inset map provides North American range of *P. sitchensis* (dark grey) and *P. glauca* (light grey), as well as source of Eastern North American population.

Figure 3.2 Interspecific heterozygosity (vertical axis) vs. hybrid index (horizontal axis) using 268 candidate gene SNPs (A) and a subset of 31 SNPs that exhibited an allele frequency differential (δ) greater than 0.90 (B). Plot based on all loci has been used to assign individuals to genotypic classes, where hybrid index 0=white spruce and 1=Sitka spruce. Individuals are classified as pure parental species (Sitka or white), backcrosses towards either Sitka spruce (BCS) or white spruce (BCW), or advanced generation hybrids (FN), based on clear discontinuities in two-dimensional space. See text for details.

Figure 3.3 Relationship between 268 SNP-based hybrid index (0=white spruce, 1=Sitka spruce) and geographic and climatic variables across 721 individuals spanning the Sitka-white introgression zone, including drainage distance (km), mean annual temperature (°C), mean annual precipitation (mm), and continentality (°C).

Figure 3.4 Boxplots illustrating variation in phenotypic traits among genotypic classes including pure parentals (Sitka and white), backcrosses to either Sitka (BCS) or white spruce (BCW), and advanced-generation hybrids (FN). Phenotypic traits are cold injury at -8°C and -18°C estimated in artificial freeze tests, and height at age ten in the common garden experiment. Boxplots with same letter are not significantly different based on Tukey's comparison of means.

4. Differential introgression reveals candidates for selection across a spruce (*Picea sitchensis* x *P. glauca*) hybrid zone

4.1 Introduction

The genetic architecture of hybrid zones offers the unique opportunity to evaluate how gene flow and selection facilitate the spread of genetic material in a natural setting (Barton and Hewitt, 1985; Rieseberg *et al.*, 1999b; Lexer *et al.*, 2005; Lexer *et al.*, 2010). While much work has addressed broad-scale patterns of introgression between species we are only beginning to resolve fine-scale patterns, identifying both candidates within select regions of the genome contributing to the maintenance of species barriers and candidates for divergent selection (Strasburg *et al.*, 2012). Heterogeneous patterns of introgression across the genome reveal those regions that may be influenced by divergent selection and regions important to local adaptation, providing a fine-scale depiction of genomic patterns influenced by both extrinsic and intrinsic factors contributing to the evolution and maintenance of species barriers (Wu, 2001; Gompert *et al.*, 2012).

In tree species, where the formation of multiple generations of artificial crosses remains largely impractical due to long-generation times, hybrid zones provide a range of natural recombinants in which to examine introgression between species where barriers are permeable (Martinsen *et al.*, 2001; Scotti-Saintagne *et al.*, 2004; Lexer *et al.*, 2005). Incomplete barriers to gene exchange may permit gene flow between species, although gene flow will be restricted if recombination is limited or there is strong selection against hybridization (Barton and Hewitt, 1985; Scotti-Saintagne *et al.*, 2004; Kane *et al.*, 2009). Outside of these regions, however, both neutral, and positively selected loci may move across the permeable genome through hybridization and introgression (Kane *et al.*, 2009). In contrast to the traditional whole-genome view of isolation advocated by the biological species concept, this perspective identifies the gene as the unit of differentiation between species (Wu, 2001). Thus, locus-specific patterns of introgression across a porous genome distinguish those regions that may be candidates for differentiation (Wu, 2001; Nosil and Schluter, 2011).

Genome scans examining interspecific differentiation, along with patterns of introgression at both landscape- and genome-levels, have been applied across a number of hybrid zones, identifying candidate regions important to the maintenance of species barriers (Scotti-Saintagne *et al.*, 2004; Yatabe *et al.*, 2007; Carling and Brumfield, 2009; Nolte *et al.*, 2009; Lexer *et al.*, 2010; Teeter *et al.*, 2010; Renault *et al.*, 2012). Combining these approaches

provides a powerful approach in identifying both broad- and fine-scale consequences of gene exchange between species that may be important both in terms of local adaptation and the maintenance of species barriers (Lexer *et al.*, 2004; Payseur, 2010; Gompert *et al.*, 2012).

The effects of natural selection are revealed at a longer time scale in near-equilibrium situations through estimation of interspecific differentiation (Campbell and Bernatchez, 2004), while more contemporary influences are evident in the degree of introgression across natural hybrid zones (Lexer *et al.*, 2010). At a contemporary scale, introgression reveals the influence of extrinsic or intrinsic factors contributing to fitness variation amongst natural genetic recombinants both across the landscape and within the genome itself.

Differential patterns of introgression indicate the spatial movement of alleles across the landscape, as well as movement of genetic material into varying genomic backgrounds (Gompert *et al.*, 2012). Traditional cline analysis examines the change in allele frequencies across the landscape, estimating the strength of intrinsic and extrinsic (environmental) selection using a wealth of theoretical background to describe cline shape (Barton and Hewitt, 1985; Payseur, 2010). More recently, the application of novel admixture mapping techniques has started to examine introgression of individual genotypes against levels of admixture (Buerkle and Lexer, 2008; Gompert and Buerkle, 2010; Payseur, 2010). Selection may promote introgression of select genes or genomic regions into new genetic backgrounds or enhance barriers to gene flow resulting in differential patterns of introgression across the genome (Gompert *et al.*, 2012).

While hybrid zone clines reveal broad, recent selection dynamics across the hybrid zone, patterns of interspecific genetic differentiation as estimated by locus-specific F_{ST} reveal a longer time-scale of selection (Whitlock, 1992; Lexer *et al.*, 2010). The strength of selection is revealed by the extent to which opposing gene flow homogenizes the accumulation of differentiation over time (Whitlock, 2008). These patterns, combined, are complementary and informative at different time scales, providing greater power to detect segments of the genome that may be important in terms of adaptation and maintenance of species barriers (Bierne *et al.*, 2011).

With an increasing call to explore new forest management strategies in the face of climate change, a move towards marker-assisted selection in forest trees will benefit from multilocus genome scans across hybrid zones. Analysis of fine-scale transfer of candidate genes linked to functionally adaptive phenotypes will allow inference regarding both selective forces at play in the maintenance of the introgression zone, the potential transfer of adaptations, and identification of regions of the genome that maintain species. Furthermore, these data will inform our projections regarding the capacity of hybrid populations to adapt to future climates.

Adaptation from standing variation may benefit from admixture where beneficial alleles have been ‘selectively filtered’ in past environments or in separate parts of species ranges, contributing to increased variation available for natural selection (Barrett and Schluter, 2008; De Carvalho *et al.*, 2010). Determining how long-term and more contemporary processes influence and maintain patterns of genomic variation and differentiation, and consequently their evolutionary potential, remains an important challenge.

In this paper we examine the fine-scale transfer of genetic material between two genetically and ecologically distinct species, Sitka spruce (*Picea sitchensis*) and white spruce (*P. glauca*). The broad goals of the study are to examine the genetic consequences of hybridization. Specifically, we explore variable patterns of introgression and differentiation across candidate gene loci putatively associated with adaptive phenotypic traits in forest tree species.

Sitka spruce, the largest of *Picea* species, exhibits a narrow latitudinal distribution along the Pacific Coast, while the distribution of white spruce spans the boreal forests of North America. These species come into contact along the Nass and Skeena river valleys of northwestern British Columbia, and into Alaska, where hybrids are referred to as *P. lutzii*. The contact zone in British Columbia spans an ecological gradient from the maritime climate of Sitka spruce to the continental climate of white spruce (Roche, 1969; O'Neill *et al.*, 2002; Bennuah *et al.*, 2004). While the maritime climate exhibits up to two metres of precipitation annually, and mean annual temperatures of approximately 8°C (Pojar *et al.*, 1991), the continental climate is characterized by seasonal extremes. Hot, dry summers with temperatures exceeding 10°C and cold, moist winters with temperatures below 0°C are characteristic of the continental climate (Ketcheson *et al.*, 1991).

The introgression zone between Sitka and white spruce offers the opportunity to examine gene flow and selection across this ecological gradient. Recent identification of candidate gene single nucleotide polymorphic (SNPs) markers putatively associated with adaptive phenotypes in spruce (Namroud *et al.*, 2008; Holliday *et al.*, 2010a) has made feasible the evaluation of locus-specific patterns of introgression and differentiation that may be important to local adaptation.

Taking a genome scan approach, we tested for patterns of differentiation amongst loci identified as candidates for selection or candidates linked to regions influenced by selection using a combination of approaches. We used F_{ST} -outlier tests to test for signatures of divergent or balancing selection among loci between parent species under near-equilibrium conditions. We then combined geographic and genomic clines to test both for the strength of selection and identify deviations from the genome-wide average as estimated with the studied SNPs.

Consequently, locus-specific patterns provide a fine-scale representation of both spatially-varying selective pressures and potential intrinsic barriers, identifying those regions of the genome that may be responsible for reproductive isolation and local adaptation across this hybrid zone.

4.2 Materials and methods

4.2.1 Sampling

Open-pollinated seeds were sampled from the upper canopy of three mature trees across each of 29 locations spanning the contact zone between *Picea sitchensis* and *P. glauca* along the Nass and Skeena rivers in northwestern British Columbia in 1997. Allopatric reference populations of *P. sitchensis* on Haida Gwaii (HG) and *P. glauca* from the Ottawa valley region, Ontario were also sampled (ENA; Table 3.1 and Figure 3.1). For each seed parent, ten open-pollinated progeny were planted in a nursery for one year and then transplanted into a common garden near Kitwanga, British Columbia within the hybrid zone by British Columbia Ministry of Forests, Land and Natural Resource Operations (55°17'N, 128°10'W, Figure 3.1). Newly flushed lateral buds were sampled across 721 individuals within the common garden in June 2009 and frozen in liquid nitrogen for subsequent DNA extractions and genetic analysis.

4.2.2 Genetic data analysis

DNA extraction was performed using a modified CTAB protocol (Doyle and Doyle, 1990) in which 50 mg of tissue was ground in liquid nitrogen for extraction using a Mixer Mill MM 400 (Retsch). Individuals were genotyped for 384 SNPs (single nucleotide polymorphisms) identified from previous studies of parental species as putative candidates for roles in budset timing and development of cold-hardiness (Holliday *et al.*, 2008; Holliday *et al.*, 2010a), weevil resistance (K. Ritland *et al.* unpublished data) and growth (J. Bousquet *et al.* unpublished data). SNP selection and quality criteria are described in detail in Hamilton *et al.* (submitted). SNPs were genotyped using the Illumina bead array platform in conjunction with the GoldenGate allele-specific assay (Shen *et al.*, 2005; Fan *et al.*, 2006). Genotyping quality was evaluated using the Genomestudio Genotyping Module (v1.0) and described in detail in J. Hamilton *et al.* (submitted). Of the initial 384 SNPs genotyped, 268 met the high quality standards for subsequent analysis.

4.2.3 Descriptive statistics

Global and locus-specific estimates of F_{ST} , expected heterozygosity (H_e), and inbreeding coefficient (F_{IS}) were calculated across all populations and between allopatric populations using *Arlequin 3.1* (Excoffier *et al.*, 2005). The significance of the inbreeding coefficient was examined using Fisher's exact tests to test for departures from Hardy-Weinberg equilibrium.

4.2.4 Identifying outlier loci potentially affected by selection

To identify candidate gene SNPs that deviated from a null hypothesis of selective neutrality at near-equilibrium conditions between allopatric reference populations, we used an F_{ST} -outlier detection method to assess interspecific patterns of differentiation. Loci that exhibit extreme values of genetic differentiation relative to average patterns of the majority of loci may be candidates for diversifying selection, while those loci that exhibit unusually low levels of differentiation may be under balancing selection (Holsinger and Weir, 2009).

We used a Bayesian likelihood method developed by Foll and Gaggiotti (2008) implemented in the program *Bayescan* to examine patterns of differentiation. This approach teases apart the influence of two models, differentiating locus-specific (α_i) and population-specific (β_j) effects using logistic regression to calculate locus-specific estimates of F_{ST} (Foll and Gaggiotti, 2008). This approach, unlike FDIST2 (Beaumont and Nichols, 1996) which evaluates genetic differentiation based on the island model, may be more appropriate where gene flow is non-symmetrical (Beaumont and Nichols, 1996; Nielsen *et al.*, 2009). Previous research across this hybrid zone suggests gene flow is not symmetrical (Hamilton *et al.* submitted).

Positive values of the locus-specific parameter (α_i) indicate that locus i may be under diversifying selection, while a negative value indicates balancing selection. The model including only population-specific (β_j) effects was compared with the model including both locus-specific and population (α_i and β_j) effects to test for selection. The ratio of posterior probabilities of the two models provides a Bayes' factor (BF, (Foll and Gaggiotti, 2008), which may be interpreted in terms of level of support for loci being subject to selection (very strong support between $\log_{10}(\text{BF})$ 1.5 and 2; decisive support for $\log_{10}(\text{BF}) > 2$) corresponding to posterior probabilities of the model of selection (substantial evidence between 0.91 and 0.99, decisive between 0.99 and 1).

Following 20 pilot runs of 5,000 iterations, we used 50,000 additional iterations (sample size of 5,000 and thinning interval of 10) to identify loci under selection. While these tests are informative, they may be subject to the influence of population demography, and hierarchical

population structure can result in over-estimation of selection (Excoffier *et al.*, 2009; Siol *et al.*, 2010).

4.2.5 Genomic cline analysis

Genome-wide estimates of admixture were calculated using the R package *Introgress* (Gompert and Buerkle, 2010), which uses a maximum likelihood approach to estimate ancestries of individuals summarized in a hybrid index. The program requires *a priori* knowledge of pure parental individuals, here identified as those reference populations – HG (Sitka) and ENA (white) to estimate the proportion of ancestry attributed to either parent. A hybrid index of zero corresponds to pure white spruce, while a hybrid index of one corresponds to pure Sitka spruce.

Employing the genomic clines method of Gompert and Buerkle (2009) implemented in *Introgress* (Gompert and Buerkle, 2010), genomic clines were estimated for individual loci based on the probability of observing three possible genotypes at each locus, WW (homozygous white spruce), WS (heterozygote genotype) or SS (homozygous Sitka spruce) as a function of genome-wide estimates of admixture (hybrid index). These genomic clines are estimated using multinomial regression of observed single-locus genotypes as a function of genome-wide admixture (Gompert and Buerkle, 2009). The likelihood of the regression model is compared with a null model of genotype frequencies simulating neutral introgression using parametric simulations as described in Gompert and Buerkle (2010).

Deviations from the genome-wide average as estimated with the studied SNPs are summarized through comparison(s) of the probability density of observed genotypic data with those of the neutral model of introgression following 1,000 stochastic parametric simulations. Evidence for positive (+) or negative (-) selection for homozygous genotypes (WW and SS) corresponds, respectively, to an increase or decrease in the total probability density of the observed genotype compared to the corresponding probability density of the homozygous genotype of the neutral model (Gompert and Buerkle, 2009; Nolte *et al.*, 2009; Gompert and Buerkle, 2010). Similarly, evidence for overdominance (WS+) or underdominance (WS-) is obtained from the comparison of observed heterozygote genotypes (WS) with the probability density of heterozygote genotypes under the neutral model.

4.2.6 Geographic cline analysis

The geographic transect was estimated using the non-linear drainage distance up river valleys from the Pacific Ocean. Previous studies have shown that genomic composition as

estimated by hybrid index is best predicted by drainage distance through this mountainous region (O'Neill *et al.*, 2002; Bennuah *et al.*, 2004). Drainage distance characterizes the likely corridor of gene flow along valleys through which pollen, and to a lesser extent seed dispersal, occur between the coastal maritime climate and the interior continental climate (Bennuah *et al.*, 2004). Transect distance (in kilometers) was calculated from the coast inland using ArcGIS (Version 10) replicating methods from Benuah *et al.* (2004). The pure white spruce reference population from Ontario (ENA) was excluded from this analysis due to its long distance from all other populations in this geographic transect.

Geographic cline parameters were estimated individually for each SNP using a modified approach to traditional analysis. Traditional cline analysis assumes a sigmoidal cline that describes sharp changes in allele frequencies near the centre of the cline and shallower asymptotes towards the edges of the contact zone (Barton and Hewitt, 1985; Payseur, 2010). The most common parameters used to describe the shape of the cline are cline width and cline centre (Porter *et al.*, 1997; Carling and Brumfield, 2009; Payseur, 2010; Teeter *et al.*, 2010). The existing field experiment sampled for this study was established to assess seed transfer for reforestation within the hybrid zone and not in allopatric areas, and thus did not sample either a pure white spruce reference population from British Columbia, or populations at the eastern margin of the hybrid zone that may also include introgression from a third species (*P. engelmannii*, (Roche, 1969). The lack of pure white spruce is evidenced by an exponential increase in allele frequencies to the east without an upper asymptote. Consequently, we developed an alternate approach, estimating proxies for traditional parameters, to account for the distortion in allele frequencies attributed to incomplete sampling of parent populations.

This novel approach has been adapted from parameter estimation using equations found in biological growth curves that estimate maximum slope of a non-linear regression and “length of lag phase” (Kahm *et al.*, 2010; Paine *et al.*, 2011). We interpret the maximum slope as an estimate of the strength of selection, comparable to the traditional estimate of width (slope=1/width) in a sigmoid curve (Porter *et al.*, 1997). Greater slope values indicate stronger barriers to introgression between species, whereas lesser slope values may indicate either extensive allele-sharing through recurrent gene flow or shared polymorphisms. “Length of lag” is interpreted as the geographic distance accumulated from the coast prior to the reaching the maximum rate of change in allele frequency. This parameter provides a geographic estimate of the location on the landscape associated with this maximum rate of change.

Linear regressions were compared with non-linear regressions estimated using both asymptotic (logistic) and non-asymptotic (exponential) functions (Paine et al. 2011). The exponential model assumes a constant rate of increase with respect to change in allele frequency and drainage distance with no asymptote, whereas the logistic model assumes an initial asymptote without a constant rate of increase. All regressions were implemented in R using the self-starting routines (*SSlin*, *SSlogis* and *SSexp*) following recommendations from Paine *et al.* (2011). AIC and R^2 values were compared between regression models to assess the best-fit model. Where the logistic model failed to converge, the exponential model was used as the sole comparison with the linear model. Using the fitted data from the best-fit model, we fit a smoothed cubic spline using the *gcFitSpline* function within the R package *grofit* (Kahm *et al.*, 2010) to estimate maximum slope and lag phase. The advantage of this approach is the step-by-step development of a best-fit-model for each SNP using the approach described in Paine *et al.* (2011).

Congruency was inspected between genomic and geographic clines identifying those loci that may be targets of selection. Enrichment of under- or over-dominance amongst heterozygotes for genomic clines and directional shifts were assessed in geographic and genomic clines to identify those loci that are involved in reproductive isolation or adaptation.

4.3 Results

The global estimate of mean genetic differentiation (F_{ST}) across all populations sampled was 0.18, ranging from 0.02 to 0.47 for individual loci (Table S2). Expected heterozygosity across all loci ranged from 0.04 to 0.5 (mean $H_e = 0.31$). The global estimate of the inbreeding coefficient (F_{IS}) was 3.73×10^{-5} (ranging from -0.29 to 0.29 for individual loci), indicating overall low levels of inbreeding. No loci exhibited significant deviations from Hardy-Weinberg equilibrium based on exact tests ($p < 0.05$).

Comparison of allopatric populations using AMOVA indicates that the majority of variation is accounted for between species (58.7%, Table 4.2). Overall, global estimates of F_{ST} between allopatric *P. sitchensis* (HG) and *P. glauca* (ENA) were 0.59 and F_{IS} was 0.005, indicating interspecific differentiation exceeded those values evaluated across all populations and low levels of inbreeding remained. These results are in line with locus-specific estimates of interspecific differentiation values between the allopatric populations, ranging from zero to one.

4.3.1 Detection of loci under selection

Using *Bayescan*, which uses a Bayesian decision-making approach to model choice in the form of a “Bayes factor” to provide evidence in support of a model of selection or neutrality, we observed three loci with a high posterior probability (>0.90 or $\log_{10}(\text{BF}) > 1$) of being subject to selection. Of the 265 SNPs examined, excluding those SNPs that show no interspecific differentiation, there are three SNPs (SNP53, 55, and 106) considered candidates for or linked to candidates for diversifying selection (Figure 4.1). There were, as expected, no candidates for balancing selection given that loci were initially selected based on degree of interspecific differentiation in SNP frequencies. There are few outlier loci between parental species, suggesting parent populations are likely close to equilibrium due to rapid decay of linkage disequilibrium in forest trees (Neale and Ingvarsson, 2008; Neale and Kremer, 2011).

4.3.2 Geographic clines

While patterns of interspecific differentiation may reveal long-term influences of selection, direct examination of the introgression zone using clinal analysis may provide additional power to detect targets of selection under non-equilibrium conditions. Geographic cline parameters, slope and “lag phase” (λ), were evaluated for each SNP (Table 4.1, Figure S3). Cline parameters reflect the change in frequency of the minor ‘Sitka spruce’ allele and, respectively alternate ‘white spruce’ allele with increasing drainage distance from the ocean. Maximum slope of regressions, describing changes in allele frequency per km of drainage distance, ranged from -0.001 to 0.012, with an average slope of 0.003 (Table 4.1). λ is the intercept of maximum slope in km, indicating the point along the drainage distance transect from the ocean at which the greatest rate of change in allele frequency occurs. Locus-specific estimates ranged from zero km to 236 km, with an average at 119 km across all loci. Negative drainage distance values were converted to zero and λ values were not estimated where the slope was equal to zero for 22 and 45 loci respectively. Few loci exhibited maximum cline slope greater than 0.011 suggesting the absence of strong or complex barriers to gene flow between the two species.

4.3.3 Genomic cline analyses

Extensive heterogeneity was observed among loci for genomic clines across the introgression zone, indicating loci deviate considerably from the genome-wide average as estimated from all SNPs (Table 4.1, Table S2, Figure S4). All loci were assessed for genomic

clines, excluding those loci that exhibited no allelic differentiation between parents ($n=4$). Of the remaining loci, 94 (36%) exhibited a neutral pattern of introgression, indicating no deviation from the genomewide estimate of ancestry. We observed a wide range of other patterns, including 24% of loci exhibiting evidence for excess ancestry of white spruce alleles in a Sitka spruce background (WW+, Table 4.1, Figure S4) and 67 with an excess of Sitka spruce alleles in a white spruce background (SS+, Table 4.1, Figure S4). Sixty loci had patterns consistent with underdominance (WS-), with a lack of heterozygous genotypes, while 66 loci showed overdominance (WS+), with an excess of heterozygotes compared to the model of neutral introgression (Table 4.1, Figure S4). Of these sixty loci, select loci exhibited underdominance (9 loci, WS-) and overdominance (16 loci, WS+) respectively when homozygous genotypes exhibited no deviation from genomewide ancestry. These loci may be candidates for transfer of adaptations contributing to transgression in adaptive traits where heterozygotes are over-represented, or may be candidates for regions of the genome involved in reproductive isolation where heterozygotes are under-represented. Comparing homozygote genotypes individually we observe a number of instances where there is an excess of one genotype relative to the genomewide estimate of ancestry indicating positive selection (+), while the alternate genotype is either negatively selected (-) or neutral. Evidence for directional selection across the genome, along with patterns of selection in one genetic background with neutrality in the alternate background bear further investigation with respect to extrinsic and intrinsic selective pressures.

Genetic map relationships among these markers remains unknown, therefore some values may be inflated if linkage disequilibrium exists between a marker and an adjacent genetic region under selection. Preliminary assessment of linkage relationships across a number of markers used in this study, however, indicates that loci are well-distributed throughout the genome (J. Bousquet and N. Isabel unpublished data).

4.4 Discussion

Characterization of the adaptive capacity of forest tree populations given current climate change predictions remains a large challenge within the field of forest genomics (Aitken *et al.*, 2008; Neale and Kremer, 2011). Differential patterns of introgression exhibited within this natural hybrid zone suggests that targets of selection may differ strongly both across the landscape and between divergent genomic backgrounds. Substantial heterogeneity both across the geographic landscape and within the genome itself, indicate both the general permeability of

spruce genomes to introgression, yet identify specific genomic regions and geographic barriers where interspecific recombination may be limited. Both geographic cline width estimates and patterns of underdominance in genomic clines identify some genomic regions where gene flow may be limited. Moreover, fine-scale variability amongst loci indicates that both extrinsic and intrinsic selective factors likely influence introgression. Congruency between geographic clines and genomic clines provide some parallels among SNPs, indicating strong role for extrinsic selection, particularly in association with precipitation gradients. In addition, patterns of differentiation over a longer evolutionary period viewed through outlier analysis point toward candidate gene loci that may be linked to genes that are important to local adaptation or maintenance of species barriers. These results provide a fine-scale examination of the movement of genetic material across this ecological transition zone, identifying potential targets regions of the genome that warrant further exploration in the future with respect to adaptive trait function.

4.4.1 Interspecific genetic differentiation reveals targets of selection

The influence of selection can be observed in the distribution of genetic differentiation across the genome. Candidate genes for divergent selection have a greater likelihood of residing in areas of the genome that are highly differentiated between species and may be involved in reproductive isolation or adaptation (Beaumont and Nichols, 1996; Beaumont and Balding, 2004). We observed few candidate genes for divergent selection (Figure 4.1), and not all of these candidates will necessarily be directly under divergent selection (Via, 2012). The influence of selection may be observed up to 50 cM away from the target of selection due to the influence of linkage (Gompert and Buerkle, 2009). Consequently, caution should be employed when interpreting loci as under “divergent” or “balancing” selection. However, while these tests remain subject to some caution, they also offer a useful first step in identifying those candidate genes that may be targets of or linked to targets of selection.

Two of the SNPs exhibiting a pattern of divergent selection in the outlier tests (SNP53 and 55) are found within the same gene approximately 100 bp apart. Although we expect LD to be lower in allopatric populations compared to hybrid populations, LD may still influence the accumulation of differentiation in allopatry (Bierne *et al.*, 2003). While physical proximity likely contributes to LD, the influence of selection will also likely preserve LD. The strong LD ($D' = 0.99$, $p < 0.0001$) implies that one of these SNPs or a locus linked to both is likely the target of selection.

Our results from similar latitudes across an east-west precipitation gradient identify different candidates for selection than those observed across a temperature gradient in Holliday *et al.* (2010). SNPs 53 and 55 are found in a gene homologous to *sld* identified in *Arabidopsis* (At2g46210.1) and have been assayed in previous studies examining intraspecific variation in the development of autumnal cold hardiness and budset in Sitka spruce (Holliday *et al.*, 2008; Holliday *et al.*, 2010a). This gene is believed to be a fatty acid/sphingolipid desaturase that produces divergent phenotypes in *Arabidopsis* during response to cold temperatures (Chen *et al.*, 2012). Sphingolipids likely contribute to both ion permeability and osmotic adaptation during response to both freezing and dehydration (Steponkus, 1984). Interestingly, this candidate gene was not associated with bud set or cold injury traits examined in Sitka spruce by Holliday *et al.* (2010). Holliday *et al.* (2010) studied populations spanning north-south gradients in Sitka spruce across the species' range of mean annual temperature and degree-days associated with both phenological cues and development of cold hardiness. In comparison, we sampled populations spanning steep east-west gradients in mean annual precipitation that appear to strongly influence the genomic composition of the hybrid zone (Hamilton *et al.* submitted). These results provide support for different genes being targets of selection in different environments, as observed in natural populations of *Arabidopsis thaliana* (Hancock *et al.*, 2011).

SNP106 (*GAI*), with putative function in *Arabidopsis* in gibberellic acid mediated signaling (At1g14920.1) also exhibited excess interspecific differentiation in the outlier test, suggesting that this candidate gene SNP may be linked to a region under diversifying selection. Interestingly, this candidate gene SNP was one of a number of candidate genes identified from expression studies of resistance to the white pine weevil (*Pissodes strobi*) in Sitka spruce. Given the role gibberellins have in the promotion of growth and induction of mitotic division, they may have a putative role in the formation of constitutive or traumatic resin canals. Previous studies have observed significant differentiation between parental species in the formation of these defense mechanisms (O'Neill *et al.*, 2002). These outlier loci suggest future functional investigations merit further investigation. Association tests within advanced-generation introgressed individuals are needed to further test for associations between these candidates and adaptive phenotypes. Interestingly, genomic cline analysis indicates that all three outlier loci exhibit neutral patterns of introgression. However, the different approaches comparing long-term accumulation of differentiation within allopatric populations versus more contemporary influence of locus-specific selection using genomic ancestry of hybrids may contribute to these dissimilarities.

4.4.2 Geographic clines

Geographic clines show a strong spatial relationship between allele frequencies and drainage distance up river valleys across the hybrid zone. The ecological transition from maritime to continental climates is reflected at a finer spatial scale in the corresponding substitution of largely Sitka spruce alleles for white spruce alleles. While there is variation among locus-specific clines, on average the maximum rate of change estimated by the maximum slope (0.003) occurs at approximately 120 km up drainages from the coast. Interestingly, this distance is concordant with the beginning of a sharp decline in mean annual precipitation. Our previous research indicates a sharp cline in genomewide ancestry associated with precipitation, where the intensity and direction of selection may shift beyond the threshold of approximately 1000 mm of precipitation annually (Hamilton et al., submitted). The maximum rate of change in allele frequency at 120 km inland coincides with the start of a steep decline in precipitation suggesting that moisture may be a strong selective factor across this ecological gradient.

We have estimated proxies for the most commonly used cline parameters: maximum slope, indicating the maximum rate of change in allele frequency; and lambda, the geographic point on the landscape where we observe the greatest change. Evaluation of these two parameters provides some interesting trends and identifies potential candidate regions for decreased recombination. Loci that exhibit shallower clines may be candidates for neutral introgression, while steep clines may be attributed to selection against hybridization, indicating intrinsic Dobzhansky-Mueller incompatibilities or extrinsic (environmental) incompatibilities limiting introgression. Although teasing apart the influence of intrinsic or extrinsic incompatibilities would be aided by testing the fitness of controlled crosses across multiple environments, this remains beyond the scope of this study.

Five candidate gene loci (SNPs 16, 137, 142, 153, and 255, Figure 4.2) exhibited the steepest clines (maximum cline slopes of 0.011-0.012) among all loci, and therefore represent candidates for divergent selection or reproductive isolation or are linked to candidates. Only SNP16 exhibited a pattern of selection against heterozygotes (WS-), while SNPs 137, 142, and 153 exhibited positive selection for the white spruce genotypes (WW+) and SNP255 was neutral (Figure 4.2). This may indicate that SNP16 is a candidate for or linked to a locus involved in intrinsic Dobzhansky-Mueller incompatibilities resulting in hybrid inferiority or breakdown, with a low frequency of heterozygotes compared to the expectation given genomewide ancestry. SNPs 137, 142, and 153, on the other hand, all exhibited the greatest rate of change in allele

frequency between 213 and 220 km, and an excess of white spruce genotypes and lack of heterozygotes in a white spruce background (Figure 4.2). This geographic region again corresponds to the drier end of the steep gradient in precipitation that began at approximately 120km inland from the ocean (Hamilton et al. submitted). At 200 km drainage distance from the coast, all mean annual precipitation values fall below 1000 mm of annual precipitation. Indeed, an analysis of variance comparison of homozygote and heterozygote genotypes with mean annual precipitation suggests that the distribution of genotypes across the hybrid zone is differentiated based on precipitation and each genotype has a distinct distribution in terms of precipitation values ($p < 0.0001$, Figure 4.2).

Our results indicate that while Sitka spruce genotypes and heterozygote genotypes are common at a range of precipitation values exceeding 1000 mm, there appears to be a threshold below which the white spruce is the sole genotype observed. We further compared the distribution of genotypes across correlated seasonal precipitation values, mean summer precipitation and precipitation as snow. While similar trends were observed the overall patterns were weaker, suggesting mean annual precipitation likely captures the majority of differentiation among genotypes. These results suggest the distribution of alleles across the landscape may be strongly influenced by extrinsic selection in one environment, but not in another, where they may be neutral.

Previous studies with natural populations of *Arabidopsis thaliana* have shown that across steep environmental gradients, some SNPs are neutral in one environment, but adaptive in another (Fournier-Level *et al.*, 2011; Hancock *et al.*, 2011). Our results point towards a similar dynamic within the spruce hybrid zone where some SNPs appear neutral against a Sitka spruce background, but are potentially adaptive against a white spruce background where there is strong selection for the white spruce genotype. Furthermore, comparing the genomic and geographic clines of SNPs 137, 142, and 153 we observe that where there is strong evidence of selection favoring the white spruce genotype there are fewer heterozygotes, suggesting potential divergent selection across environments (Figure 4.2). Combined with the steep geographic clines suggesting strong selection for Sitka alleles within the coastal region, these findings suggest congruency between genomic and geographic clines.

4.4.3 Genomic clines

Using genomic clines, we compared the influence of locus-specific selection on introgression of individual loci into new genetic backgrounds across the hybrid zone. Overall,

the majority of loci exhibited a diversity of patterns ranging from under-representation to over-representation of certain genotypes (170/264), although the single most common category across homozygote and heterozygote genotypes was the neutral model of introgression (WW WS SS, 94/264).

The majority of loci, however, displayed a diversity of patterns inconsistent with neutral introgression. The most common categories deviating from a pattern of neutral introgression were either negative selection (WW- WS- SS+, 28/264) or positive selection for white spruce genotypes (WW+ WS+ SS-, 21/264). These loci may be candidates for adaptive introgression. In contrast, those loci that exhibit a deficiency of white spruce genotypes may be within regions of the genome involved in reproductive isolation.

While it is not possible to determine if false positives are present amongst the patterns of deviation (Teeter *et al.*, 2010), sampling design may have influenced the patterns we detected. Few individuals were sampled from the eastern end of the introgression zone and this may have impacted the observed frequency of white spruce genotypes. However, because genomic clines reflect introgression of locus-specific genotypes with respect to genomewide ancestry across all loci, the patterns reflect the rate of introgression given present ancestry. In addition, asymmetric introgression toward Sitka spruce (Hamilton *et al.* submitted) may result in selection against white spruce genotypes where strong pre- or post-zygotic barriers exist.

Patterns of introgression consistent with selection against heterozygotes (WW WS- SS) were observed in 16 of 264 markers. These loci may be within or linked to genes that result in decreased fitness of heterozygotes, although whether this is due to intrinsic or extrinsic genetic factors remains unclear. Selection against heterozygotes may indicate either simple underdominance or intrinsic Dobzhansky-Muller incompatibilities resulting from heterozygote incompatibility (Teeter *et al.*, 2010). SNP16, which had a very steep cline, exhibits a pattern of underdominance, further supporting congruence between geographic and genomic clines.

Few cases (9 out of 264) exhibited an excess of heterozygotes that could be attributed to overdominance. These loci, however, may indicate an adaptive role for hybrid genotypes (Nolte *et al.*, 2009). Transgressive patterns exhibited in hybrids for cold tolerance at moderately cold temperatures in a previous study point towards an increased fitness of hybrid genotypes (Hamilton *et al.* submitted). This lends support to the role of hybrid zones in the transfer of adaptations between species (Rieseberg and Wendel, 1993), and may have important implications in the response of admixed populations to a changing climate.

Genomic clines tell us much about locus-specific introgression patterns, although there are some caveats. While the majority of loci exhibit deviations from a neutral model of introgression, some deviations could result from genetic drift, particularly if drift occurred independently in different populations. However, it is unlikely that the influence of drift would overwhelm the influence of selection, particularly given the large population sizes throughout the hybrid zone (Nolte *et al.*, 2009). Furthermore, the number of loci in categories deviating from a pattern of neutral introgression may be over-estimated due to LD, which is expected to be substantial within hybrid zones (Lexer *et al.*, 2007). Linkage information for the loci we studied is unavailable, although we can assess pairwise patterns of LD within candidate genes where there are multiple SNPs per gene. Estimates of LD between SNPs on the same gene indicate that SNPs are significantly linked (average $D'=0.89$ across 18 genes, $p<0.0001$). In future studies, estimation of linkage relationships and the degree of recombination following first contact will be possible as genetic maps become available for spruce. Finally, integral to the genomic clines method is distinguishing neutral introgression at select loci from those under positive or negative selection. With interspecific crosses we expect more loci to have negative fitness consequences than positive ones (Barton, 2001), thus there may be a slight bias toward over-estimating positively selected loci in the hybrid zone and underestimating negatively selected ones if rates of neutral introgression are underestimated (Rieseberg *et al.*, 1999b). However, while this analysis may underestimate the rate of neutral introgression, it suggests identification of loci exhibiting negative selection is likely conservative.

Examining differential introgression using both estimates of interspecific differentiation combined with geographic and genomic clines identifies heterogeneous patterns across the genome, supporting the porous nature of the spruce genome. Candidates for divergent selection identified using outlier tests merit further investigations using functional assays in combination with genomic association tests. Furthermore, patterns of differential introgression both across the landscape and into varying genomic backgrounds identify an important role for environmental selection, indicating adaptation across precipitation gradients. From a forest management perspective, these approaches identify a suite of candidate gene SNPs for future inclusion in association and functional assays as molecular breeding becomes increasingly feasible. Furthermore, this fine-scale approach provides context for broader questions regarding the maintenance of species barriers in these long-lived species, identifying regions of the genome where selection may prevent introgression. Identifying the positions across the genome where

differences accumulate in the future will aid this research, providing an opportunity for comparison among forest tree hybrid zones for candidate regions that maintain species.

Table 4.1 Summary statistics for genomic, geographic clines and genetic differentiation (F_{ST}) across all loci. Individual loci (N) are classified based on genomic clines indicating an excess (+) or lack (-) of homozygous white spruce (WW), Sitka spruce (SS) and heterozygous (WS) genotypes given estimates of genomewide ancestry for, or as neutral. Mean and range for each class are provided for estimates of genetic differentiation (F_{ST}), slope, and lambda.

Homozygote Genomic Clines	Heterozygote Genomic Clines	N	F_{ST}	Slope	Lambda
WW+ SS-	WS+	21	0.20 (0.07-0.36)	0.004 (0.001-0.008)	116 (0-190)
	WS	14	0.20 (0.05-0.39)	0.004 (0.000-0.008)	115 (0-201)
	WS-	8	0.12 (0.03-0.44)	0.002 (0.000-0.009)	41 (0-163)
WW- SS+	WS+	8	0.23 (0.10-0.43)	0.003 (0.000-0.005)	137 (0-221)
	WS	16	0.15 (0.04-0.47)	0.001 (-0.001-0.003)	57 (0-198)
	WS-	28	0.26 (0.10-0.47)	0.003 (0.000-0.009)	144 (34-220)
WW+ SS	WS+	4	0.10 (0.08-0.12)	0.002 (0.001-0.003)	117 (77-146)
	WS	14	0.21 (0.09-0.34)	0.006 (0.000-0.012)	167 (96-231)
	WS-	3	0.10 (0.04-0.16)	0.002 (0.000-0.003)	72.98 (0-146)
WW- SS	WS+	2	0.17 (0.05-0.29)	0.002 (0.000-0.003)	127 (127-127)
WW SS+	WS+	2	0.26 (0.13-0.38)	0.004 (0.001-0.006)	70 (0-140)
	WS	1	0.19 (0.19-0.19)	0.003 (0.003-0.003)	59 (50-50)
	WS-	12	0.23 (0.10-0.47)	0.003 (0.001-0.005)	129 (0-205)
WW SS-	WS+	5	0.19 (0.07-0.25)	0.003 (0.001-0.004)	77 (1-155)
	WS	7	0.16 (0.03-0.25)	0.003 (0.000-0.005)	100 (46-119)
WW SS	WS+	9	0.16 (0.03-0.29)	0.002 (0.000-0.006)	97 (18-174)
	WS	94	0.19 (0.02-0.43)	0.003 (0.000-0.011)	118 (0-236)
	WS-	16	0.21 (0.04-0.45)	0.003 (0.000-0.011)	159 (84-215)

Table 4.2 AMOVA for 268 candidate gene SNPs between *Picea sitchensis* and *P. glauca*

Source of variation	d.f	Sum of squares	Variance components	% of variation
Between species	1	2902.52	46.97	58.70
Among individuals within species	64	2126.18	0.17	0.21
Within populations	66	2170.00	32.88	41.09
Total	131	7198.70	80.02	

Figure 4.1 Log-transformed Bayes factors and locus-specific F_{ST} estimated from *Bayescan* provide, of 268 candidate gene SNPs, identifying potential candidates for loci exhibiting diversifying (excess differentiation) or balancing (deficient differentiation) selection across the Sitka-white spruce hybrid zone. Vertical lines mark $\log_{10}(\text{BF})$ of 1 (solid) corresponds to a posterior probabilities of 0.91, beyond which provides substantial to very strong evidence in favor of selection.

Figure 4.2 – Comparison of geographic clines, (top), genomic clines (middle) and genotype distribution for SNPs exhibiting a cline width less than 100km. Genomic clines (**a**) Genomic clines indicate locus-specific patterns of introgression using the genomewide estimate of admixture (hybrid index: 0=white spruce, 1=Sitka spruce) to estimate the probability of observing a particular genotype at that locus, *P*-values are provided in the right corner of the observed data under a model of neutral introgression. The 95% confidence envelope of the probability of the homozygous white spruce genotype (dark green) and the heterozygous genotype (light green) are based on 1000 neutral parametric simulations. Fitted genomic clines are observed for the homozygous white spruce genotype (solid line) and heterozygous genotype (dashed line), while open circles indicate observed genotypes; either white spruce (WW, top), heterozygous (WS, middle) or Sitka spruce (SS, bottom). The frequency of observed genotypes are indicated on the right of the panel. Geographic clines (**b**) indicate the relationship between the frequency of the minor Sitka spruce allele with drainage distance. Geographic cline parameters indicate the proportion of variance accounted for using a linear or non-linear (logistic or exponential) regression and maximum slope and intercept (lambda, dashed line) for fitted values (solid line). Distribution of genotypic variation (**c**) indicate the frequency of each genotype with associated mean annual precipitation (mm) values exhibited across the hybrid zone.

5. Thesis conclusions

5.1 Introduction

The single largest threat to most native tree populations globally is climate change. Today's forests are facing unprecedented rates of climate change that will result in economic and environmental consequences that we are only beginning to see the effects of. Consequently, there is an urgency to quantify the threat of climate change, and increase our understanding of the ability of trees to respond to that threat. This response will, in part, depend upon the genetic architecture of populations, particularly the genetic makeup underlying phenotypic traits related to adaptation to climate.

Species distribution models and climate warming scenarios predict dramatic changes for species given current climate predictions (Hamann and Wang, 2006; Wang *et al.*, 2012a). While this does not necessarily mean the species we see on the landscape today will no longer be there, it does suggest that these individuals may become maladapted to the changing conditions, resulting in decreased average fitness of populations (Aitken *et al.*, 2008). Traditionally, to address questions regarding the underlying mechanisms contributing to local adaptation, forest managers and researchers have explored intraspecific variation. To this end, both traditional provenance trials and association tests, using genomic tools have been utilized to better understand the molecular basis of phenotypic variation, identifying candidate genes involved in adaptation to climate (González-Martínez *et al.*, 2006; Eckert *et al.*, 2009a; Eckert *et al.*, 2009b; Eckert *et al.*, 2010). The approaches used in this thesis, therefore, represent a logical next step, using both provenance trials and genomic tools to explore interspecific variation. Utilizing the variation preserved in hybrid zones, we have an increased opportunity to explore novel genetic and phenotypic variation that may be important in terms of local adaptation.

The broad goal of this thesis has been to examine the genomic and phenotypic architecture of individuals spanning the introgression zone between Sitka spruce and white spruce. Using a number of genetic tools, including maternally- and paternally-inherited molecular markers along with putatively neutral nuclear microsatellite markers I have provided a broad-scale analysis of the neutral genetic structure of the hybrid zone, inferring the influence of historic and contemporary gene flow, as well as identifying a strong role for environmental selection. These approaches are complemented and patterns verified through the use of a number of candidate gene SNPs with putative associations with growth and cold hardiness. While I observed strong associations between climate and genomic ancestry as predicted by

SNPs, associations with phenotypes were weak, indicating that much phenotypic variation has yet to be explained within the introgression zone. Phenotypic differentiation among genotypic classes, however, indicates hybrid classes are more cold tolerant at moderately cold temperatures than either parent species class. While this pattern was not retained at colder temperatures, it does suggest that hybrid classes may be transgressive for cold tolerance at moderate temperatures, pointing towards a possible selective advantage for hybrids within a specific temperature range.

Finally, taking a fine-scale approach I have examined locus-specific introgression and patterns of genetic differentiation, assessing the influence of selection within the introgression zone. The approaches used in this thesis reveal a number of candidate genes that may be targets of extrinsic or intrinsic selection within the hybrid zone over separate time-scales. F_{ST} -outlier tests between parent species reveal the influence of long-term gene flow, informing patterns of gene flow resulting from numerous generations (Lexer *et al.*, 2010), whereas genomic and geographic clines reveal more contemporary influences of gene flow given genomic ancestry and spatial variation.

From a forest management perspective, these approaches could produce a tool to aid selection of those seed sources to be used for future climates to address the broad goal of reducing the risk of planting maladapted seedlings by being able to predict local adaptation from molecular markers at a seedling stage. Current seed transfer guidelines within this region remain fairly stringent, limiting latitudinal transfer distances to 2° North and 1° South and elevational transfer to ± 200 metres (Snetsinger, 2010). However, to ensure the transfer of well-adapted seedlings, the results from this study suggest strong associations between genomic ancestry and precipitation indicate water availability may be an important factor limiting seed transfer within this region. The steep inflection point at approximately 1000 mm of precipitation observed using both neutral microsatellite and candidate gene SNP-based hybrid indices indicates that current longitudinal transfer of hybrid individuals may span a very narrow range, above and below which individuals exhibit excess Sitka- and white spruce ancestry, respectively. However, current climatic patterns for this region and predictions into the future indicate this region is likely to become warmer and wetter (Mbogga *et al.*, 2010). Therefore, changes in moisture availability throughout this region could influence limits to longitudinal seed transfer, increasing potential movement of seed from coastal climates further inland following the increase in moisture availability.

5.2 Neutral genetic structure of the Sitka-white spruce hybrid zone

I observed putatively neutral patterns of genetic structure, both for uni-parentally inherited organelle markers and bi-parentally inherited nuclear markers, that were consistent with previous observations within the Sitka and white spruce zone of introgression (Sutton *et al.*, 1991; Sutton *et al.*, 1994; Bennuah *et al.*, 2004). The distribution of chloroplast and mitochondrial haplotypes support independent colonization routes of Sitka spruce and white spruce to the Nass-Skeena region following the Pleistocene (Soltis *et al.*, 1997; Anderson *et al.*, 2006; de Lafontaine *et al.*, 2010; Shafer *et al.*, 2010). The evidence I have presented in this thesis suggests the arrival of Sitka spruce in advance of white spruce resulted in a ‘mitochondrial capture’ of the Sitka spruce mitotype throughout the hybrid zone. Combined with geographic structuring of chloroplast haplotypes reflecting pollen-mediated gene flow, which unlike mtDNA haplotypes were not fixed throughout the hybrid zone, these patterns suggest introgression may be unidirectional.

Pre-zygotic barriers resulting from differentiation in reproductive phenology may contribute to patterns observed, however this does not preclude possible post-zygotic barriers resulting from intrinsic genetic incompatibilities. A formal test for differences in reproductive phenology, evaluating heat sum requirements for Sitka spruce, white spruce and hybrids in a common garden setting, may clarify the influence of this putative barrier. This may be possible within the current ten-year old common gardens planted within the hybrid zone. As trees become reproductively mature and begin to produce pollen and seed cones associations may be made between genotypic classes already assessed and timing of reproductive and vegetative phenology.

While the organelle genomes tease apart the long-term influence of both seed-mediated and pollen-mediated gene flow within the introgression zone, microsatellite variation reveals gene flow within the bi-parentally inherited nuclear genome. Strong associations between distance and climate with neutral genetic structure within the hybrid zone were consistent with associations observed for morphological variation, identifying a common role for distance and climatic variables in the putatively neutral structure of the hybrid zone. Position along the ecological transition from maritime to continental climates consistently predicted both genetic ancestry based on microsatellites and morphology of cone scales and bracts, pointing towards a strong role for extrinsic selection. Replicated transects along the Nass and Skeena rivers reveal congruent patterns, particularly with drainage distance and mean annual precipitation. The

concordance of these genetic and morphological clines point towards a hybrid zone maintained by selection along an ecological gradient (Brennan *et al.*, 2009).

5.3 Broad-scale genomic and phenotypic architecture of the spruce hybrid zone

The genomic architecture of the spruce hybrid zone as revealed by candidate gene SNPs was comparable to that based on separate studies using putatively neutral microsatellite markers within this thesis, and expressed sequence tag markers in an earlier study (Bennuah *et al.*, 2004). Associations between genomic ancestry as revealed by hybrid index with climate and distance variables indicate a strong role for environmental selection within the hybrid zone. The ecological transition from the maritime climate of Sitka spruce to the continental climate of white spruce correlated with increasing drainage distance up river valleys consistently predicted ancestry, similar to the pattern observed for neutral genetic markers. This is consistent with previous research where extrinsic selection has been identified as playing a significant role in hybrid zone maintenance across ecotonal regions (Moore, 1977; Wang *et al.*, 1997; Yanchukov *et al.*, 2006; Culumber *et al.*, 2011). While I have not been able to directly test the bounded hybrid superiority model of hybrid zone maintenance, the evidence presented in this thesis suggests an interaction between genomic ancestry and environment. However, these results, combined with patterns of asymmetric introgression consistent across studies, suggest that there may be intrinsic factors influencing the genomic architecture of the hybrid, in addition to extrinsic factors. More and more studies are pointing towards the combined influences of extrinsic and intrinsic selection maintaining hybrid zones emphasizing the need for reciprocal transplants combined with interspecific crosses to examine fitness across environments (Crespin *et al.*, 2002; Miglia *et al.*, 2005; Brennan *et al.*, 2009; Field *et al.*, 2010).

Surprisingly, only weak interactions were observed between putatively adaptive phenotypic traits and environment across the hybrid zone. Previous research examining intraspecific variation in adaptive phenotypic traits with climatic gradients has described strong associations between phenotypes and environment, indicating local adaptation (Howe *et al.*, 2003; Mimura and Aitken, 2007; Savolainen *et al.*, 2007; Holliday *et al.*, 2010a). However, we expect greater variation in fitness across hybrid classes than within species depending upon the influence of epistasis, dominance, or maternal effects between the species (Rieseberg and Carney, 1998; Burgess and Husband, 2006; Arnold *et al.*, 2008). Furthermore, while

intraspecific studies describe wide temperature gradients associated with growing season resulting in height and phenological differentiation (Mimura and Aitken, 2007; Savolainen *et al.*, 2007; Holliday *et al.*, 2010a), precipitation appears to be a limiting factor along the gradient spanning the Nass-Skeena region. Steep clinal gradients in ancestry associated with precipitation support previous observations of differentiation in water-use efficiency among Sitka spruce, white spruce and hybrids (Silim *et al.*, 2001). This points towards the importance of precipitation and raises the possibility of different physiological adaptations between parent species and hybrids that act to maintain hybrid zone structure (Kimball and Campbell, 2009). While all of British Columbia is predicted to become warmer over this century, this region is also predicted to become wetter, unlike some others (Mbogga *et al.*, 2010). If the present structure of the hybrid zone is maintained by adaptation to precipitation gradients, given current climate warming predictions the structure of the hybrid zone may be impacted, leading to hybrid zone movement (Buggs, 2007). While this may be a very preliminary hypothesis, there is increasing empirical evidence supporting hybrid zone movement driven by climate change (Buggs, 2007; Walls, 2009; Carson *et al.*, 2012). This will likely influence the current direction and extent of seed transfer within the hybrid zone and into the future.

Identification of genotypic classes provided the opportunity to compare phenotypic traits between recombinant types to both test assumptions of the bounded hybrid superiority model and identify potential transgressive traits. Transgressive traits are relatively common across plant taxa (Rieseberg *et al.*, 1999a). Cold tolerance appears to be a positively transgressive trait across all hybrid genotypic classes at moderate, but not at colder temperatures. Teasing apart the complex interactions between genes involved in cold adaptation contributing to transgression may require transcriptional assays during cold acclimation across genotypic classes. Dissection of the genes expressed during adaptation to moderate temperatures versus colder temperatures by genotypic class may identify gene families or linked genes that exhibit functional roles in cold adaptation, revealing the genes contributing to transgressive traits.

5.4 Fine-scale patterns of differential introgression

Examination of locus-specific introgression between Sitka and white spruce, combining outlier tests with geographic and genomic clines, revealed differential patterns of introgression for different loci. The goal of this fine-scale examination was to detect candidate gene loci that may be important in terms of local adaptation and the maintenance of species barriers. This builds from previous studies where outlier tests (Scotti-Saintagne *et al.*, 2004; Yatabe *et al.*,

2007; Lexer *et al.*, 2010; Zeng *et al.*, 2010), geographic clines (Porter *et al.*, 1997; Carling and Brumfield, 2009; Gompert *et al.*, 2010) or genomic clines (Nolte *et al.*, 2009; Teeter *et al.*, 2010) have been used to identify candidate loci that may be under selection or linked to regions of the genome that are under selection. The novelty of this thesis lies in a combined approach. Patterns of interspecific genetic differentiation yield a longer accumulation of differentiation between species and clinal approaches reveal more contemporary influences of selection allowing assessment of patterns of differentiation across time-scales,

Outlier tests in other hybrid zones have typically revealed few candidates for divergent or balancing selection between species, although a number of studies to-date have been based on putatively neutral microsatellite markers. Yatabe *et al.* (2007) observed five EST-associated microsatellites (5% of total) that exhibited an excess of interspecific differentiation between the self-compatible annuals, *Helianthus annuus* and *H. petiolaris* based on the approach advocated by Beaumont and Nichols (1996). I observed three candidate gene SNPs (1% of total) that were candidates for divergent selection. These values appear low given other interspecific studies examining differentiation in forest trees, *Quercus* and *Populus*, although methods regarding the molecular markers used and how they were selected vary among studies. Using a wide array of molecular markers Scotti-Saintagne *et al.* (2004) identified 12% of the loci as outliers, whereas a similar comparison in *Populus* revealed 35% of the loci examined had estimates of genetic differentiation exceeding that expected under the neutral model of evolution (Lexer *et al.*, 2010).

My data point towards weak barriers to gene flow between Sitka and white spruce, and long-term persistent introgression of species that may have come into contact repeatedly over long evolutionary time scales. Ancestral polymorphisms may contribute to the patterns presented in this thesis as both Sitka spruce and white spruce are closely related based on phylogenetic assessment (Ran *et al.*, 2006). Although candidate gene SNP selection based on $F_{ST} > 0.40$ should have biased our assessment to SNPs to those under divergent selection, comparison with similar studies suggest that our results may be conservative. In the future as linkage map information becomes accessible for spruce we may be able to evaluate the recombination history of the hybrid zone. Evaluating the ancestry of genomic blocks will aid in identifying whether variation within this hybrid zone is consistent with interspecific recombination or ancestral polymorphisms (Lexer *et al.*, 2010), and also provide information on the duration of secondary contact between these species. Furthermore, while outlier tests provide direction with respect to potential targets of natural selection, association with relevant phenotypes remains an area to be addressed.

Geographic and genomic clines provide a unique perspective of contemporary introgression between Sitka spruce and white spruce, revealing the influence of selection at both the landscape and genomic level (Payseur, 2010). Clinal congruence points toward the influence of extrinsic or intrinsic selection at a fine scale. Few studies have combined methods, however, Teeter *et al.* (2010) examined differential patterns of introgression using both genomic and geographic clines across replicate transects between *Mus musculus* and *M. domesticus*. While they identified relatively few candidate gene SNPs with narrow clines (12%), they suggested these regions may have roles in reproductive isolation between these mice species. Furthermore, of those loci that exhibited narrow cline width, three had similar patterns of introgression across replicate transects identified using the genomic clines method (Teeter *et al.*, 2010). This study varies from my thesis research, however, as it emphasized the role of intrinsic selection, in particular the Dobzhansky-Muller model of reproductive isolation, and identified no associations between geography and the position of the hybrid zone (Teeter *et al.*, 2010). In my own study I cannot exclude the possibility that intrinsic selection and Dobzhansky-Muller incompatibilities contribute to the genetic structure of the Sitka-white spruce hybrid zone. However, the results I have presented throughout this dissertation emphasize extrinsic factors, specifically the importance of geography and climate across this ecological transition zone.

My results based on hundreds of SNPs are comparable to previous studies using thousands of SNPs that have examined adaptation across environments within *Arabidopsis thaliana*, identifying adaptive SNPs in one environment that are neutral in another (Fournier-Level *et al.*, 2011; Hancock *et al.*, 2011). These studies present major advances to identifying those genes involved in adaptation to climate within species (Savolainen, 2011). Using the genomic clines method in the present thesis, I identify candidate gene SNPs that may be under selection within a Sitka spruce genomic background associated with a maritime climate, or a white spruce genomic background associated with a continental climate. While this method is fairly preliminary compared to those achieved in model organisms with shorter generation times, my results support the finding in *Arabidopsis* that different genes may be targets of selection in different environments, and an adaptive polymorphism in one environment may be neutral in another (Hancock *et al.*, 2011).

5.5 Limitations of the current study

While this study identifies potential forest management implications and practical applications with regards to direction of seed transfer and environment associations, there are

limitations to working with a long-lived non-model organism. One of these limitations is disentangling the influence of pre-zygotic and post-zygotic barriers to reproduction. Evidence of asymmetric introgression suggests that there are likely barriers to reproduction between Sitka spruce and white spruce. While I have inferred phenological differences as a likely pre-zygotic barrier, a formal test using the common garden as individuals become reproductively mature would be required to confirm this. Surveys of phenological stages in situ would be informative, as phenology is influenced by both genetics and environment. While the opportunity for crossing within a single environment will be largely genetically determined, opportunities for crossing of individuals in different populations will be affected by both the genetics of the individual, and the environments of the populations.

However, identifying post-zygotic barriers, be they extrinsic or intrinsic remains problematic. In species with shorter time to reproductive maturity, artificial crosses could generate a range of genotypic classes, including F1s, F2s, and beyond to identify putative extrinsic or intrinsic post-zygotic barriers (Rieseberg and Blackman, 2010). Furthermore, the availability of recombinant classes formed through controlled crossings would permit evaluation of fitness in a range of environments (Kimball and Campbell, 2009). It would be valuable to assess the fitness of an F1-generation in a range of controlled or field environments to evaluate the possibility of incorporating hybrids into tree breeding programs. Bearing in mind that F1s may exhibit heterosis for phenotypic traits (Rieseberg *et al.*, 1999a), however, it will be equally important to evaluate fitness of advanced generation crosses for potential hybrid breakdown in F2 generations and beyond.

Formal testing of the bounded hybrid superiority model requires testing parent species and recombinant classes using reciprocal transplants into the native habitat of both parental species and hybrids (Arnold and Hodges, 1995). The common garden used in this experiment was initially established to measure traits of economic value and pest-resistance at multiple sites within the hybrid zone, but not within parental habitats. Given this, I have been limited in my ability to formally test the bounded hybrid superiority model of hybrid zone maintenance.

This study has also been limited by sampling design of individuals within the pre-existing common garden. The sampling design of the common garden experiment reflects an initial goal to compare local adaptation and productivity among seed sources for reforestation within the hybrid zone. Based on this goal, sampling was focused within the hybrid zone and to a lesser extent its periphery and beyond. This resulted in a lack of individuals sampled to the colder eastern (white spruce) end of the introgression zone. Furthermore, eastern sampling of

populations was limited to avoid possible influence of a third high-elevational spruce species, Engelmann spruce, which hybridizes with white spruce in mountainous regions of eastern British Columbia. This resulted in the use of FN and ENA populations, well-separated from the hybrid zone, representing pure white spruce reference populations in the first and final two data chapters, respectively.

The common garden experiment within the hybrid zone permitted evaluation of phenotypic traits important to adaptation in conifers, however, this study would benefit from the inclusion of additional phenotypic traits, replication of select traits, and traits that in retrospect would have been useful. Bud phenology (bud break and bud set) within the common garden would be a useful additional trait to explore for possible differentiation in frost-free growth periods among genotypic classes, as differentiation has been observed within intraspecific studies (Mimura and Aitken, 2007; Holliday *et al.*, 2010a). In addition, replication of cold hardiness experiments at different time periods during cold acclimation using additional temperatures could provide a finer-scale evaluation of cold tolerance of individual genotypic classes. This could elaborate on the current observation of transgression for cold tolerance for hybrid individuals at moderately cold temperatures, but not at colder temperatures. Finally, in retrospect, we have identified the important influence of water availability in the structure of the hybrid zone. Stable carbon isotope ratios ($\delta^{13}\text{C}$) may be used in the future to measure water use efficiency, testing for variation in photosynthetic capacity of Sitka spruce, white spruce and controlled hybrid crosses in a controlled setting (Sun *et al.*, 1996; Silim *et al.*, 2001).

This study would benefit greatly from genetic map information for the candidate SNPs used. Although there is a genetic mapping study in *Picea glauca* currently underway for a number of the candidate gene SNPs used in this study (J. Bousquet, N. Isabel and B. Pelgas unpublished data), these data have yet to become available, and not all SNPs will be included on the genetic map. Linkage is expected to be substantial in hybrid zones (Lexer *et al.*, 2007). The evaluation of linkage relationships can be used to estimate degree of recombination and address questions regarding the presence of ancestral polymorphisms versus consistent interspecific recombination (Lexer *et al.*, 2006). Furthermore, coarse-scale genetic maps providing a location for loci across linkage groups will aid in identifying regions of the genome that exhibit extreme levels of differentiation, and therefore may be involved in reproductive isolation and important in terms of the maintenance of species barriers (Lexer *et al.*, 2010).

One of the limitations to working in conifers is that, as of yet, there has been no conifer genome sequenced. While a number of projects are underway to sequence the genome in *Picea*

abies, *Pinus taeda* and *Picea glauca*, large genome sizes ranging from 21 to 40 Gb, much larger than the closest related reference in *Populus* (485 Mb, (Neale and Kremer, 2011), this remains a limitation within this study. The evolutionary distance separating conifers from angiosperms is approximately 300 million years, presenting additional genome complexity (Ritland, 2012). Within this context, functional orthologs for candidate gene SNPs in conifers are identified from model systems, most commonly with *Arabidopsis*. This remains a concern in the present analysis, and therefore caution must be taken when making functional inferences regarding the candidate gene SNPs.

While I have surveyed variation in a number of molecular markers, this is only a small fraction of the spruce genome. With the rapid advancement of new ‘next-generation’ sequencing technologies whole genome scans offer an increased opportunity to explore differential introgression with increased resolution at a fraction of the cost (Ritland, 2012). The value of the present study, with respect to candidate gene SNP selection, is that it has targeted candidate genes expressed in previous studies of *Picea* with putative adaptive trait function. However, patterns of differentiation in forest trees are likely a result of many loci of small effect (Kremer *et al.*, 2012), suggesting that increasing the density of markers with new technologies, such as genotyping by sequencing (GBS) or restriction site associated DNA (RAD) tags, will aid in identifying those regions of the genome responsible for adaptive differentiation and the maintenance of species barriers (Neale and Kremer, 2011).

5.6 Future research

This study adds to our understanding of the genetic basis underlying local adaptation within conifer hybrid zones by identifying genotypes with appropriate climate-related genetic variation that may be suitable for replanting in the ecotonal area described herein, informing direction of seed transfer, and identifying a suite of candidate genes that may be used in ecologically relevant climate-based seed management. While this research represents a significant step forward in terms of characterizing the dynamic interactions between gene flow and natural selection within this spruce hybrid zone, and bridges the gap between previous studies that have assessed intraspecific variation (Mimura and Aitken, 2007; Holliday *et al.*, 2008; Namroud *et al.*, 2008; Holliday *et al.*, 2010a), with increasing technological advances future opportunities will allow us to expand upon the present questions among others at different magnitudes of scale .

5.6.1 *Test for asymmetrical barriers to genetic exchange*

The establishment of long-term common garden experiments within the spruce introgression zone offers the opportunity not only to evaluate fitness of these long-lived species directly in the field, but will also offer the opportunity to address questions regarding recombination over multiple generations once trees become reproductively mature. The average age at sexual maturity in *Picea* is approximately 14.5 years (Verdu, 2002). The common garden was initially established in 1998, and anecdotal evidence suggests a few individuals are beginning to produce reproductive seed and pollen cones (Hamilton pers. obs.), providing the next generation of hybrid seed. This resource will facilitate the evaluation of putative intrinsic versus extrinsic mechanisms resulting in asymmetrical barriers to gene exchange.

There are a number of approaches that could be used, however a relatively simple first step would be to evaluate seed production of individuals across the observed genotypic classes (Sitka, white, BCS, BCW and FN). Using the 31 SNPs that exhibited an allele frequency differential greater than 0.90 should readily be able to identify genotypic classes based on methods from Chapter 3 of seed produced by individuals from the common garden. Following from this seed collected could be evaluated for viability using germination tests or evaluating seed size and other correlates to seed viability (Tiffin *et al.*, 2001). To differentiate the maternal and paternal contribution to subsequent generations we may take a molecular approach, evaluating the distribution of nuclear markers between parent and progeny. Teasing apart the distribution of chloroplast haplotypes between parents and progeny may point towards the most frequent pollen donor class, which likely contributes to population structure. The value of having reproductively mature Sitka spruce and white spruce individuals within the common garden also offers the opportunity to make controlled crosses in the field and evaluate post-pollination pre-zygotic barriers. The products or lack thereof of controlled crosses may allow us to evaluate whether asymmetric gene flow is a result of pre-pollination pre-zygotic barriers (e.g. temporal isolation), or conspecific pollen precedence and unilateral incompatibilities resulting from post-pollination isolating mechanisms (Rieseberg and Blackman, 2010).

5.6.2 *Gene expression and association analysis with drought tolerance*

With the current rapid rate of advancement in molecular technologies, including high-throughput genotyping and sequencing, evaluating the complex architecture of adaptation in non-model organisms has become feasible (Brunner *et al.*, 2007; Neale and Kremer, 2011). Interesting outcomes from the present study worth investigating further include the strong clinal

gradients and significant genotypic differentiation associated with precipitation gradients. Previous research has characterized water-use efficiency (WUE) in Sitka spruce, white spruce and their hybrids, noting highest WUE of white spruce followed by hybrids under low-water availability (Silim *et al.*, 2001). These observations are consistent with adaptation of Sitka and white spruce to maritime and continental climates, respectively. However, understanding the underlying physiological mechanisms contributing to adaptation to moisture would benefit from both transcriptome and association analyses. Consistent patterns observed between hybrids and reference parent species may point towards candidate genes or physiological responses relevant to intraspecific studies. This analysis could be aided by evaluating specific genotypic classes (Sitka, White, BCW, BCS, FN, and F1 if present from progeny in common garden) to compare physiological responses to low-water conditions within multiple controlled growth chamber environments.

To assess the expression of genes involved in drought stress response, gene expression could be assessed over an appropriate time frame. Advances in whole transcriptome sequencing facilitate such studies, comparing gene expression in seedlings grown under different conditions. The advantage to this approach is the comparison between genotypic classes where it may be possible to identify genotypic class-specific patterns of gene expression. Much work has been done evaluating the response of trees to drought conditions (Street *et al.*, 2006; Hamanishi *et al.*, 2010; Hamanishi and Campbell, 2011). These studies have identified many candidate genes with putative roles in drought response, including dehydrins, heat-shock proteins, and genes with functional roles in photosynthesis, carbohydrate metabolism, cell wall synthesis and plant defense (Hamanishi and Campbell, 2011). Parent and hybrid individuals could be sequenced for genes with a putative role in drought tolerance based on annotation or gene expression. Where polymorphisms are observed, associations with phenotypic or physiological responses can be tested. For example, Gonzelez-Martinez *et al.* (2006) tested for signatures of selection in 18 candidate genes related to drought-stress response in *Pinus taeda*. They identified two potentially adaptive genes, *early-response-to-drought-3 (erd3)* gene and *Caffeoyl-CoA-O-methyltransferase (ccoamt-1)*. Interestingly, in our own study, SNP255, which exhibited one of the steepest geographic clines has homology to *early-response-to-drought-6* gene (*erd6*, AT1G08930.2) in *Arabidopsis*, suggesting that this gene may be important in terms of drought response across species. New targeted gene capture technologies, such as Agilent SureSelect (Agilent, 2011), may in the future permit resequencing of whole genes capturing all SNPs within

them, providing us with a wealth of information on gene diversity and the influence of that diversity on adaptation (Grover *et al.*, 2012).

Ongoing research within the *Pinus contorta* - *Pinus banksiana* hybrid zone indicates that water deficit influences gene expression associated with formation of constitutive and induced defenses to mountain pine beetle (J. Cooke, personal communication). One of the major limiting factors to re-planting Sitka spruce is its susceptibility to the white pine shoot tip weevil. Previous research indicates that under controlled conditions, the spruce hybrids produce an intermediate defense response relative to parent species, less susceptible than Sitka spruce, but more susceptible than white spruce (O'Neill *et al.*, 2002). While there are large differences between the effects of mountain pine beetle and white pine shoot tip weevil on their respective tree hosts, it may be of interest to evaluate whether low-water conditions, strongly associated with more resistant white spruce genotypes, influences constitutive or induced defense responses to the weevil.

5.6.3 Predicting phenotypes using genomic ancestry

In Chapter 4 I presented a precursor to admixture mapping, in which I used genomic clines to evaluate introgression of individual genotypes against genomewide ancestry. The next step in this analysis will be to evaluate whether locus-specific ancestries predict phenotypes. This may be possible combining outputs from *Introgres*, where ancestry is defined as 0=Sitka spruce, 1=heterozygote or 2=white spruce using linear regression with adaptive phenotypes (e.g. cold injury, height). Of particular interest will be to examine whether the addition of an allele from either parent (producing a heterozygote) appear transgressive. However, because of the asymmetry of introgression, there may be an excess of Sitka spruce alleles, therefore an appropriate question to ask may be whether the addition of a white spruce allele compared to pure Sitka spruce significantly influences phenotypes. Following from this, evaluation of candidate gene function of those SNPs that exhibit differentiation for adaptive phenotypes based on ancestry may point towards candidate gene families or common physiological functions important in terms of adaptation and worth pursuing in future studies.

5.6.4 Genomewide resequencing and clinal variation

With cost limitations to whole genome resequencing decreasing, in the future it will be possible to map maximal slope along individual linkage groups, as in Carling and Brumfield (2009) to compare introgression by genomic location on a much finer scale. This approach will

be particularly useful in identifying those regions of the genome that exhibit reduced introgression, identifying potential candidate gene regions involved in reproductive isolation (Payseur, 2010). Given the spruce genome is over 20 billion basepairs, the current dataset will likely provide the opportunity to test the feasibility of this approach. Identifying locus-specific patterns contributing to interspecific variation along linkage groups depends on the locus coverage and density along linkage groups (Buerkle and Lexer, 2008). As genetic map information becomes available, patterns presented here may direct the focus of future efforts to those linkage groups, through a combination of targeted gene capture and next-generation sequencing approaches, that harbour candidates exhibiting greater differentiation or reduced gene flow, as in the sex chromosomes of the *Passerina* and *Mus* hybrid zones (Macholán *et al.*, 2007; Carling and Brumfield, 2009; Teeter *et al.*, 2010). As the capacity for whole genome resequencing increases, the scale at which we examine the transfer of genetic material between species will become increasingly fine. These technologies are leading to association of landscape-level processes to gene-level responses on a scale that is transforming the field of landscape genomics.

5.6.5 Conclusions

As molecular approaches evolve, our understanding of the genomic architecture underlying adaptive phenotypic traits will similarly evolve. The application of new cost-effective sequencing technologies opens the door to a wealth of information on a scale that until only recently was not feasible in non-model systems (Neale and Kremer, 2011). Application of these approaches within hybrid zones will continue to provide useful insight into the genetic architecture of important traits involved in adaptation and those involved in the maintenance of species barriers. Combined, these advanced technical approaches and large collaborative research efforts point towards a bright future for teasing apart the genomic complexity of phenotypic traits important to local adaptation in forest tree species and their hybrids.

References

- A'Hara SW, Cottrell JE. 2004.** A set of microsatellite markers for use in Sitka spruce (*Picea sitchensis*) developed from *Picea glauca* ESTs. *Molecular Ecology Notes* **4**(4): 659-663.
- A'Hara SW, Cottrell JE. 2007.** Characterization of a suite of 40 EST-derived microsatellite markers for us in Sitka spruce (*Picea sitchensis* (Bong.) Carr). *Silvae Genetica* **56**(3-4): 138-141.
- Agilent 2011.** SureSelect Target Enrichment System Kits.In.
- Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S. 2008.** Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications* **1**(1): 95-111.
- Allendorf FW, Leary RF, Spruell P, Wenburg JK. 2001.** The problems with hybrids: setting conservation guidelines. *Trends in Ecology and Evolution* **16**(11): 613-622.
- Anderson EC, Thompson EA. 2002.** A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* **160**: 1217-1229.
- Anderson LL, Hu FS, Nelson DM, Petit RJ, Paige KN. 2006.** Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. *Proceedings of the National Academy of Sciences* **103**(33): 12447-12450.
- Arnold ML. 1997.** *Natural hybridization and evolution*. New York: Oxford University Press.
- Arnold ML, Cornman RS, Martin NH. 2008.** Hybridization, hybrid fitness and the evolution of adaptations. *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology* **142**(1): 166-171.
- Arnold ML, Hodges SA. 1995.** Are natural hybrids fit or unfit relative to their parents. *Trends in Ecology and Evolution* **10**(2): 67-71.
- Barrett R, Schluter D. 2008.** Adaptation from standing genetic variation. *Trends in Ecology and Evolution* **23**(1): 38-44.
- Barton N, Gale KS 1993.** Genetic analysis of hybrid zones. In: Harrison R ed. *Hybrid zones and the evolutionary process*. New York: Oxford University Press.
- Barton N, Hewitt GM. 1981.** A chromosomal cline in the grasshopper *Podisma pedestris*. *Evolution* **35**: 1008-1018.
- Barton NH. 2001.** The role of hybridization in evolution. *Molecular Ecology* **10**: 551-568.
- Barton NH, Hewitt GM. 1985.** Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16**: 113-148.
- Beaumont MA, Balding DJ. 2004.** Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13**(4): 969-980.
- Beaumont MA, Nichols RA. 1996.** Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London, B* **263**: 1619-1626.
- Bennuah SY, Wang T, Aitken SN. 2004.** Genetic analysis of the *Picea sitchensis* x *glauca* introgression zone in British Columbia. *Forest Ecology and Management* **197**(1-3): 65-77.
- Bierne N, Daguin C, Bonhomme F, David P, Borsa P. 2003.** Direct selection on allozymes is not required to explain heterogeneity among marker loci across a *Mytilus* hybrid zone. *Molecular Ecology* **12**: 2505-2510.
- Bierne N, Welch J, Loire E, Bonhomme F, David P. 2011.** The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology* **20**: 2044-2072.

- Bomblies K, Weigel D. 2007.** Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nature Reviews Genetics* **8**(5): 382-393.
- Brennan AC, Bridle JR, Wang A-L, Hiscock SJ, Abbott RJ. 2009.** Adaptation and selection in the *Senecio* (Asteraceae) hybrid zone on Mount Etna, Sicily. *New Phytologist* **183**(3): 702-717.
- Bridle JR, Vines TH. 2006.** Limits to evolution at range margins: when and why does adaptation fail? *Trends in Ecology and Evolution* **22**(3): 140-147.
- Brunner A, DiFazio SP, Groover AT. 2007.** Forest genomics grows up and branches out. *New Phytologist* **173**(710-713).
- Buerkle C, Lexer C. 2008.** Admixture as the basis for genetic mapping. *Trends in Ecology and Evolution* **23**(12): 686-694.
- Buerkle CA. 2005.** Maximum-likelihood estimation of a hybrid index based on molecular markers. *Molecular Ecology Notes* **5**(3): 684-687.
- Buggs RJA. 2007.** Empirical study of hybrid zone movement. *Heredity* **99**: 301-312.
- Burgess KS, Husband BC. 2006.** Habitat differentiation and the ecological costs of hybridization: the effects of introduced mulberry (*Morus alba*) on a native congener (*M. rubra*). *Journal of Ecology* **94**(6): 1061-1069.
- Campbell D, Bernatchez L. 2004.** Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Molecular Biology and Evolution* **21**(5): 945-956.
- Campbell DR, Waser NM. 2001.** Genotype-by-environment interaction and the fitness of plant hybrids in the wild. *Evolution* **55**(4): 669-676.
- Cannell MGR, Smith RI. 1983.** Thermal time, chill days and prediction of budburst in *Picea sitchensis*. *Journal of Applied Ecology* **20**(3): 951-963.
- Carling MD, Brumfield RT. 2009.** Speciation in *Passerina* buntings: introgression patterns of sex-linked loci identify a candidate gene region for reproductive isolation. *Molecular Ecology* **18**(5): 834-847.
- Carney SE, Wolfe DE, Rieseberg LH. 2000.** *Hybridisation and forest conservation*. Australia: CSIRO Publishing Group.
- Carson EW, Tobler M, Minckley WL, Ainsworth RJ, Dowling TE. 2012.** Relationships between spatio-temporal environmental and genetic variation reveal an important influence of exogenous selection in a pupfish hybrid zone. *Molecular Ecology* **21**: 1209-1222.
- Chapuis M-P, Estoup A. 2007.** Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* **24**(3): 621-631.
- Chen M, Markham JE, Cahoon EB. 2012.** Sphingolipic 8 unsaturation is important for glucosylceramide biosynthesis and low-temperature performance in Arabidopsis. *The Plant Journal* **69**: 769-781.
- Chourmouzis C, Yanchuk AD, Hamann A, Smets P, Aitken S 2009.** Forest tree genetic conservation status report 1: In situ conservation status of all indigenous British Columbia species. In. Victoria: Centre for Forest Conservation Genetics, Forest Genetics Council of BC, and BC Ministry of Forest and Range, Forest Science Program.
- Crespin L, Berrebi P, Lebreton J-D. 2002.** Spatially varying natural selection in a fish hybrid zone. *Journal of Fish Biology* **61**(696-711).
- Culumber ZW, Fisher HS, Tobler M, Mateos M, Barber PH, Sorenson MD, Rosenthal GG. 2011.** Replicated hybrid zones of *Xiphophorus* swordtails along an elevational gradient. *Molecular Ecology* **20**: 342-356.

- Currat M, Ruedi M, Petit RJ, Excoffier L. 2008.** The hidden side of invasions: massive introgression by local genes. *Evolution* **62**(8): 1908-1920.
- Daubenmire R. 1967.** Some geographic variations in *Picea sitchensis* and their ecologic interpretation. *Canadian Journal of Botany* **46**: 787-798.
- Davis MB, Shaw RG. 2001.** Range shifts and adaptive responses to Quaternary climate change. *Science* **292**(5517): 673-679.
- De Carvalho D, Ingvarsson PK, Joseph JA, Suter L, Sedivy C, Macaya-Sanz D, Cottrell JE, Heinze B, Schanzer I, Lexer C. 2010.** Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree. *Molecular Ecology* **19**: 1638-1650.
- de Lafontaine G, Turgeon J, Payette S. 2010.** Phylogeography of white spruce (*Picea glauca*) in eastern North America reveals contrasting ecological trajectories. *Journal of Biogeography* **37**: 741-751.
- Doyle JJ, Doyle JL. 1990.** Isolation of plant DNA from fresh tissue. *Focus* **12**: 13-15.
- Dumolin-Lapegue S, Demesure B, Fineschi S, Le Corre V, Petit RJ. 1997.** Phylogeographic structure of white oaks throughout the European continent. *Genetics* **146**: 1475-1487.
- Dungey HS. 2001.** Pine hybrids - a review of their use performance and genetics. *Forest Ecology and Management* **148**: 243-258.
- Eckert AJ, Bower AD, González-Martínez SC, Wegrzyn JL, Coop G, Neale DB. 2010.** Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). *Molecular Ecology* **19**: 3789-3805.
- Eckert AJ, Bower AD, Wegrzyn JL, Pande B, Jermstad KD, Krutovsky KV, St. Clair B, Neale D. 2009a.** Association genetics of coastal Douglas Fir (*Pseudotsuga menziesii* var *menziesii*, Pinaceae). I. Cold-hardiness related traits. *Genetics* **182**: 1289-1302.
- Eckert AJ, Pande B, Ersoz ES, Wright MH, Rashbrook VK, Nicolet CM, Neale DB. 2009b.** High-throughput genotyping and mapping of single nucleotide polymorphisms in loblolly pine (*Pinus taeda* L.). *Tree Genetics and Genomes* **5**: 225-234.
- Emms SK, Arnold ML. 1997.** The effect of habitat on parental and hybrid fitness: transplant experiments with Louisiana Irises. *Evolution* **51**(4): 1112-1119.
- Endler JA. 1977.** Geographic variation, speciation, and clines. *Monographs in Population Biology*.
- Excoffier L, Hofer T, Foll M. 2009.** Detecting loci under selection in a hierarchically structured population. *Heredity* **103**(4): 285-298.
- Excoffier L, Laval G, Schneider S. 2005.** Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**: 47-50.
- Falush D, Stephens M, Pritchard JK. 2007.** Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* **7**(4): 574-578.
- Fan J, Gunderson K, Bibikova M, Yeakley J, Chen J, Wickhamgarcia E, Lebruska L, Laurent M, Shen R, Barker D. 2006.** Illumina Universal Bead Arrays. *Methods in Enzymology* **410**: 57-73.
- Fan S, Grossnickle SC, Sutton BCS. 1997.** Relationships between gas exchange adaptation of Sitka x interior spruce genotypes and ribosomal DNA markers. *Tree Physiology* **17**: 115-123.
- Field DL, Ayre DJ, Whelan RJ, Young AG. 2010.** Patterns of hybridization and asymmetrical gene flow in hybrid zones of the rare *Eucalyptus aggregata* and common *E. rubida*. *Heredity* **106**(5): 841-853.

- Field DL, Ayre DJ, Whelan RJ, Young AG. 2011.** The importance of pre-mating barriers and the local demographic context for contemporary mating patterns in hybrid zones of *Eucalyptus aggregata* and *Eucalyptus rubida*. *Molecular Ecology* **20**(11): 2367-2379.
- Flint HL, Boyce BR, Beattie DJ. 1967.** Index of injury - a useful expression of freezing injury to plant tissues as determined by the electrolytic method. *Canadian Journal of Plant Science* **47**: 229-230.
- Foll M, Gaggiotti O. 2008.** A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a bayesian perspective *Genetics* **180**: 977-993.
- Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM. 2011.** A map of local adaptation in *Arabidopsis thaliana*. *Science* **334**(6052): 86-89.
- Freeman JS, Jackson HD, Steane DA, McKinnon GE, Dutkowski GW, Potts BM, Vaillancourt RE. 2001.** Chloroplast DNA phylogeography of *Eucalyptus globulus*. *Australian Journal of Botany* **49**: 585-596.
- Friedmann M, Ralph S, Aeschliman D, Zhuang J, Ritland K, Ellis BE, Bohlmann J, Douglas CJ. 2007.** Microarray gene expression profiling of the developmental transition from primary to secondary growth and constitutive defense in Sitka spruce (*Picea sitchensis*) apical shoots. *Journal of Experimental Botany* **58**: 593-614.
- Fritsche F, Kaltz O. 2000.** Is the *Prunella* (Lamiaceae) hybrid zone structured by an environmental gradient? Evidence from a reciprocal transplant experiment. *American Journal of Botany* **87**(7): 995-1003.
- Fritz RS, Hochwender CG, Albrechtsen BR, Czesak ME. 2006.** Fitness and genetic architecture of parent and hybrid willows in common gardens. *Evolution* **60**(6): 1215-1227.
- Gapare WJ, Aitken SN. 2005.** Strong spatial genetic structure in peripheral but not core populations of Sitka spruce [*Picea sitchensis* (Bong.) Carr.]. *Molecular Ecology* **14**(9): 2659-2667.
- Gompert Z, Buerkle CA. 2009.** A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology* **18**(6): 1207-1224.
- Gompert Z, Buerkle CA. 2010.** introgress: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources* **10**(2): 378-384.
- Gompert Z, Buerkle CA. 2011.** Bayesian estimation of genomic clines. *Molecular Ecology* **20**(10): 2111-2127.
- Gompert Z, Lucas LK, Fordyce JA, Forister ML, Nice CC. 2010.** Secondary contact between *Lycaeides idas* and *L. melissa* in the Rocky Mountains: extensive admixture and a patchy hybrid zone. *Molecular Ecology* **19**(15): 3171-3192.
- Gompert Z, Parchman TL, Buerkle CA. 2012.** Genomics of isolation in hybrids. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**: 439-450.
- González-Martínez SC, Ersoz E, Brown GR, Wheeler NC, Neale D. 2006.** 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.
- Grossnickle SC, Sutton BCS, Folk RS, Gawley RJ. 1996.** Relationship between nuclear DNA markers and physiological parameters in Sitka x interior spruce populations. *Tree Physiology* **16**: 547-555.
- Grover CE, Salmon A, Wendel JF. 2012.** Targeted sequene capture as a powerful tool for evolutionary analysis. *American Journal of Botany* **99**(2): 312-319.

- Hamanishi ET, Campbell MM. 2011.** Genome-wide responses to drought in forest trees. *Forestry* **84**(3): 273-283.
- Hamanishi ET, Raj S, Wilkins O, Thomas BR, Mansfield SD, Plant AL, Campbell MM. 2010.** Intraspecific variation in the *Populus balsamifera* drought transcriptome. *Plant, Cell and Environment* **33**: 1742-1755.
- Hamann A, Wang T. 2006.** Potential effects of climate change on ecosystem and tree species distribution in British Columbia. *Ecology* **87**(11): 2773-2786.
- Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F, Bergelson J. 2011.** Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* **334**(6052): 83-86.
- Hannerz M, Aitken SN, King JN, Budge S. 1999.** Effects of genetic selection for growth on frost hardiness in western hemlock. *Canadian Journal for Forest Research* **29**: 509-516.
- Hanninen H, Tanino K. 2011.** Tree seasonality in a warming climate. *Trends in Plant Science* **16**(8): 412-416.
- Harrison RG. 1990.** Hybrid zones: windows on evolutionary process. *Oxford Surveys of Evolutionary Biology* **7**: 69-128.
- Harrison RG. 1993.** *Hybrids and hybrid zones: historical perspective*. New York: Oxford University Press.
- Hellmann J, Pinedakrch M. 2007.** Constraints and reinforcement on adaptation under climate change: Selection of genetically correlated traits. *Biological Conservation* **137**(4): 599-609.
- Hewitt GM. 1988.** Hybrid zones - natural laboratories for evolutionary studies. *Trends in Ecology and Evolution* **3**(7): 158-167.
- Hodgetts RB, Aleksiuik MA, Brown A, Clarke C, Macdonald E, Nadeem S, Khasa D. 2001.** Development of microsatellite markers for white spruce (*Picea glauca*) and related species. *Theoretical and Applied Genetics* **102**: 1252-1258.
- Holliday JA, Ralph SG, White R, Bohlmann J, Aitken SN. 2008.** Global monitoring of autumn gene expression within and among phenotypically divergent populations of Sitka spruce (*Picea sitchensis*). *New Phytologist* **178**(1): 103-122.
- Holliday JA, Ritland K, Aitken SN. 2010a.** Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (*Picea sitchensis*). *New Phytologist* **188**(2): 501-514.
- Holliday JA, Suren H, Aitken SN. 2011.** Divergent selection and heterogeneous migration rates across the range of Sitka spruce (*Picea sitchensis*). *Proceedings of the Royal Society B: Biological Sciences* **279**(1734): 1675-1683.
- Holliday JA, Yuen M, Ritland K, Aitken SN. 2010b.** Postglacial history of a widespread conifer produces inverse clines in selective neutrality tests. *Molecular Ecology* **19**(18): 3857-3864.
- Holsinger KE, Weir BS. 2009.** Genetics in geographically structured populations: defining, estimating and interpreting F_{st}. *Nature Reviews Genetics* **10**: 639-650.
- Howe GT, Aitken SN, Neale DB, Jermstad KD, Wheeler NC, Chen THH. 2003.** From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees. *Canadian Journal of Botany* **81**(12): 1247-1266.
- Jaramillo-Correa JP, Beaulieu J, Bousquet J. 2001.** Contrasting evolutionary forces driving population structure at expressed sequence tag polymorphisms, allozymes and quantitative traits in white spruce. *Molecular Ecology* **10**: 2729-2740.

- Jaramillo-Correa JP, Bousquet J. 2003.** New evidence from mitochondrial DNA of a progenitor-derivative species relationship between black spruce and red spruce (Pinaceae). *American Journal of Botany* **90**(12): 1801-1806.
- Jaramillo-Correa JP, Bousquet J. 2005.** Mitochondrial genome recombination in the zone of contact between two hybridizing Conifers. *Genetics* **171**(4): 1951-1962.
- Jensen JL, Bohonak AJ, Kelley ST. 2005.** Isolation by distance, web service. *BMC Genetics* **6**: 13.
- Jump AS, Penuelas J. 2005.** Running to stand still: adaptation and the response of plants to rapid climate change. *Ecology Letters* **8**(1010-1020).
- Kahm M, Hasenbrink G, Lichtenberg-Frate H, Ludwig J, Kschischo M. 2010.** grofit: Fitting biological growth curves with R. *Journal of Statistical Software* **33**(7): 1-21.
- Kane NC, King MG, Barker MS, Raduski A, Karrenberg S, Yatabe Y, Knapp SJ, Rieseberg LH. 2009.** Genetic analyses indicate highly porous genomes and high levels of gene flow between divergent *Helianthus* species. *Evolution* **63**(8): 2061-2075.
- Kawakami T, Butlin R, Adams M, Paull DJ, Cooper SJB. 2009.** Genetic analysis of a chromosomal hybrid zone in the Australian Morabine grasshoppers (Vanciemenella, Viatica species group). *Evolution* **63**(1): 139-152.
- Keeling CI, Weisshaar S, Ralph S, Jancsik S, Hamberger B, Dullat HK, Bohlmann J. 2011.** Transcriptome mining, functional characterization and phylogeny of a large terpene synthase gene family in spruce (*Picea* spp.). *BMC Plant Biology* **11**: 43-55.
- Keim P, Paige KN, Whitham TG, Lark KG. 1989.** Genetic analysis of an interspecific hybrid swarm of *Populus*: occurrence of unidirectional introgression. *Genetics* **123**: 557-565.
- Ketcheson MV, Braudmandl TF, Meidinger D, Utzig G, Demarchi DA, Wikeem BM. 1991.** Interior Cedar-Hemlock Zone. *Ecosystems of British Columbia*.
- Kimball S, Campbell D. 2009.** Physiological differences among two *Penstemon* species and their hybrids in field and common garden environments. *New Phytologist* **181**: 478-488.
- Kimball S, Campbell DR, Lessin C. 2008.** Differential performance of reciprocal hybrids in multiple environments. *Journal of Ecology* **96**(6): 1306-1318.
- Krakowski J, Chourmouzis C, Yanchuk A, Kolotelo D, Hamann A, Aitken S 2009.** Forest tree genetic conservation status report 2: genetic conservation status of operational tree species. In: Victoria: Centre for Forest Conservation Genetics, Forest Genetics Council of BC, and BC Ministry of Forest and Range, Forest Science Program.
- Kremer A, Ronce O, Robledo-Arnuncio JJ, Guillaume F, Bohrer G, Nathan R, Bridle JR, Gomulkiewicz R, Klein ER, Ritland K, Kuparinen A, Gerber S, Schueler S. 2012.** Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecology Letters* **15**: 378-392.
- Kruuk LEB, Baird SJE, Gale KS, Barton NH. 1999.** A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. *Genetics* **153**: 1959-1971.
- Latta RG, Gardner KM, Johansen-Morris AD. 2006.** Hybridization, recombination, and the genetic basis of fitness variation across environments in *Avena barbata*. *Genetica* **129**(2): 167-177.
- Lewontin RC, Birch LC. 1966.** Hybridization as a source of variation for adaptation to new environments. *Evolution* **20**(3): 315-336.

- Lexer C, Buerkle CA, Joseph JA, Heinze B, Fay MF. 2007.** Admixture in European *Populus* hybrid zones makes feasible the mapping of loci that contribute to reproductive isolation and trait differences. *Heredity* **98**(2): 74-84.
- Lexer C, Fay MF. 2005.** Adaptation to environmental stress: a rare or frequent driver of speciation? *Journal of Evolutionary Biology* **18**(4): 893-900.
- Lexer C, Fay MF, Joseph JA, Nica MS, Heinze B. 2005.** Barrier to gene flow between two ecologically divergent *Populus* species, *P. alba* (white poplar) and *P. tremula* (European aspen): the role of ecology and life history in gene introgression. *Molecular Ecology* **14**(4): 1045-1057.
- Lexer C, Heinze B, Alia R, Rieseberg LH. 2004.** Hybrid zones as a tool for identifying adaptive genetic variation in outbreeding forest trees: lessons from wild annual sunflowers (*Helianthus* spp.). *Forest Ecology and Management* **197**(1-3): 49-64.
- Lexer C, Joseph JA, van Loo M, Barbara T, Heinze B, Bartha D, Castiglione S, Fay MF, Buerkle CA. 2010.** Genomic admixture analysis in European *Populus* spp. reveals unexpected patterns of reproductive isolation and mating. *Genetics* **186**(2): 699-712.
- Lexer C, Kremer A, Petit RJ. 2006.** Shared alleles in sympatric oaks: recurrent gene flow is a more parsimonious explanation than ancestral polymorphism. *Molecular Ecology* **15**(7): 2007-2012.
- Little EL. 1953.** A natural hybrid in Alaska. *Journal of Forestry* **51**: 745-747.
- Macholán M, Munclinger P, Šugerková M, Dufková P, Bímová B, Božíková E, Zima J, Piálek J. 2007.** Genetic analysis of autosomal and X-Linked markers across a mouse hybrid zone. *Evolution* **61**(4): 746-771.
- Martinsen GD, Whitham TG, Turek RJ, Keim P. 2001.** Hybrid populations selectively filter gene introgression between species. *Evolution* **55**(7): 1325-1335.
- Marty JP. 2001.** Ecological consequences of recent climate change. *Conservation Biology* **15**: 320-331.
- Mayr E. 1963.** *Animal Species and Evolution*. Cambridge: Belknap Press.
- Mbogga MS, Wang X, Hamann A. 2010.** Bioclimate envelope model predictions for natural resource management: dealing with uncertainty. *Journal of Applied Ecology* **47**: 731-740.
- McKinnon GE, Jordan GJ, Vaillancourt R, Steane DA, Potts B. 2004.** Glacial refugia and reticulate evolution: the case of the Tasmanian eucalypts. *Philosophical Transactions of the Royal Society B: Biological Sciences* **359**: 275-284.
- McKinnon GE, Smith JJ, Potts B. 2010.** Recurrent nuclear DNA introgression accompanies chloroplast DNA exchange between two eucalypt species. *Molecular Ecology* **19**: 1367-1380.
- McKinnon GE, Vaillancourt RE, Jackson HD, Potts BM. 2001.** Chloroplast sharing in the Tasmanian Eucalypts. *Evolution* **55**(4): 703-711.
- Meirmans PG, Tienderen V. 2004.** GENOTYPE and GENODIVE: two programs for the analysis of genetic diversify of asexual organisms. *Molecular Ecology Notes* **4**: 792-794.
- Miglia KJ, McArthur ED, Moore WS, Wang H, Graham JH, Freeman DC. 2005.** Nine-year reciprocal transplant experiment in the gardens of the basin and mountain big sagebrush (*Artemisia tridentata*: Asteraceae) hybrid zone of Salt Creek Canyon: the importance of multiple-year tracking of fitness. *Biological Journal of the Linnean Society* **86**: 213-225.

- Milne RI, Abbott RJ. 2008.** Reproductive isolation among two interfertile *Rhododendron* species: low frequency of post-F1 hybrid genotypes in alpine hybrid zones. *Molecular Ecology* **17**: 1108-1121.
- Mimura M, Aitken SN. 2007.** Adaptive gradients and isolation-by-distance with postglacial migration in *Picea sitchensis*. *Heredity* **99**(2): 224-232.
- Mitchell-Olds T, Schmitt J. 2006.** Genetic mechanisms and evolutionary significance of natural variation in *Arabidopsis*. *Nature* **441**(7096): 947-952.
- Moehring AJ. 2011.** Heterozygosity and its unexpected correlations with hybrid sterility. *Evolution* **65**(9): 2621-2630.
- Moore WS. 1977.** An evaluation of narrow hybrid zones in vertebrates. *The Quarterly Review of Biology* **52**(3): 263-277.
- Morgenstern EK. 1996.** Geographic variation in forest trees. *UBC Press*.
- Namroud M-C, Beaulieu J, Juge N, Laroche J, Bousquet J. 2008.** Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Molecular Ecology* **17**(16): 3599-3613.
- Neale D, Ingvarsson P. 2008.** Population, quantitative and comparative genomics of adaptation in forest trees. *Current Opinion in Plant Biology*.
- Neale DB, Kremer A. 2011.** Forest tree genomics: growing resources and applications. *Nature Reviews Genetics* **12**(2): 111-122.
- Neale DB, Sederoff RR. 1989.** Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in loblolly pine. *Theoretical and Applied Genetics* **77**: 212-216.
- Nielsen EE, Hemmer-Hansen J, Poulsen NA, Loeschcke V, Moen T, Johansen T, Mittelholzer C, Taranger G, Ogden R, Carvalho GR. 2009.** Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology* **9**: 276-287.
- Niklas KJ. 1984.** The motion of windborne pollen grains around conifer ovulate cones: implications on wind pollination. *American Journal of Botany* **71**(3): 356-374.
- Niklas KJ, Tha Paw U K. 1982.** Pollination and airflow patterns around conifer ovulate cones. *Science* **217**(4458): 442-444.
- Nolte AW, Gompert Z, Buerkle CA. 2009.** Variable patterns of introgression in two sculpin hybrid zones suggest that genomic isolation differs among populations. *Molecular Ecology* **18**(12): 2615-2627.
- Nosil P, Schluter D. 2011.** The genes underlying the process of speciation. *Trends in Ecology and Evolution* **26**(4): 160-167.
- O'Neill GA, Aitken SN, King JN, Alfaro RI. 2002.** Geographic variation in resin canal defenses in seedlings from the Sitka spruce x white spruce introgression zone. *Canadian Journal of Forest Research* **32**(3): 390-400.
- O'Neill GA, Hamann A, Wang T. 2008a.** Accounting for population variation improves estimates of the impact of climate change on species' growth and distribution. *Journal of Applied Ecology* **45**: 1040-1049.
- O'Neill GA, Ukrainetz NK, Carlson MR, Cartwright CV, Jaquish BC, King JN, Krakowski J, Russell JH, Stoehr MU, Xie C, Yanchuk A 2008b.** Assisted migration to address climate change in British Columbia: recommendations for interim seed transfer standards. In. Victoria: BC Ministry of Forest and Range, Research Branch.
- Paine CET, Marthews TR, Vogt DR, Purves D, Rees M, Hector A, Turnbull LA. 2011.** How to fit nonlinear plant growth models and calculate growth rates: an update for ecologists. *Methods in Ecology and Evolution*.

- Parmesan C. 2006.** Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology, Evolution, and Systematics* **37**(1): 637-669.
- Pavy N, Pelgas B, Beauseigle S, Blais S, Gagnon F, Gosselin I, Lamothe M, Isabel N, Bousquet J. 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**(1): 21.
- Payseur BA. 2010.** Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular Ecology Resources* **10**(5): 806-820.
- Pelgas B, Bousquet J, Meirmans PG, Ritland K, Isabel N. 2011.** QTL mapping in white spruce: gene maps and genomic regions underlying adaptive traits across pedigrees, years and environments. *BMC Genomics* **12**(1): 145.
- Pertoldi C, Bijlsma R, Loeschcke V. 2007.** Conservation genetics in a globally changing environment: present problems, paradoxes and future challenges. *Biodiversity and Conservation* **16**(14): 4147-4163.
- Petit RJ, Csaikl UM, Bordacs S, Burg K, Coart E, Cottrell J, Dam Bv, Deans JD, Dumolin-Lapegue S, Fineschi S, Finkeldey R, Gillies A, Glaz I, Goicoechea PG, Jensen JS, König AO, Lowe AJ, Madsen SF, Matyas G, Munro RC, Olaldo M, Pemonge M-H, Popescu F, Slade D, Tabbener H, Turchini D, de Vries SGM, Ziegenhagen B, Kremer A. 2002.** Chloroplast DNA variation in European white oaks phylogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecology and Management* **156**: 5-26.
- Petit RJ, Excoffier L. 2009.** Gene flow and species delimitation. *Trends in Ecology and Evolution* **24**(7): 386-393.
- Petit RJ, Pineau E, Demesure B, Bacilieri R, Ducousso A, Kremer A. 1997.** Chloroplast DNA footprints of postglacial recolonization by Oaks. *Proceedings of the Academy of Science USA* **94**(18): 9996-10001.
- Pielou EC. 1991.** After the Ice Age: the Return of Life to Glaciated North America. *University of Chicago Press, Chicago*.
- Pisaric MFJ, Holt C, Szeicz JM, Karst T, Smol JP. 2003.** Holocene treeline dynamics in the mountains northeastern British Columbia, Canada, inferred from fossil pollen and stomata. *The Holocene* **13**(2): 161-173.
- Pojar J, Klinka K, Demarchi DA. 1991.** Coastal Western Hemlock Zone.
- Porter AH, Wenger R, Geiger H, Scholl A, Shapiro AM. 1997.** The *Pontia daplidice-edusa* hybrid zone in Northwestern Italy. *Evolution* **51**(5): 1561-1573.
- Pritchard JK, Stephens M, Donnelly P. 2000.** Inference of population structure using multilocus genotype data. *Genetics* **155**: 945-959.
- Pyhäjärvi T, Salmela MJ, Savolainen O. 2007.** Colonization routes of *Pinus sylvestris* inferred from distribution of mitochondrial DNA variation. *Tree Genetics and Genomes* **4**(2): 247-254.
- Rajora OP, Rahman MH, Dayanandan S, Mosseler A. 2001.** Isolation, characterization, inheritance and linkage of microsatellite DNA markers in white spruce (*Picea glauca*) and their usefulness in other spruce species. *Molecular and General Genetics* **264**: 871-882.
- Ralph SG, Yueh H, Friedmann M, Aeschliman D, Zeznik JA, Nelson CC, Butterfield YSN, Kirkpatrick R, Liu J, Jones SJM, Marra MA, Douglas CJ, Ritland K, Bohlmann J. 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**(8): 1545-1570.
- Ran J-H, Wei X-X, Wang X-Q. 2006.** Molecular phylogeny and biogeography of *Picea* (Pinaceae): Implications for phylogeographical studies using cytoplasmic haplotypes. *Molecular Phylogenetics and Evolution* **41**(2): 405-419.
- Raymond M, Rousset F. 1995.** GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**: 248-249.
- Renault S, Maillet N, Normandeau E, Sauvage C, Derome N, Rogers SM, Bernatchez L. 2012.** Genome-wide patterns of divergence during speciation: the lake whitefish case study. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**: 354-363.
- Rieseberg L. 1995.** The role of hybridization in evolution: old wine in new skins. *American Journal of Botany* **82**(7): 944-953.
- Rieseberg LH, Archer MA, Wayne RK. 1999a.** Transgressive segregation, adaptation and speciation. *Heredity* **83**: 363-372.
- Rieseberg LH, Blackman BK. 2010.** Speciation genes in plants. *Annals of Botany* **106**: 439-455.
- Rieseberg LH, Carney SE. 1998.** Plant hybridization. *New Phytologist* **140**: 599-624.
- Rieseberg LH, Ellstrand NC. 1993.** What can molecular and morphological markers tell us about plant hybridization. *Critical Reviews in Plant Sciences* **12**(3): 213-241.
- Rieseberg LH, Soltis DE. 1991.** Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* **5**(1): 65-84.
- Rieseberg LH, Welch ME 2002.** Gene transfer through introgressive hybridization: history evolutionary significance, and phylogenetic consequences. In: Syvanen M, Kado CI eds. *Horizontal Gene Transfer*. London: Academic Press, 199-216.
- Rieseberg LH, Wendel JF. 1993.** Introgression and its consequences in plants. *Hybrid zones and the evolutionary process*, Oxford University Press, New York.
- Rieseberg LH, Whitton J, Gardner KA. 1999b.** Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* **152**(713-727).
- Rieseberg LH, Widmer A, Arntz AM, Burke B. 2003.** The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. *Philosophical Transactions of the Royal Society B: Biological Sciences* **358**(1434): 1141-1147.
- Ritland K. 2012.** Genomics of a phylum distant from flowering plants: conifers. *Tree Genetics and Genomes*.
- Roberts DR, Hamann A. 2012.** Predicting potential climate change impacts with bioclimate envelope models: a palaeoecological perspective. *Global Ecology and Biogeography* **21**: 121-133.
- Roche L. 1969.** A genecological study of the genus *Picea* in British Columbia. *New Phytologist* **68**(2): 505-554.
- Rungis D, Berube Y, Zhang J, Ralph S, Ritland CE, Ellis BE, Douglas C, Bohlmann J, Ritland K. 2004.** Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. *Theoretical and Applied Genetics* **109**(6): 1283-1294.
- Rweyongeza DM, Dhir NK, Barnhardt LK, Hansen C, Yang R-C. 2007.** Population differentiation of the lodgepole pine (*Pinus contorta*) and jack pine (*Pinus banksiana*) complex in Alberta: growth, survival, and responses to climate. *Canadian Journal of Botany* **85**: 545-556.

- Sambatti JBM, Ortiz-Barrientos D, Baack EJ, Rieseberg LH. 2008.** Ecological selection maintains cytonuclear incompatibilities in hybridizing sunflowers. *Ecology Letters* **11**(10): 1082-1091.
- Sapir Y, Moody ML, Brouillette LC, Donovan LA, Rieseberg LH. 2007.** Patterns of genetic diversity and candidate genes for ecological divergence in a homoploid hybrid sunflower, *Helianthus anomalus*. *Molecular Ecology* **16**(23): 5017-5029.
- Savolainen O. 2011.** The genomic basis of local climatic adaptation. *Science* **334**: 49-50.
- Savolainen O, Pyhäjärvi T, Knürr T. 2007.** Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution, and Systematics* **38**(1): 595-619.
- Scotti-Saintagne C, Mariette S, Porth I, Goicoechea PG, Barreneche T, Bodenes C, Burg K, Kremer A. 2004.** Genome scanning for Interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.]. *Genetics* **168**(3): 1615-1626.
- Seehausen O. 2004.** Hybridization and adaptive radiation. *Trends in Ecology and Evolution* **19**(4): 198-207.
- Shafer ABA, Cullingham CI, CÔTÉ SD, Coltman DW. 2010.** Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology* **19**(21): 4589-4621.
- Shen R, Fan J, Campbell D, Chang W, Chen J, Doucet D, Yeakley J, Bibikova M, Wickhamgarcia E, McBride C. 2005.** High-throughput SNP genotyping on universal bead arrays. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **573**(1-2): 70-82.
- Silim SN, Guy RD, Patterson TB, Livingston NJ. 2001.** Plasticity in water-use efficiency of *Picea sitchensis*, *P. glauca* and their natural hybrids. *Oecologia* **128**(3): 317-325.
- Siol M, Wright SI, Barrett SCH. 2010.** The population genomics of plant adaptation. *New Phytologist* **188**: 313-332.
- Slatkin M. 1995.** A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**: 1463.
- Snetsinger J 2010.** Chief Forester's standards for seed use. In. Victoria: Ministry of Forests.
- Soltis D, Gitzendanner M, Streng D, Soltis P. 1997.** Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. *Plant Systematics and Evolution* **206**: 353-373.
- Spitze K. 1993.** Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation. *Genetics* **135**: 367-374.
- Stebbins GL. 1959.** The role of hybridization in evolution. *Proceedings of the American Philosophical Society* **103**(2): 231-251.
- Steponkus PL. 1984.** Role of the plasma membrane in freezing injury and cold acclimation. *Annual Review of Plant Physiology* **35**(543-584).
- Strasburg JL, Sherman NA, Wright KM, Moyle LC, Willis JH, Rieseberg LH. 2012.** What can patterns of differentiation across plant genomes tell us about adaptation and speciation ? *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**: 364-373.
- Street NR, Skogstrom O, Sjodin A, Tucker J, Rodriguez-Acosta M, Nilsson P, Jansson S, Taylor G. 2006.** The genetics and genomics of the drought response in *Populus*. *The Plant Journal* **48**: 321-341.
- Sun Z, J., Livingston NJ, Guy RD, Ethier GJ. 1996.** Stable carbon isotopes as indicators of increased water use efficiency and productivity in white spruce (*Picea glauca* (Moench) Voss) seedlings. *Plant, Cell and Environment* **19**: 887-894.

- Sutton BCS, Flanagan DJ, Gawley JR, Newton CH, Lester DT, El-Kassaby YA. 1991.** Inheritance of chloroplast and mitochondrial DNA in *Picea* and composition of hybrids from introgression zones. *Theoretical and Applied Genetics* **82**: 242-248.
- Sutton BCS, Pritchard SC, Gawley JR, Newton CH. 1994.** Analysis of Sitka spruce - interior spruce introgression in British Columbia using cytoplasmic and nuclear DNA probes. *Canadian Journal for Forest Research* **24**(278-285).
- Szymura JM, Barton NH. 1986.** Genetic Analysis of a Hybrid Zone Between the Fire-Bellied Toads, *Bombina bombina* and *B. variegata*, Near Cracow in Southern Poland. *Evolution* **40**(6): 1141-1159.
- Taberlet P, Gielly L, Pautou G, Bouvet J. 1991.** Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17**: 1105-1109.
- Teeter KC, Thibodeau LM, Gompert Z, Buerkle CA, Nachman MW, Tucker PK. 2010.** The variable genomic architecture of Isolation between hybridizing species of house mice. *Evolution* **64**(2): 472-485.
- Tiffin P, Olson MS, Moyle LC. 2001.** Asymmetrical crossing barriers in agiosperms. *Proceedings of the Royal Society B: Biological Sciences* **268**(861-867).
- Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. 2004.** Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**: 535-538.
- Vendramin GG, Lelli L, Rossi P, Morgante M. 1996.** A set of primers for the amplification of 20 chloroplast microsatellites in Pinaceae. *Molecular Ecology* **5**: 595-598.
- Verdu M. 2002.** Age at maturity and diversification in woody agiosperms. *Evolution* **56**(7): 1352-1361.
- Via S. 2012.** Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene flow. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**: 451-460.
- Walls SC. 2009.** The role of climate in the dynamics of a hybrid zone in Appalachian salamanders. *Global Change Biology* **15**: 1903-1910.
- Wang H, McArthur ED, Sanderson SC, Graham JH, Freeman DC. 1997.** Narrow hybrid zone between two subspecies of Big Sagebrush (*Artemisia tridentata*: Asteraceae). IV. Reciprocal transplant experiments. *Evolution* **51**(1): 95-102.
- Wang T, Campbell EM, O'Neill GA, Aitken SN. 2012a.** Projecting future distributions of ecosystem climate niches: uncertainties and management applications. *Forest Ecology and Management* **279**: 128-140.
- Wang T, Hamann A, Spittlehouse D, Aitken SN. 2006a.** Development of scale-free climate data for western Canada for use in resource management. *International Journal of Climatology* **26**(3): 383-397.
- Wang T, Hamann A, Spittlehouse D, Murdock TQ. 2012b.** ClimateWNA - High-resolution spatial climate data for western North America. *Journal of Applied Meteorology and Climatology* **51**: 16-29.
- Wang T, Hamann A, Yanchuk A, O'Neill GA, Aitken SN. 2006b.** Use of response functions in selecting lodgepole pine populations for future climates. *Global Change Biology* **12**(12): 2404-2416.
- Weir BS. 1996.** *Genetic data analysis II*. Sunderland, Massachusetts, USA: Sinauer.
- Welch ME, Rieseberg LH. 2002.** Habitat divergence between a homoploid hybrid sunflower species, *Helianthus paradoxus* and its progenitors. *American Journal of Botany* **89**: 472-478.

- Wheeler NC, Guries RP. 1987.** A quantitative measure of introgression between lodgepole and jack pines. *Canadian Journal of Botany* **65**(1876-1885).
- White TL, Adams WT, Neale DB. 2007.** *Forest Genetics*. Cambridge: CABI Publishing.
- Whitlock MC. 1992.** Temporal fluctuations in demographic parameters and the genetic variance among populations. *Evolution* **46**(3): 608-615
- .
- Whitlock MC. 2008.** Evolutionary inference from QST. *Molecular Ecology* **17**(8): 1885-1896.
- Whittemore AT, Schaal BA. 1991.** Interspecific gene flow in sympatric oaks. *Proceedings of the National Academy of Sciences* **88**(6): 2540-2544.
- Wilhelm R, Hilbisch TJ. 1998.** Assessment of natural selection in a hybrid population of mussels: evaluation of exogenous vs endogenous selection models. *Marine Biology* **131**: 505-514.
- Woods J 1988.** Nursery trials of Sitka-Interior spruce hybrids. In Lands MoFa. *British Columbia*.
- Wright S. 1951.** The genetical structure of populations. *Annals of Eugenics* **15**: 323-354.
- Wu C. 2001.** The genic view of the process of speciation. *Journal of Evolutionary Biology* **14**: 851-865.
- Wu CA, Campbell DR. 2005.** Cytoplasmic and nuclear markers reveal contrasting patterns of spatial genetic structure in a natural *Ipomopsis* hybrid zone. *Molecular Ecology* **14**(3): 781-792.
- Yanchukov A, Hofman S, Szymura JM, Mezhzherin SV, Morozov-Leonov SY, Barton N, Nurnberger B. 2006.** Hybridization of *Bombina bombina* and *B. variegata* (Anura, Discoglossidae) at a sharp ecotone in western Ukraine: comparisons across transects and over time. *Evolution* **60**(3): 583-600.
- Yatabe Y, Kane NC, Scotti-Saintagne C, Rieseberg LH. 2007.** Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics* **175**(4): 1883-1893.
- Yeh FC, Arnott JT. 1986.** Electrophoretic and morphological differentiation of *Picea sitchensis*, *Picea glauca*, and their hybrids. *Canadian Journal of Forest Research* **16**: 791-798.
- Zeng Y, Liao W, Petit RJ, Zhang D. 2010.** Exploring species limits in two closely related Chinese Oaks. *PLoS ONE* **5**(11): 1-15.

Appendix: supplementary tables and figures

Table S1 Principal components analysis (PC) describing morphological variation with eigenvector coefficients, eigenvalues, proportion of variation explained.

	PC1 (Length) Eigenvector Coefficient	PC2 (Width) Eigenvector Coefficient
L1	0.68	0.31
L2	0.36	0.44
L3	0.60	-0.50
W1	0.03	0.51
W2	-0.17	0.44
W3	0.12	-0.11
Eigenvalue	0.05	0.03
% of total variance	50.80	26.90

Table S2 - Summary of locus-specific patterns across the Sitka-white spruce zone of introgression, including genetic data analysis, geographic and genomic cline analysis for 268 candidate gene SNPs. Candidate gene SNPs used in the study include locus number, name, heterozygosity (H_e), genetic differentiation (F_{ST}), and inbreeding coefficient (F_{IS}). Geographic cline parameters included identify regression model, R^2 , AIC, maximum slope, and lambda (λ) of the fit data. Genomic cline selection categories indicate over (+) or under (-) representation of genotypic classes based on a model of neutral introgression.

#	Locus Name	H_e	F_{ST}	F_{IS}	Reg ^a	R^2	Geographic Clines			Genomic Clines		
							AIC	Max slope	λ	Selection Categories ^b		
1	0_10112.contig2.C2.352	0.33	0.34	-0.05	Log	0.79	-57.73	0.004	111	WW	WS-	SS
2	0_10754.contig2.C1.179	0.19	0.03	0.00	Lin	0.10	-75.49	0	-	-	-	-
3	0_12681.contig2.C2.315	0.07	0.04	-0.04	Lin	0.11	-90.33	0	-	WW	WS+	SS
4	0_13680.contig2.C1.149	0.26	0.42	0.05	Log	0.83	-82.63	0.004	152	WW-	WS-	SS+
5	0_13680.contig2.NC1.68	0.33	0.37	0.00	Exp	0.86	-71.74	0.01	203	WW	WS	SS
6	0_14976.contig2.NC1.354	0.05	0.05	0.03	Lin	0.39	-118.44	0	-	WW	WS	SS
7	0_15075.contig2.C2.341	0.11	0.04	0.08	Lin	0.08	-80.53	0	-	WW+	WS-	SS
8	0_16142.contig2.C1.266	0.23	0.47	0.04	Exp	0.80	-84.67	0.009	220	WW-	WS-	SS+
9	0_17017.contig2.C1.225	0.34	0.43	0.04	Log	0.63	-30.07	0.004	106	WW	WS	SS
10	0_17017.contig2.NC1.250	0.34	0.43	0.03	Log	0.65	-31.17	0.004	110	WW	WS	SS
11	0_17238.contig2.NC1.122	0.31	0.18	-0.03	Exp	0.72	-73.64	0.004	161	WW	WS	SS
12	100_316_NS	0.42	0.19	0.01	Exp	0.20	-45.29	0.001	< 0	WW-	WS+	SS+
13	103_455_NS	0.44	0.07	-0.08	Lin	0.06	-38.24	0	-	WW+	WS	SS-
14	114_248_S	0.41	0.11	-0.03	Lin	0.34	-47.85	0.001	-	WW	WS	SS
15	124_495_S	0.47	0.12	-0.02	Exp	0.40	-34.66	0.002	8	WW	WS	SS
16	125_312_S	0.29	0.45	0.13	Exp	0.75	-56.79	0.011	214	WW	WS-	SS
17	127_273_S	0.27	0.19	-0.03	Exp	0.54	-71.05	0.002	140	WW-	WS-	SS+
18	13_496_NS	0.45	0.12	-0.06	Exp	0.33	-29.25	0.002	18	WW	WS	SS
19	13_632_S	0.50	0.16	-0.02	Exp	0.39	-26.08	0.002	< 0	WW+	WS-	SS
20	132_78_S	0.42	0.15	-0.02	Exp	0.42	-33.02	0.003	81	WW+	WS+	SS-
21	133_39_S	0.21	0.18	-0.08	Exp	0.34	-57.67	0.002	155	WW	WS	SS
22	133_418_S	0.41	0.14	-0.12	Log	0.63	-43.09	0.003	< 0	WW+	WS-	SS-
23	133_553_NS	0.42	0.11	-0.01	Exp	0.52	-46.91	0.003	77	WW+	WS+	SS-
24	135_122_NS	0.44	0.02	0.03	Lin	0.00	-50.70	0	-	WW	WS	SS
25	14_248_NS	0.50	0.12	-0.02	Exp	0.34	-35.56	0.002	< 0	WW-	WS	SS+
26	14_301_NS	0.50	0.11	-0.03	Exp	0.37	-37.74	0.002	< 0	WW-	WS	SS+

Table S2 - Cont'd

#	Locus Name	H_e	F_{ST}	F_{IS}	Reg ^a	R^2	Geographic Clines		λ	Genomic Clines		
							AIC	Max slope		Selection Categories ^b		
27	141_349_S	0.27	0.16	-0.29	Log	0.77	-70.72	0.003	81	WW-	WS+	SS+
28	144_441_S	0.46	0.44	-0.02	Exp	0.61	-13.82	0.009	163	WW+	WS-	SS-
29	162_199_S	0.49	0.04	0.03	Lin	0.12	-40.39	0	-	WW+	WS-	SS-
30	164_465_S	0.49	0.13	0.04	Lin	0.06	-31.02	0	-	WW-	WS	SS+
31	169_375_NS	0.23	0.13	0.06	Log	0.53	-59.94	0.002	67	WW	WS	SS
32	179_114_S	0.41	0.09	0.00	Lin	0.43	-44.48	0.001	-	WW+	WS	SS
33	179_319_NS	0.47	0.13	-0.08	Exp	0.20	-23.81	0.001	< 0	WW	WS+	SS+
34	179_699_S	0.49	0.20	-0.08	Exp	0.48	-28.30	0.003	43	WW	WS	SS
35	19_567_S	0.46	0.05	-0.08	Lin	0.10	-54.08	0	-	WW-	WS+	SS
36	191_162_S	0.26	0.06	-0.04	Exp	0.48	-66.57	0.002	124	WW	WS-	SS
37	194_470_S	0.49	0.15	-0.01	Exp	0.67	-53.05	0.003	50	WW	WS	SS
38	195_356_NS	0.25	0.47	0.02	Log	0.67	-60.94	0.005	177	WW-	WS-	SS+
39	198_447_S	0.04	0.03	0.10	Lin	0.02	-113.04	0	-	-	-	-
40	20_374_NS	0.42	0.36	0.02	Log	0.74	-34.76	0.007	146	WW+	WS+	SS-
41	206_435_NS	0.38	0.10	0.00	Exp	0.34	-44.34	0.002	44	WW	WS	SS
42	208PG02825j	0.30	0.07	-0.01	Exp	0.16	-44.91	0.001	< 0	WW	WS	SS
43	208PG04280j	0.24	0.16	0.25	Exp	0.18	-49.54	0.001	84	WW	WS-	SS
44	208PG08590a	0.42	0.12	-0.01	Exp	0.27	-37.00	0.002	< 0	WW	WS	SS
45	208pg10495g	0.26	0.36	0.05	Exp	0.78	-65.62	0.01	213	WW	WS	SS
46	208pg10524e	0.27	0.24	0.01	Log	0.53	-56.26	0.002	60	WW	WS	SS
47	208pg10802g.1	0.31	0.29	0.03	Log	0.71	-51.55	0.005	158	WW	WS	SS
48	208pg12875c	0.23	0.23	-0.04	Log	0.88	-76.21	0.005	152	WW-	WS+	SS+
49	208pg13043k	0.21	0.19	-0.04	Exp	0.23	-55.06	0.001	124	WW-	WS-	SS+
50	208PG13612a	0.15	0.07	-0.07	Lin	0.01	-55.48	0	-	WW	WS-	SS
51	209_523_S	0.38	0.14	-0.02	Exp	0.44	-43.43	0.003	94	WW	WS	SS
52	213_153_S	0.24	0.33	-0.01	Log	0.73	-61.78	0.005	152	WW	WS	SS
53	213_330_S	0.31	0.40	0.07	Log	0.77	-58.81	0.004	124	WW	WS	SS
54	213_468_NS	0.23	0.33	-0.01	Log	0.73	-61.38	0.004	134	WW	WS	SS
55	213_72_S	0.30	0.42	0.08	Log	0.77	-59.58	0.004	127	WW	WS	SS
56	214_180_S	0.39	0.29	-0.06	Exp	0.71	-40.80	0.008	177	WW+	WS	SS
57	214_558_S	0.37	0.31	-0.03	Exp	0.72	-43.22	0.008	185	WW+	WS	SS

Table S2 - Cont'd

#	Locus Name	H_e	F_{ST}	F_{IS}	Reg ^a	R^2	AIC	Geographic Clines		Genomic Clines		
								Max slope	λ	Selection Categories ^b		
58	215_132_S	0.37	0.18	-0.05	Exp	0.49	-46.58	0.003	117	WW	WS	SS
59	222_305_S	0.50	0.10	-0.12	Lin	0.07	-35.39	0	-	WW-	WS+	SS+
60	222_370_S	0.13	0.07	0.00	Lin	0.00	-62.60	0	-	WW	WS	SS
61	242_241_S	0.25	0.20	0.00	Log	0.50	-75.72	0.001	53	WW-	WS-	SS+
62	244_118_NS	0.29	0.04	-0.04	Lin	0.03	-50.82	0	-	WW	WS	SS
63	245_170_NS	0.36	0.25	-0.05	Exp	0.51	-43.88	0.004	141	WW-	WS-	SS+
64	245_281_S	0.48	0.15	0.04	Exp	0.27	-30.39	0.002	< 0	WW	WS	SS
65	245_98_NS	0.29	0.16	-0.02	Exp	0.42	-43.37	0.003	154	WW	WS-	SS
66	249_648_S	0.50	0.11	0.00	Exp	0.09	-33.06	0.001	< 0	WW-	WS	SS+
67	252_200_NS	0.26	0.07	0.07	Exp	0.37	-66.36	0.002	98	WW	WS	SS
68	259_736_NS	0.44	0.25	-0.03	Log	0.65	-38.59	0.003	74	WW	WS+	SS-
69	260_264_S	0.44	0.25	-0.02	Log	0.66	-39.59	0.003	82	WW	WS+	SS-
70	260_84_S	0.44	0.25	-0.03	Log	0.65	-38.59	0.003	71	WW	WS+	SS-
71	27_420_S	0.50	0.10	0.00	Lin	0.29	-44.27	0	-	WW	WS	SS-
72	27_711_S	0.31	0.11	0.08	Exp	0.45	-46.14	0.003	146	WW+	WS-	SS
73	27_99_S	0.15	0.04	-0.02	Lin	0.02	-72.16	0	-	WW	WS	SS
74	273507_S	0.50	0.04	0.04	Lin	0.04	-34.19	0	-	WW	WS	SS-
75	288_302_NS	0.47	0.19	-0.02	Log	0.67	-44.06	0.003	50	WW	WS	SS+
76	29_177_S	0.50	0.13	0.06	Lin	0.18	-28.89	0	-	WW-	WS+	SS+
77	29_592_S	0.46	0.19	0.08	Exp	0.39	-28.60	0.003	46	WW	WS	SS-
78	295_78_S	0.43	0.19	0.00	Exp	0.46	-40.51	0.003	76	WW	WS	SS
79	2iTC2438a	0.41	0.14	0.00	Exp	0.47	-35.82	0.003	96	WW+	WS	SS
80	2pa08pg12519k	0.31	0.20	-0.06	Log	0.67	-51.70	0.002	74	WW	WS+	SS
81	2TC7674e	0.16	0.12	0.01	Exp	0.30	-64.34	0.002	15	WW	WS	SS
82	30_423_S	0.44	0.03	-0.01	Lin	0.01	-43.14	0	-	WW	WS	SS
83	41_150_NS	0.50	0.08	-0.03	Lin	0.26	-39.55	0	-	WW+	WS	SS-
84	45_1067_NS	0.21	0.03	-0.01	Lin	0.11	-69.89	0	-	WW	WS+	SS
85	46_575_NS	0.43	0.08	0.00	Lin	0.34	-47.05	0	-	WW+	WS-	SS-
86	46_623_NS	0.46	0.11	-0.01	Exp	0.37	-39.14	0.001	< 0	WW+	WS	SS-
87	50_135_S	0.20	0.07	-0.01	Exp	0.26	-70.15	0.001	100	WW	WS+	SS
88	51_36_S	0.35	0.31	0.05	Exp	0.67	-46.52	0.007	187	WW	WS-	SS

Table S2 - Cont'd

#	Locus Name	H_e	F_{ST}	F_{IS}	Reg ^a	R^2	Geographic Clines			Genomic Clines		
							AIC	Max slope	λ	Selection Categories ^b		
89	68_286_S	0.44	0.28	-0.01	Exp	0.80	-48.66	0.008	163	WW+	WS+	SS-
90	69_753_S	0.38	0.06	0.19	Lin	0.05	-35.94	0	-	WW-	WS	SS+
91	71_365_NS	0.50	0.12	0.00	Exp	0.32	-28.42	0.002	< 0	WW	WS	SS
92	84_261_S	0.35	0.05	0.01	Lin	0.17	-54.77	0	-	WW	WS	SS
93	84_370_NS	0.46	0.25	-0.02	Exp	0.59	-32.94	0.005	117	WW	WS	SS-
94	85_279_S	0.33	0.35	0.08	Log	0.56	-33.93	0.002	87	WW	WS	SS
95	86_438_S	0.32	0.34	0.13	Log	0.73	-54.70	0.003	107	WW	WS	SS
96	89_300_NS	0.46	0.25	-0.02	Exp	0.59	-33.66	0.005	119	WW	WS	SS-
97	89_37_NS	0.46	0.25	-0.02	Exp	0.59	-33.66	0.005	119	WW	WS	SS-
98	97_489_S	0.45	0.10	0.04	Lin	0.54	-50.21	0.001	-	WW	WS	SS
99	BB.PF00643.12e	0.15	0.08	0.01	Lin	0.22	-57.59	0.001	-	WW+	WS+	SS
100	BB.PF0139.20e	0.50	0.12	0.29	Exp	0.33	-24.71	0.002	< 0	WW+	WS-	SS-
101	C13628.contig2.C4.584	0.34	0.32	0.04	Exp	0.81	-67.63	0.008	192	WW	WS	SS
102	C14881.contig5.C1.273	0.18	0.43	-0.05	Exp	0.43	-63.52	0.004	202	WW-	WS+	SS+
103	C1498.contig1.NC1.839	0.37	0.34	-0.03	Log	0.67	-31.34	0.005	137	WW+	WS+	SS-
104	C1498.contig1.NC2.1166	0.28	0.08	-0.01	Exp	0.10	-42.56	0.001	< 0	WW+	WS-	SS-
105	C16679.contig1.C1.315	0.45	0.07	-0.03	Exp	0.30	-37.75	0.002	< 0	WW+	WS	SS-
106	C18467.contig1.NC2.168	0.34	0.37	-0.03	Exp	0.51	-40.87	0.005	163	WW	WS	SS
107	C20322.contig1.NC3.296	0.37	0.31	0.01	Log	0.75	-40.82	0.007	107	WW+	WS	SS
108	C20925.contig1.NC3.450	0.21	0.08	-0.10	Exp	0.51	-65.03	0.003	174	WW	WS+	SS
109	C2211.contig1.C5.1435	0.09	0.16	0.07	Exp	0.50	-93.41	0.004	215	WW	WS-	SS
110	C2270.contig1.NC1.384	0.31	0.39	0.06	Log	0.66	-44.62	0.003	104	WW	WS	SS
111	C2285.contig1.C2.449	0.12	0.16	-0.03	Exp	0.24	-80.06	0.001	155	WW-	WS-	SS+
112	C2319.contig2.NC1.360	0.24	0.29	0.02	Log	0.82	-67.12	0.006	170	WW+	WS	SS
113	C24607.contig1.NC4.1208	0.32	0.31	0.00	Log	0.86	-68.62	0.004	108	WW	WS	SS
114	C3300.contig1.NC4.640	0.48	0.03	0.01	Lin	0.07	-43.11	0	-	WW	WS	SS-
115	C4447.contig1.C2.631	0.36	0.30	-0.04	Log	0.73	-49.86	0.005	149	WW	WS	SS
116	C4545.contig1.C1.200	0.33	0.03	0.02	Lin	0.03	-55.95	0	-	WW+	WS-	SS-
117	C4575.contig1.C2.853	0.33	0.09	0.01	Lin	0.29	-45.30	0	-	WW+	WS	SS
118	C4773.contig1.NC1.338	0.22	0.31	0.10	Log	0.53	-49.31	0.002	112	WW	WS-	SS
119	C4944.contig2.C2.472	0.11	0.19	-0.01	Exp	0.66	-115.58	0.003	213	WW-	WS-	SS+

Table S2 - Cont'd

#	Locus Name	H_e	F_{ST}	F_{IS}	Reg ^a	R^2	Geographic Clines		λ	Genomic Clines		
							AIC	Max slope		Selection Categories ^b		
120	C4944.contig2.C4.573	0.24	0.19	0.06	Exp	0.60	-73.56	0.003	170	WW	WS-	SS
121	C4944.contig2.C5.740	0.40	0.07	-0.01	Lin	0.10	-31.30	0.001	-	WW+	WS+	SS-
122	C5104.contig1.C1.624 ⁱ	0.38	0.13	0.07	Exp	0.54	-44.53	0.003	119	WW+	WS	SS
123	C6522.contig1.NC1.269 ⁱ	0.49	0.04	-0.02	Lin	0.02	-35.90	0	-	WW	WS	SS
124	C6814.contig1.NC8.578	0.50	0.04	-0.04	Lin	0.00	-40.37	0	-	WW-	WS	SS+
125	C6847.contig1.C2.1238	0.26	0.05	0.02	Lin	0.45	-74.21	0.001	-	WW	WS	SS
126	C717.contig2.NC2.162	0.49	0.04	0.09	Lin	0.02	-34.82	0	-	WW	WS	SS
127	C7807.contig1.C1.230	0.15	0.12	0.01	Exp	0.27	-79.31	0.001	145	WW	WS	SS
128	C8159.contig1.NC7.1499	0.50	0.04	0.11	Lin	0.09	-43.82	0	-	WW	WS	SS
129	C9634.contig2.NC2.1086	0.10	0.07	0.01	Lin	0.43	-100.82	0	-	WW	WS	SS
130	C996.contig1.NC1.663	0.37	0.10	-0.01	Lin	0.11	-69.87	0	-	WW-	WS-	SS+
131	C996.contig1.NC4.945	0.34	0.14	0.10	Lin	0.19	-67.26	0	-	WW-	WS-	SS+
132	CL1458Contig1.contig2.C2.311	0.39	0.29	-0.12	Exp	0.71	-50.78	0.006	161	WW	WS+	SS
133	CL1458Contig1.contig2.C3.377	0.19	0.03	0.00	Lin	0.10	-75.02	0	-	-	-	-
134	CO481261.contig1.NC7.671	0.12	0.04	-0.01	Lin	0.05	-75.73	0	-	WW	WS	SS
135	CO484662.contig1.C1.269	0.19	0.03	-0.01	Lin	0.15	-70.34	0	-	WW	WS	SS
136	P03539.4	0.24	0.11	-0.06	Log	0.62	-56.76	0.002	77	WW+	WS+	SS
137	p09832.2	0.18	0.19	0.05	Exp	0.79	-72.42	0.011	221	WW+	WS	SS
138	P15825.2	0.21	0.24	-0.07	Log	0.47	-53.01	0.002	129	WW	WS	SS
139	P4800.3	0.19	0.30	0.04	Exp	0.78	-83.33	0.009	221	WW	WS	SS
140	P6937.1	0.37	0.13	0.03	Exp	0.41	-40.63	0.003	102	WW-	WS-	SS+
141	P7108.2	0.45	0.12	0.04	Exp	0.69	-57.93	0.003	82	WW+	WS	SS-
142	P9580.1	0.22	0.24	0.10	Exp	0.87	-87.61	0.011	219	WW+	WS	SS
143	PTC9341	0.25	0.09	-0.06	Exp	0.45	-68.20	0.002	129	WW	WS	SS
144	SNP_GQ0013.BR.1_E01.Contig1.1146	0.24	0.47	0.19	Log	0.68	-58.71	0.004	143	WW	WS-	SS+
145	SNP_GQ0014.BR_A18.Contig1.666	0.41	0.33	0.13	Log	0.76	-41.65	0.005	126	WW+	WS+	SS-
146	SNP_GQ0021.B3.r_E11.Contig1.558	0.20	0.26	-0.08	Log	0.76	-60.19	0.005	162	WW+	WS+	SS-
147	SNP_GQ0021.BR.1_O06.Contig1.333	0.17	0.20	0.12	Log	0.50	-78.31	0.001	86	WW	WS-	SS+
148	SNP_GQ0031.TB_K19.Contig2.238	0.11	0.19	-0.06	Exp	0.78	-98.36	0.01	236	WW	WS	SS
149	SNP_GQ0043.TB_G16.Contig2.1226	0.26	0.19	-0.05	Log	0.70	-64.44	0.002	84	WW	WS+	SS
150	SNP_GQ0044.B3.r_K18.Contig1.396	0.27	0.33	-0.04	Log	0.75	-60.62	0.003	125	WW	WS	SS

Table S2 - Cont'd

#	Locus Name	H_e	F_{ST}	F_{IS}	Reg ^a	R^2	Geographic Clines		λ	Genomic Clines		
							AIC	Max slope		Selection Categories ^b		
151	SNP_GQ0044.B3.r_N02.Contig1.846	0.22	0.04	0.01	Lin	0.06	-59.18	0	-	WW	WS-	SS
152	SNP_GQ0046.B3_H01.Contig1.506	0.27	0.21	-0.01	Exp	0.22	-46.80	0.001	95	WW-	WS	SS+
153	SNP_GQ0048.B3.r_I01.Contig1.195	0.16	0.19	0.05	Exp	0.78	-83.01	0.012	231	WW+	WS	SS
154	SNP_GQ00612.B3_G14.Contig1.819	0.09	0.33	0.06	Exp	0.55	-116.28	0.003	221	WW-	WS+	SS+
155	SNP_GQ00612.B3_J14.Contig1.472	0.26	0.32	0.01	Log	0.84	-63.46	0.006	159	WW+	WS	SS-
156	SNP_GQ0072.B3.r_I18.Contig1.409	0.19	0.05	0.02	Exp	0.47	-82.96	0.002	138	WW	WS	SS
157	SNP_GQ0074.B3.r_L04.Contig1.773	0.37	0.36	-0.02	Log	0.78	-45.01	0.005	123	WW+	WS	SS-
158	SNP_GQ0178.B7_E07.Contig1.180	0.22	0.31	0.07	Log	0.52	-54.34	0.002	104	WW	WS	SS
159	SNP_GQ02010.B3.r_E06.Contig1.520	0.32	0.15	0.02	Exp	0.59	-54.44	0.004	155	WW	WS+	SS-
160	SS_CO483349.contig3.496	0.46	0.14	-0.08	Exp	0.17	-29.46	0.001	< 0	WW-	WS	SS+
161	WS.2.0.GQ0011.B3.r.O22.2.439	0.34	0.33	-0.09	Exp	0.76	-54.40	0.009	199	WW	WS	SS
162	WS.2.0.GQ0013.BR.1.F05.1.445	0.21	0.19	0.01	Log	0.85	-78.82	0.008	190	WW+	WS+	SS-
163	WS.2.0.GQ0013.BR.1.F24.1.457	0.17	0.47	0.06	Exp	0.39	-71.27	0.003	198	WW-	WS	SS+
164	WS.2.0.GQ0013.BR.1.H07.1.1246	0.20	0.33	0.01	Exp	0.26	-52.82	0.002	166	WW-	WS-	SS+
165	WS.2.0.GQ0014.B3.r.K03.1.350	0.35	0.21	0.01	Log	0.56	-30.56	0.004	135	WW+	WS+	SS-
166	WS.2.0.GQ0015.BR.F19.1.1238	0.21	0.18	0.05	Exp	0.71	-67.28	0.008	214	WW	WS	SS
167	WS.2.0.GQ0021.BR.1.G04.1.641	0.27	0.30	0.00	Log	0.79	-66.42	0.005	163	WW	WS	SS
168	WS.2.0.GQ0021.BR.1.I14.1.917	0.12	0.15	0.04	Log	0.70	-83.52	0.003	173	WW+	WS	SS
169	WS.2.0.GQ0023.B3.r.A10.1.304	0.46	0.06	0.02	Exp	0.26	-43.34	0.001	< 0	WW	WS	SS
170	WS.2.0.GQ0024.B3.r.O14.1.374	0.33	0.19	-0.02	Exp	0.72	-73.13	0.004	154	WW	WS	SS
171	WS.2.0.GQ0024.BR.K09.4.220	0.14	0.19	0.03	Exp	0.73	-114.05	0.003	205	WW	WS-	SS+
172	WS.2.0.GQ0025.BR.I12.1.575	0.14	0.14	0.05	Log	0.74	-83.02	0.005	187	WW	WS-	SS+
173	WS.2.0.GQ0025.BR.J23.1.1534	0.23	0.13	0.00	Exp	0.71	-74.06	0.006	194	WW	WS	SS
174	WS.2.0.GQ0031.B3.r.N13.1.1210	0.41	0.23	-0.03	Exp	0.60	-42.05	0.004	132	WW	WS	SS
175	WS.2.0.GQ0031.TB.F08.2.1213	0.26	0.28	0.03	Exp	0.66	-48.92	0.01	214	WW	WS	SS
176	WS.2.0.GQ0032.TB.K21.1.136	0.22	0.13	0.07	Log	0.53	-67.77	0.002	140	WW	WS-	SS+
177	WS.2.0.GQ0033.TB.D14.1.699	0.36	0.28	-0.10	Log	0.63	-40.89	0.003	68	WW	WS+	SS
178	WS.2.0.GQ0034.B3.r.M12.1.702	0.31	0.17	0.03	Exp	0.79	-66.96	0.007	187	WW+	WS+	SS-
179	WS.2.0.GQ0041.BR.J16.4.199	0.29	0.22	0.03	Log	0.56	-42.60	0.004	148	WW	WS	SS
180	WS.2.0.GQ00410.B3.P11.1.1618	0.38	0.39	0.04	Log	0.68	-26.22	0.005	121	WW+	WS	SS-
181	WS.2.0.GQ00411.B3.J14.1.1171	0.18	0.23	-0.01	Log	0.33	-55.46	0.001	92	WW	WS-	SS+

Table S2 - Cont'd

#	Locus Name	H_e	F_{ST}	F_{IS}	Reg ^a	R^2	Geographic Clines		Genomic Clines			
							AIC	Max slope	λ	Selection Categories ^b		
182	WS.2.0.GQ00412.B3.E01.1.1202	0.20	0.28	-0.08	Log	0.72	-71.44	0.005	180	WW	WS-	SS+
183	WS.2.0.GQ00412.B3.K07.1.1479	0.14	0.13	0.03	Exp	0.47	-86.79	0.002	184	WW	WS	SS
184	WS.2.0.GQ00412.B3.M21.1.371	0.39	0.10	0.02	Exp	0.16	-32.80	0.001	< 0	WW	WS-	SS+
185	WS.2.0.GQ00412.B3.P24.3.109	0.35	0.14	-0.05	Exp	0.49	-37.15	0.004	143	WW+	WS	SS-
186	WS.2.0.GQ0043.BR.J01.2.228	0.19	0.03	-0.12	Lin	0.12	-77.05	0	-	WW	WS	SS
187	WS.2.0.GQ0044.B3.r.L23.1.678	0.33	0.21	0.03	Exp	0.55	-59.63	0.003	140	WW-	WS-	SS+
188	WS.2.0.GQ0045.B3.G10.1.344	0.35	0.31	0.03	Log	0.81	-58.33	0.004	126	WW	WS-	SS+
189	WS.2.0.GQ0045.B3.I14.1.573	0.20	0.37	-0.09	Log	0.55	-63.94	0.003	162	WW-	WS-	SS+
190	WS.2.0.GQ0045.B3.N03.1.416	0.33	0.07	-0.05	Exp	0.22	-45.81	0.001	1	WW	WS+	SS-
191	WS.2.0.GQ0045.B3.N10.1.1522	0.21	0.28	-0.28	Log	0.63	-62.20	0.004	165	WW-	WS+	SS+
192	WS.2.0.GQ0045.B3.P14.1.834	0.29	0.40	0.01	Exp	0.66	-54.76	0.007	199	WW-	WS-	SS+
193	WS.2.0.GQ0046.B3.C03.1.551	0.42	0.05	0.00	Exp	0.25	-52.08	0.001	< 0	WW	WS	SS
194	WS.2.0.GQ0047.B3.F06.1.894	0.41	0.21	0.01	Exp	0.72	-44.38	0.007	160	WW+	WS	SS
195	WS.2.0.GQ0049.B3.A02.1.657	0.32	0.18	-0.01	Exp	0.42	-38.25	0.003	146	WW	WS	SS
196	WS.2.0.GQ0061.B3.r.G16.3.334	0.25	0.29	-0.03	Log	0.80	-71.79	0.003	127	WW-	WS+	SS
197	WS.2.0.GQ00611.B3.H11.1.1029	0.22	0.27	0.03	Log	0.77	-67.36	0.007	187	WW	WS	SS
198	WS.2.0.GQ00611.B3.J20.1.130	0.26	0.12	0.01	Log	0.69	-67.67	0.003	146	WW+	WS+	SS
199	WS.2.0.GQ00611.B3.L10.2.622	0.30	0.31	0.00	Log	0.71	-51.79	0.003	109	WW	WS	SS
200	WS.2.0.GQ00612.B3.L21.1.172	0.19	0.31	0.01	Log	0.70	-78.61	0.003	156	WW-	WS-	SS+
201	WS.2.0.GQ0064.B3.r.I13.1.1236	0.40	0.23	-0.05	Exp	0.42	-36.90	0.003	100	WW-	WS	SS+
202	WS.2.0.GQ0064.TB.H03.2.370	0.42	0.35	0.04	Log	0.73	-35.77	0.005	115	WW+	WS+	SS-
203	WS.2.0.GQ0072.B3.r.P11.1.1000	0.49	0.13	-0.04	Exp	0.57	-48.65	0.002	11	WW-	WS	SS+
204	WS.2.0.GQ0073.TB.L02.2.233	0.29	0.15	-0.05	Exp	0.59	-61.27	0.004	157	WW	WS	SS
205	WS.2.0.GQ0073.TB.M05.1.1123	0.16	0.14	0.02	Exp	0.56	-94.38	0.003	188	WW	WS-	SS
206	WS.2.0.GQ0085.B3.r.O08.1.222	0.25	0.18	0.02	Log	0.42	-56.48	0.001	39	WW	WS	SS
207	WS.2.0.GQ0131.B3.E24.1.1764	0.33	0.14	0.00	Log	0.46	-36.59	0.002	15	WW+	WS+	SS-
208	WS.2.0.GQ0133.B7.1.D11.1.1584	0.15	0.22	-0.10	Log	0.77	-96.92	0.003	163	WW-	WS-	SS+
209	WS.2.0.GQ0134.B7.1.L07.1.1358	0.45	0.38	0.00	Exp	0.49	-13.31	0.006	140	WW	WS+	SS+
210	WS.2.0.GQ0161.TB.B13.1.1161	0.20	0.09	0.02	Log	0.75	-87.42	0.002	127	WW+	WS+	SS
211	WS.2.0.GQ0163.TB.B18.1.1080	0.28	0.37	0.08	Log	0.63	-49.22	0.004	148	WW	WS-	SS
212	WS.2.0.GQ0165.B3.F11.2.34	0.43	0.06	0.03	Lin	0.14	-43.76	0	-	WW	WS	SS

Table S2 - Cont'd

#	Locus Name	H_e	F_{ST}	F_{IS}	Reg ^a	R^2	Geographic Clines		λ	Genomic Clines		
							AIC	Max slope		Selection Categories ^b		
213	WS.2.0.GQ0168.B3.J12.1.1192	0.17	0.28	0.05	Log	0.74	-85.35	0.004	172	WW	WS-	SS
214	WS.2.0.GQ0168.B3.N16.1.556	0.17	0.26	0.00	Log	0.51	-75.20	0.001	108	WW-	WS-	SS+
215	WS.2.0.GQ0173.TB.A04.4.594	0.30	0.11	0.00	Exp	0.62	-53.39	0.005	166	WW+	WS+	SS-
216	WS.2.0.GQ0175.B7.K18.1.223	0.09	0.12	-0.04	Log	0.60	-125.84	0.001	140	WW	WS-	SS
217	WS.2.0.GQ0177.B7.K12.1.501	0.24	0.36	-0.01	Log	0.82	-75.92	0.004	141	WW	WS	SS
218	WS.2.0.GQ0178.B7.A11.1.460	0.35	0.25	-0.09	Exp	0.53	-43.66	0.004	149	WW	WS	SS
219	WS.2.0.GQ0187.T24.A06.1.1353	0.19	0.26	-0.05	Exp	0.46	-73.37	0.003	182	WW-	WS-	SS+
220	WS.2.0.GQ0193.B3.r.A11.3.420	0.11	0.15	0.03	Exp	0.50	-103.63	0.002	200	WW-	WS-	SS+
221	WS.2.0.GQ0195.B3.D14.1.174	0.39	0.23	0.05	Log	0.77	-48.56	0.005	128	WW+	WS	SS-
222	WS.2.0.GQ0197.B3.G24.1.764	0.14	0.32	0.03	Exp	0.64	-111.89	0.003	210	WW-	WS-	SS+
223	WS.2.0.GQ0198.B3.P03.1.170	0.32	0.10	0.00	Exp	0.52	-55.03	0.003	129	WW	WS	SS
224	WS.2.0.GQ02010.B3.r.N03.1.1528	0.18	0.17	0.03	Log	0.63	-61.83	0.003	144	WW	WS	SS
225	WS.2.0.GQ02010.B7.H23.1.251	0.49	0.05	-0.07	Lin	0.21	-46.99	0.001	-	WW+	WS	SS-
226	WS.2.0.GQ02011.B3.r.B09.2.447	0.15	0.06	-0.01	Lin	0.09	-85.63	0	-	WW	WS	SS
227	WS.2.0.GQ02013.TB.O16.1.231	0.20	0.11	0.08	Exp	0.22	-56.14	0.001	112	WW	WS	SS
228	WS.2.0.GQ02014.B3.r.H08.1.644	0.19	0.20	-0.09	Log	0.49	-63.86	0.002	116	WW	WS	SS
229	WS.2.0.GQ02015.TB.B10.1.1440	0.28	0.13	0.01	Log	0.62	-48.63	0.002	89	WW	WS-	SS+
230	WS.2.0.GQ02016.B3.r.F09.1.1121	0.28	0.04	-0.02	Lin	0.10	-58.42	0	-	WW+	WS-	SS-
231	WS.2.0.GQ0202.B3.O09.3.261	0.26	0.22	0.00	Log	0.74	-61.13	0.004	152	WW+	WS+	SS-
232	WS.2.0.GQ0204.B3.H10.1.662	0.40	0.25	-0.12	Log	0.61	-41.90	0.002	18	WW	WS+	SS
233	WS.2.0.GQ0204.B3.P14.2.925	0.50	0.08	0.04	Lin	0.25	-36.60	0	-	WW-	WS	SS+
234	WS.2.0.GQ0206.B3.P13.1.173	0.36	0.27	-0.09	Log	0.67	-36.72	0.003	74	WW+	WS+	SS-
235	WS.2.0.GQ0208.B3.P21.1.535	0.30	0.38	0.01	Log	0.68	-39.84	0.006	159	WW+	WS	SS-
236	WS.2.0.GQ0222.B7.B17.1.379	0.37	0.11	0.05	Exp	0.45	-56.24	0.002	65	WW-	WS-	SS+
237	WS.2.0.GQ0222.B7.P03.4.50	0.22	0.32	-0.03	Log	0.55	-40.59	0.003	136	WW	WS	SS
238	WS.2.0.GQ0226.B7.D08.1.418	0.33	0.08	0.01	Exp	0.41	-57.95	0.002	74	WW	WS	SS
239	WS.2.0.GQ0226.B7.D16.1.397	0.26	0.19	-0.04	Log	0.74	-72.08	0.003	114	WW	WS	SS
240	WS.2.0.GQ02511.B3.A11.2.431	0.26	0.21	-0.03	Exp	0.66	-52.44	0.008	201	WW+	WS	SS-
241	WS.2.0.GQ0253.B7.G03.1.1020	0.50	0.07	0.01	Lin	0.33	-38.18	-0	-	WW-	WS	SS+
242	WS.2.0.GQ0255.B3.P02.1.233	0.18	0.24	-0.05	Log	0.52	-81.11	0.001	101	WW-	WS-	SS+
243	WS.2.0.GQ0258.B3.B12.1.786	0.25	0.28	0.00	Exp	0.66	-60.38	0.007	203	WW	WS	SS

Table S2 - Cont'd

#	Locus Name	H_e	F_{ST}	F_{IS}	Reg ^a	R^2	Geographic Clines			Genomic Clines		
							AIC	Max slope	λ	Selection Categories ^b		
244	WS.2.0.GQ02801.B7.O14.1.512	0.37	0.17	-0.03	Log	0.29	-39.78	0.001	< 0	WW-	WS	SS+
245	WS.2.0.GQ02805.B7.J24.2.535	0.28	0.16	-0.02	Exp	0.35	-42.32	0.003	141	WW-	WS-	SS+
246	WS.2.0.GQ02807.B7.A19.1.869	0.45	0.04	-0.02	Lin	0.03	-36.48	0	-	-	-	-
247	WS.2.0.GQ02808.B7.O03.2.818	0.49	0.05	-0.03	Exp	0.18	-36.65	0.001	< 0	WW	WS	SS
248	WS.2.0.GQ02815.B7.M19.1.534	0.16	0.25	-0.08	Log	0.27	-63.02	0.001	76	WW-	WS-	SS+
249	WS.2.0.GQ02819.B7.K02.2.592	0.23	0.20	0.13	Log	0.47	-40.16	0.002	85	WW+	WS+	SS-
250	WS.2.0.GQ02823.SP6.H05.1.827	0.42	0.18	-0.01	Exp	0.49	-31.34	0.004	110	WW+	WS+	SS-
251	WS.2.0.GQ02827.B7.B09.1.298	0.26	0.34	0.05	Log	0.74	-46.72	0.005	149	WW+	WS	SS
252	WS.2.0.GQ02830.B7.N19.1.816	0.19	0.30	-0.04	Log	0.76	-81.91	0.003	161	WW	WS-	SS+
253	WS.2.0.GQ02903.B7.B21.1.1399	0.23	0.37	-0.10	Log	0.60	-59.93	0.005	151	WW-	WS-	SS+
254	WS.2.0.GQ02905.B7.P10.1.849	0.33	0.17	0.02	Exp	0.35	-56.54	0.002	72	WW-	WS	SS+
255	WS.2.0.GQ03101.B7.A12.1.268	0.22	0.26	-0.05	Exp	0.80	-77.33	0.011	222	WW	WS	SS
256	WS.2.0.GQ03101.B7.M09.1.229	0.35	0.07	-0.05	Exp	0.51	-57.96	0.002	87	WW+	WS+	SS-
257	WS.2.0.GQ03105.B7.N08.1.636	0.29	0.25	0.01	Log	0.74	-71.10	0.002	81	WW	WS	SS
258	WS.2.0.GQ03108.B7.H08.1.831	0.31	0.27	0.04	Log	0.76	-44.31	0.005	144	WW+	WS	SS-
259	WS.2.0.GQ03115.B7.P17.1.1218	0.31	0.15	-0.01	Exp	0.65	-56.52	0.005	168	WW	WS	SS
260	WS.2.0.GQ03118.B7.C03.1.798	0.27	0.35	-0.01	Log	0.92	-83.57	0.007	161	WW	WS	SS
261	WS.2.0.GQ03125.B7.D11.2.871	0.27	0.31	0.05	Log	0.77	-60.11	0.004	140	WW	WS-	SS+
262	WS.2.0.GQ03126.B7.M13.1.633	0.16	0.18	0.02	Exp	0.79	-110.53	0.004	204	WW	WS-	SS
263	WS.2.0.GQ03226.B7.M05.1.485	0.33	0.27	0.07	Log	0.52	-52.01	0.002	34	WW-	WS-	SS+
264	WS.2.0.GQ03409.B7.H11.1.187	0.37	0.14	-0.04	Exp	0.62	-50.20	0.004	140	WW	WS	SS
265	WS.2.0.GQ03516.B7.I16.1.170	0.47	0.08	-0.03	Log	0.64	-49.13	0.002	< 0	WW+	WS+	SS-
266	WS.2.0.GQ03614.B7.C22.1.141	0.47	0.15	0.03	Exp	0.59	-28.93	0.005	101	WW+	WS+	SS-
267	WS00841.B21_O11.contig1.NC1.149	0.38	0.07	-0.08	Lin	0.03	-32.77	0	-	WW	WS	SS
268	WS01026.B21_I20.contig1.C1.288	0.12	0.16	-0.05	Exp	0.25	-82.09	0.001	156	WW-	WS	SS+

^a Regression type used in geographic cline analysis, where Log is logistic, Lin is linear and Exp is exponential

^b Genomic cline analysis indicating over (+) or underrepresentation (-) of individual genotypic classes, where WW are interclass homozygotes for white spruce; WS, interclass heterozygotes; and SS interclass homozygotes for Sitka spruce

Figure S1 Relationships between morphological PC1 (Length, **A**), and morphological PC2 (Width, **B**) with geographic and climatic variables spanning the Sitka-white spruce zone of introgression, including drainage distance (km), mean annual temperature ($^{\circ}\text{C}$), mean annual precipitation (mm), and degree days below 0°C

Figure S2 Regression analysis of hybrid index (Q-index), morphology PC1, and morphology PC2 of Sitka x white spruce (*Picea sitchensis* x *P. glauca*) individuals in separate river transects, the Skeena and Nass respectively with drainage distance (km, **A**), mean annual temperature (°C, **B**), mean annual precipitation (mm, **C**), and degree days below 0°C (**D**) The use of linear or non-linear relationships is outlined in Table 3, while the regression slope and intercept, along with significance of the comparison between rivers is outlined in Table 4.

Figure S3 Locus-specific geographic clines for all loci indicating the relationship between minor Sitka spruce allele frequency and drainage distance. Geographic cline parameters indicate the proportion of variance accounted for using a linear or non-linear (logistic or exponential) regression and maximum slope and intercept (lambda, dashed line) for fitted values (solid line).

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S3 Cont'd

Figure S4 Genomic clines for all loci indicating locus-specific patterns of introgression using the genomewide estimate of admixture (hybrid index: 0=white spruce, 1=Sitka spruce) to estimate the probability of observing a particular genotype at that locus, P -values are provided in the right corner of the observed data under a model of neutral introgression. The 95% confidence envelope of the probability of the homozygous white spruce genotype (dark green) and the heterozygous genotype (light green) are based on 1000 neutral parametric simulations. Fitted genomic clines are observed for the homozygous white spruce genotype (solid line) and heterozygous genotype (dashed line), while open circles indicate observed genotypes; either white spruce (WW, top), heterozygous (WS, middle) or Sitka spruce (SS, bottom). The frequency of observed genotypes are indicated on the right of the panel.

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd

Figure S4 Cont'd