Differential introgression reveals candidate genes for selection across a spruce (Picea sitchensis × P. glauca) hybrid zone

Jill A. Hamilton1, Christian Lexer2 and Sally N. Aitken1

1Centre for Forest Conservation Genetics and Department of Forest Sciences, University of British Columbia, Vancouver, BC, Canada; 2Department of Biology, Unit of Ecology and Evolution, University of Fribourg, Fribourg, Switzerland

Summary

- Differential patterns of introgression between species across ecological gradients provide a fine-scale depiction of extrinsic and intrinsic factors that contribute to the maintenance of species barriers and adaptation across heterogeneous environments.
- Introgression was examined for 721 individuals collected from the ecological transition zone spanning maritime to continental climates within the Picea sitchensis–Picea glauca contact zone using a panel of 268 candidate gene single nucleotide polymorphisms.
- Geographic clines showed a strong spatial relationship between allele frequencies and both distance from the ocean along major rivers and mean annual precipitation, indicating a strong role for environmental selection. Interspecific patterns of differentiation using outlier tests revealed three candidate genes that may be targets of long-term divergent selection between the parental species, although contemporary genomic clines within the hybrid zone suggested neutral patterns of introgression for these genes.
- This study provides a fine-scale analysis of locus-specific introgression, identifying a suite of candidate loci that may be targets of extrinsic or intrinsic selection, with broad application in understanding local adaptation to climate.

Introduction

The genetic architecture of hybrid zones offers a unique opportunity to evaluate how gene flow and selection facilitate the spread of genetic material in a natural setting (Barton & Hewitt, 1985; Rieseberg et al., 1999; Lexer et al., 2005, 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 under divergent selection and important to local adaptation. Such analyses provide a fine-scale depiction of genomic patterns influenced by extrinsic or intrinsic factors that contribute to the evolution and maintenance of species barriers (Wu, 2001; Gompert et al., 2012).

Incomplete barriers to reproduction may permit gene flow between species, although genetic exchange will be restricted if recombination is limited or there is strong selection against hybridization (Barton & 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 can identify those regions that may be candidates for differentiation (Wu, 2001; Nosil & Schluter, 2011).

Genome scans examining interspecific differentiation, along with analyses of introgression at both landscape and genome levels, have identified candidate regions important to the maintenance of species barriers in a number of hybrid zones (Scotti-Saintagne et al., 2004; Yatabe et al., 2007; Carling & 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 for identifying genomic regions between species that may be important for local adaptation or the maintenance of species barriers (Lexer et al., 2004; Payseur, 2010; Gompert et al., 2012).

Differential patterns of introgression among loci 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 uses the changes in allele frequencies across the landscape to describe cline shape and estimate the strength of intrinsic and extrinsic (environmental) selection (Barton & Hewitt, 1985; Brennan et al., 2009; Payseur, 2010). More recently, novel admixture mapping techniques have facilitated analysis of introgression of individual genotypes against genome-wide average amounts of admixture...
(Buerkle & Lexer, 2008; Gompert & 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 reflect broad, recent selection dynamics across the hybrid zone, patterns of interspecific differentiation as estimated by locus-specific $F_{ST}$ are informative on a longer timescale (Whitlock, 1992; Lexer et al., 2010). The strength of selection is indicated by the extent to which gene flow counters the accumulation of differentiation over time (Whitlock, 2008). These interspecific and intrahybrid zone patterns combined are complementary and informative at different timescales, 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).

In tree species, where the creation of multiple generations of artificial crosses remains largely impractical because of long generation times, natural hybrid zones provide a range of recombinants with which to examine introgression (Martinsen et al., 2001; Scotti-Saintagne et al., 2004; Lexer et al., 2005). With an increasing call to explore new forest management strategies in the face of climate change, understanding local adaptation in a changing environment will benefit from multilocus genome scans across hybrid zones. Analysis of fine-scale introgression of candidate genes linked to functionally adaptive phenotypes will allow inference of selective forces at play in the maintenance of the introgression zone, the potential transfer of adaptations between species, and identification of genomic regions that maintain species. Furthermore, these data will inform our projections regarding the capacity of natural 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 & Schluter, 2008; De Carvalho et al., 2010). Determining how long-term and more contemporary processes influence and maintain patterns of genomic variation and 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 (*Picea glauca*). 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. 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). The mild, wet maritime climate receives up to 2 m of precipitation annually, with a mean annual temperature of c. 8°C (Pojar et al., 1991). The continental climate is characterized by comparatively lower mean annual precipitation (MAP), and while average annual temperatures are somewhat lower, the temperature range is greater as a result of seasonal extremes: hot, dry summers with average temperatures exceeding 10°C and cold, moist winters with temperatures below 0°C are the norm (Ketcheson et al., 1991).

The objective of this research is to explore variable patterns of introgression and differentiation in this spruce hybrid zone across candidate gene loci putatively associated with adaptive phenotypic traits in forest tree species. Recent identification of candidate gene single nucleotide polymorphic markers (SNPs) putatively associated with adaptive phenotypes in spruce (Namroud et al., 2008; Holliday et al., 2010) has made it feasible to evaluate locus-specific patterns of introgression, differentiation, and adaptation. An earlier study (Hamilton et al., 2012) estimated hybrid index and characterized genome-wide patterns of admixture as well as phenotypic variation across this hybrid zone. Taking a genome scan approach in the current study, we tested for patterns of differentiation amongst loci identified as candidates for differential selection or linked to regions influenced by selection using a combination of approaches. We used $F_{ST}$-outlier tests to identify signatures of divergent or balancing selection among loci between parent species under near-equilibrium conditions. We then combined geographic and genomic clines to test the strength of selection and identify deviations from the genome-wide average as estimated with the studied SNPs. Locus-specific patterns depict fine-scale patterns of spatially varying selection and may identify those regions of the genome that are responsible for reproductive isolation or those important to local adaptation across this hybrid zone.

**Materials and Methods**

**Sampling**

Open-pollinated seeds were sampled from the upper canopy of mature trees across each of 29 locations spanning the contact zone between *P. 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 (Eastern North America, ENA; Table 1, Fig. 1). For each seed parent, 10 open-pollinated progeny were planted in a nursery for 1 yr and then transplanted into a common garden near Kitwanga, British Columbia, within the hybrid zone by the British Columbia Ministry of Forests, Land and Natural Resource Operations (55°17'N, 128°10'W, Fig. 1). Newly flushed lateral buds were sampled from 721 individuals within the common garden in June 2009 and frozen in liquid nitrogen for subsequent DNA extraction and genetic analysis. The same genetic materials were analyzed in Hamilton et al. (2012).

**Genetic data analysis**

DNA extraction was performed using a modified CTAB protocol (Doyle & Doyle, 1990). Fifty milligrams of tissue were ground in liquid nitrogen for extraction using a Mixer Mill MM 400 (Retsch, Newtown, PA, USA). Individuals were genotyped for 384 SNPs identified from previous intraspecific studies of
Descriptive statistics

Global and locus-specific estimates of \( F_{ST} \), expected heterozygosity (\( H_e \)), and inbreeding coefficient (\( F_{IS} \)) were calculated for 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. Linkage disequilibrium (\( D' \)) was estimated for SNPs within the same gene using *R* (R Development Core Team, 2007). Within-population maternal family structure may impact allele frequencies; however, a previous study in oaks indicates that mixing five to six families yields a population sample that is in perfect Hardy–Weinberg equilibrium (Streiff et al., 1999). Furthermore, previous research near the hybrid zone indicates that the effective number of pollen parents is very high (18–20 per mother tree), suggesting the use of open-pollinated progeny likely offers a large sample of paternal haplotypes (Mimura & Aitken, 2007).

Identifying outlier loci potentially affected by selection

To identify candidate gene SNPs that deviate from a null hypothesis of selective neutrality at near-equilibrium conditions between allopatric reference populations (HG and ENA), we used an \( F_{ST} \)-outlier detection method to assess interspecific patterns of differentiation. Based on indications of limited population structure among white spruce populations using a subset of 120 SNPs, we assumed the eastern Canadian white spruce populations are representative of British Columbia populations (Hamilton et al., 2012). Loci that exhibit extreme genetic differentiation relative to parental populations as putative candidates for roles in bud set timing and development of cold-hardiness (Holliday et al., 2008, 2010), weevil resistance (Porth et al., 2011) and growth traits (Namroud et al., 2008; Pavy et al., 2008; Pelgas et al., 2011). SNP selection and quality criteria are described in Hamilton et al. (2012). SNPs were genotyped using the Illumina bead array platform in conjunction with the GoldenGate allele-specific assay (Shen et al., 2012). Of the initial 384 SNPs genotyped, 268 met our quality standards for subsequent analysis.

### Table 1 Origins of Sitka × white spruce (*Picea sitchensis × Picea glauca*) study populations, sample size (\( N \)) and associated geographic distance and climatic variables

<table>
<thead>
<tr>
<th>Population</th>
<th>( N )</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Elevation</th>
<th>Drainage distance (km)</th>
<th>MAT (°C)</th>
<th>MAP (mm)</th>
<th>Continentality (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haida Gwaii (HG)(^1)</td>
<td>24</td>
<td>53.65</td>
<td>−132.20</td>
<td>10</td>
<td>0</td>
<td>7.5</td>
<td>1482</td>
<td>13.6</td>
</tr>
<tr>
<td>Mt Kaien</td>
<td>18</td>
<td>54.30</td>
<td>−130.33</td>
<td>600</td>
<td>0</td>
<td>6.2</td>
<td>3065</td>
<td>10.8</td>
</tr>
<tr>
<td>Seal Cove</td>
<td>18</td>
<td>54.32</td>
<td>−130.32</td>
<td>30</td>
<td>0</td>
<td>8</td>
<td>2901</td>
<td>10.2</td>
</tr>
<tr>
<td>Bella Bella</td>
<td>16</td>
<td>52.17</td>
<td>−128.16</td>
<td>60</td>
<td>0</td>
<td>8.6</td>
<td>2821</td>
<td>11.6</td>
</tr>
<tr>
<td>Ocean Falls</td>
<td>17</td>
<td>52.36</td>
<td>−127.69</td>
<td>60</td>
<td>25</td>
<td>7.8</td>
<td>4437</td>
<td>13.4</td>
</tr>
<tr>
<td>Ramsden Point</td>
<td>30</td>
<td>55.02</td>
<td>−130.13</td>
<td>20</td>
<td>43</td>
<td>7</td>
<td>3086</td>
<td>14.3</td>
</tr>
<tr>
<td>Kinkolth – low</td>
<td>30</td>
<td>55.00</td>
<td>−129.93</td>
<td>30</td>
<td>51</td>
<td>6.6</td>
<td>2979</td>
<td>14.8</td>
</tr>
<tr>
<td>Extew – low</td>
<td>18</td>
<td>54.42</td>
<td>−129.07</td>
<td>30</td>
<td>85</td>
<td>5.6</td>
<td>1691</td>
<td>19.6</td>
</tr>
<tr>
<td>Eskers O/P – low</td>
<td>24</td>
<td>54.41</td>
<td>−129.03</td>
<td>100</td>
<td>85</td>
<td>5.6</td>
<td>1639</td>
<td>19.3</td>
</tr>
<tr>
<td>Eskers O/P – middle</td>
<td>18</td>
<td>54.41</td>
<td>−129.02</td>
<td>450</td>
<td>85</td>
<td>3.9</td>
<td>1581</td>
<td>19.8</td>
</tr>
<tr>
<td>Greenville – low</td>
<td>24</td>
<td>55.09</td>
<td>−129.50</td>
<td>220</td>
<td>87</td>
<td>5.2</td>
<td>1700</td>
<td>18.8</td>
</tr>
<tr>
<td>Kitimat – low</td>
<td>30</td>
<td>54.07</td>
<td>−128.54</td>
<td>100</td>
<td>108</td>
<td>6.3</td>
<td>2722</td>
<td>18.5</td>
</tr>
<tr>
<td>Hagensburg</td>
<td>18</td>
<td>52.39</td>
<td>−126.58</td>
<td>60</td>
<td>115</td>
<td>7</td>
<td>1787</td>
<td>16.6</td>
</tr>
<tr>
<td>Kitimat – high</td>
<td>30</td>
<td>54.20</td>
<td>−128.55</td>
<td>250</td>
<td>119</td>
<td>5.4</td>
<td>2044</td>
<td>19.1</td>
</tr>
<tr>
<td>Nass Camp – low</td>
<td>30</td>
<td>55.30</td>
<td>−129.05</td>
<td>185</td>
<td>126</td>
<td>5.5</td>
<td>989</td>
<td>19.9</td>
</tr>
<tr>
<td>Nass</td>
<td>29</td>
<td>55.28</td>
<td>−129.02</td>
<td>100</td>
<td>127</td>
<td>5.5</td>
<td>1020</td>
<td>20.3</td>
</tr>
<tr>
<td>Douglas Creek – high</td>
<td>30</td>
<td>54.83</td>
<td>−128.73</td>
<td>800</td>
<td>150</td>
<td>2.5</td>
<td>1396</td>
<td>19.8</td>
</tr>
<tr>
<td>Douglas Creek – middle</td>
<td>30</td>
<td>54.83</td>
<td>−128.69</td>
<td>450</td>
<td>153</td>
<td>4.6</td>
<td>1686</td>
<td>19.3</td>
</tr>
<tr>
<td>Kiteen – low</td>
<td>29</td>
<td>55.44</td>
<td>−128.83</td>
<td>310</td>
<td>156</td>
<td>4.4</td>
<td>1083</td>
<td>20.3</td>
</tr>
<tr>
<td>Doneen – low</td>
<td>24</td>
<td>54.83</td>
<td>−128.31</td>
<td>220</td>
<td>166</td>
<td>5</td>
<td>965</td>
<td>20</td>
</tr>
<tr>
<td>Talchako River – low</td>
<td>18</td>
<td>52.15</td>
<td>−125.96</td>
<td>305</td>
<td>180</td>
<td>4.9</td>
<td>907</td>
<td>19.7</td>
</tr>
<tr>
<td>Cedarvale – low</td>
<td>18</td>
<td>50.62</td>
<td>−128.35</td>
<td>110</td>
<td>188</td>
<td>5.4</td>
<td>832</td>
<td>20.3</td>
</tr>
<tr>
<td>Talchako River – high</td>
<td>18</td>
<td>52.06</td>
<td>−125.92</td>
<td>488</td>
<td>191</td>
<td>3.6</td>
<td>985</td>
<td>19.8</td>
</tr>
<tr>
<td>Cedarvale – middle</td>
<td>18</td>
<td>55.03</td>
<td>−128.42</td>
<td>230</td>
<td>192</td>
<td>4.5</td>
<td>894</td>
<td>20.1</td>
</tr>
<tr>
<td>Donaker Creek</td>
<td>18</td>
<td>57.45</td>
<td>−131.08</td>
<td>300</td>
<td>217</td>
<td>1.8</td>
<td>692</td>
<td>26.5</td>
</tr>
<tr>
<td>Kitwanga – low</td>
<td>24</td>
<td>55.37</td>
<td>−128.14</td>
<td>450</td>
<td>220</td>
<td>3.4</td>
<td>640</td>
<td>21.8</td>
</tr>
<tr>
<td>Kitwanga Lake – middle</td>
<td>24</td>
<td>55.36</td>
<td>−128.13</td>
<td>680</td>
<td>221</td>
<td>2.8</td>
<td>673</td>
<td>21.8</td>
</tr>
<tr>
<td>Skeena Crossing</td>
<td>18</td>
<td>55.10</td>
<td>−127.81</td>
<td>210</td>
<td>222</td>
<td>4.9</td>
<td>590</td>
<td>22.8</td>
</tr>
<tr>
<td>Meziadin</td>
<td>18</td>
<td>56.10</td>
<td>−129.30</td>
<td>265</td>
<td>234</td>
<td>4.7</td>
<td>908</td>
<td>21.1</td>
</tr>
<tr>
<td>Moricetown</td>
<td>18</td>
<td>55.03</td>
<td>−127.34</td>
<td>440</td>
<td>277</td>
<td>3.7</td>
<td>540</td>
<td>23.1</td>
</tr>
<tr>
<td>Eastern North America (ENA)(^1)</td>
<td>42</td>
<td>45.48</td>
<td>−76.72</td>
<td>79.2</td>
<td>NA</td>
<td>6.3</td>
<td>914.2</td>
<td>NA</td>
</tr>
</tbody>
</table>

MAT, mean annual temperature; MAP, mean annual precipitation. NA, not applicable.

Reproduced, with permission, from Hamilton et al. (2012).

1Parental populations used to calculate hybrid index.
to average patterns of the majority of loci are candidates for diversifying selection, while those loci that exhibit unusually low degrees of differentiation may be under balancing selection.

We used a Bayesian likelihood method developed by Foll & Gaggiotti (2008) implemented in the program BayeScan to examine patterns of differentiation. This approach teases apart the influence of two models, differentiating locus-specific ($\alpha_i$) and population-specific ($\beta_j$) effects using logistic regression to calculate locus-specific estimates of $FST$ (Foll & Gaggiotti, 2008). This approach is appropriate where gene flow is nonsymmetrical (Beaumont & Nichols, 1996; Nielsen et al., 2009), as previous research has suggested for this hybrid zone (Hamilton et al., 2012).

Positive values of the locus-specific parameter ($\alpha_i$) indicate that locus $i$ may be more differentiated on average, while negative values indicate less differentiation on average. The model including both locus-specific and population ($\alpha_i$ and $\beta_j$) effects was compared with the model specifically testing loci against the neutral model of introgression following 1000 stochastic parametric simulations. A false discovery rate for a $p$-value of 0.05 was calculated using qvalue (Storey, 2002). Evidence for an excess (+) or deficit (−) of homozygous genotypes (WW and SS) corresponds, respectively, to an increase or decrease in the total probability density of the observed genotype compared with the corresponding probability density of the homozygous genotype of the neutral model (Gompert & Buerkle, 2009, 2010; Nolte et al., 2009). Similarly, evidence for an excess (+) or deficit (−) of heterozygous genotypes (WS) is obtained from comparing the observed frequency of heterozygous genotypes with the probability density of heterozygous genotypes under the neutral model.

Genomic cline analysis

Genomic-wide estimates of admixture were calculated using the R package Introgress (Gompert & Buerkle, 2010), which uses a maximum-likelihood approach to estimate ancestries of individuals summarized in a hybrid index. This program requires a priori knowledge of pure parental individuals, here identified as those reference populations HG (Sitka spruce) and ENA (white spruce), 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 & Buerkle (2009) implemented in Introgress (Gompert & Buerkle, 2010), genomic clines were estimated for individual loci based on the probabilities of observing three possible genotypes at each locus as a function of genome-wide estimates of admixture (hybrid index): WW (homozygous white spruce), WS (heterozygous genotype), or SS (homozygous Sitka spruce). These genomic clines are estimated using multinomial regression of observed single-locus genotypes as a function of genome-wide admixture (Gompert & 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 & Buerkle (2010).

Deviations from the genome-wide average as estimated with the SNPs are based on departures from the 95% confidence envelope for the logistic regressions as summarized through comparisons of the probability density of observed genotypic data with those of the neutral model of introgression following 1000 stochastic parametric simulations. A false discovery rate for a $p$-value of 0.05 was calculated using qvalue (Storey, 2002). Evidence for an excess (+) or deficit (−) of homozygous genotypes (WW and SS) corresponds, respectively, to an increase or decrease in the total probability density of the observed genotype compared with the corresponding probability density of the homozygous genotype of the neutral model (Gompert & Buerkle, 2009, 2010; Nolte et al., 2009). Similarly, evidence for an excess (+) or deficit (−) of heterozygous genotypes (WS) is obtained from comparing the observed frequency of heterozygous genotypes with the probability density of heterozygous genotypes under the neutral model.

Geographic cline analysis

The geographic gradient was estimated using drainage distance along 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; Hamilton et al., 2012). Drainage distance characterizes the likely corridor of gene flow along valleys through which pollen, and to a lesser extent seed dispersal, occurs between the coastal maritime climate and the interior continental climate (Bennuah et al., 2004). Drainage distance was calculated from the coast inland using ArcGIS (Version 10), replicating methods from Bennuah et al. (2004). The pure white spruce reference population from Ontario (ENA) was excluded from this analysis as a result of its long distance from all other populations.

Geographic cline parameters were estimated individually for each SNP using a modified approach to traditional analysis. Traditional cline analysis assumes a sigmoidal shape that describes sharp changes in allele frequencies near the centre of the cline and shallower asymptotes towards the edges of the contact zone (Barton & Hewitt, 1985; Payseur, 2010). The most common parameters used to describe cline shape are width and centre (Porter et al., 1997; Carling & 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 (Picea engelmannii; Roche, 1969). The lack of pure white spruce populations in our sample is evidenced by an exponential increase in white spruce 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 white spruce parent populations.

This novel approach has been adapted from equations found in biological growth curves that allow estimation of maximum slope of a nonlinear regression and ‘length of lag phase’ or lambda (Kahn 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). Steeper slopes indicate stronger barriers to introgression between species, whereas flatter slopes may indicate either extensive allele-sharing through recurrent gene flow or shared polymorphisms. ‘Length of lag’ is interpreted as the drainage distance accumulated from the coast to the point of maximum rate of change in allele frequency. This parameter associates a geographic estimate on the landscape with maximum rate of change in allele frequency.

Linear regressions were compared with nonlinear regressions estimated using both asymptotic (logistic) and nonasymptotic (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). Akaike information criterion (AIC) and R² values were compared between regression models to identify 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 fitted a smoothed cubic spline using the gcFitSpline function within the R package grofit (Kahn 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).

Enrichment of heterozygotes and homozygotes for genomic clines and directional shifts were assessed in geographic and genomic clines to identify those loci that may be candidates for involvement in reproductive isolation or adaptation. Where similar patterns were observed in geographic and genomic clines, the distribution of genotypes was associated with a previously identified important climatic variable, MAP (Hamilton et al., 2012) across the hybrid zone.

Results

The global estimate of mean genetic differentiation (FST) across all populations sampled was 0.18, ranging from 0.02 to 0.47 for individual loci (Supporting Information, Table S1). Expected heterozygosity (HE) averaged 0.31 across all loci, ranging from 0.04 to 0.5. The inbreeding coefficient (FIS) averaged 3.73 × 10⁻³, 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 indicated that the majority of variation is accounted for between species (Table 2). Overall, the estimate of FST between allopatric Sitka spruce (HG) and white spruce (ENA) was 0.59, indicating that interspecific differentiation exceeded those where all populations, sympatric and allopatric, were evaluated. These results are in line with locus-specific estimates of interspecific differentiation values between the allopatric populations, ranging from zero to one.

Outlier analysis

Using BayeScan, which uses a Bayesian decision-making approach to model choice with 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}(BF) > 1) of being subject to selection. All three SNPs (53, 10

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>df</th>
<th>Sum of squares</th>
<th>Variance components</th>
<th>% of variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between species</td>
<td>1</td>
<td>2902.52</td>
<td>46.97</td>
<td>58.70</td>
</tr>
<tr>
<td>Among individuals</td>
<td>64</td>
<td>2126.18</td>
<td>0.17</td>
<td>0.21</td>
</tr>
<tr>
<td>within species</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within populations</td>
<td>66</td>
<td>2170.00</td>
<td>32.88</td>
<td>41.09</td>
</tr>
<tr>
<td>Total</td>
<td>131</td>
<td>7198.70</td>
<td>80.02</td>
<td></td>
</tr>
</tbody>
</table>

Table 2 AMOVA for 268 candidate gene single nucleotide polymorphisms (SNPs) between Picea sitchensis and Picea glauca
55, and 106) are candidates for or are linked to regions involved in the transfer of adaptations contributing to transgression in adaptive traits. Where heterozygotes are underrepresented, these loci may be candidates for regions of the genome involved in reproductive isolation.

Genetic map relationships among these markers remain unknown, and therefore some values may be inflated if linkage disequilibrium (LD) exists between a marker and an adjacent genetic region under selection. SNP frequencies within the same candidate gene indicate that physical proximity likely contributes to LD (i.e., SNP 53 and 55 – , ). Preliminary assessment of genetic map locations for a subset of SNPs used in this study suggests the loci used are well distributed throughout the genome (J. Bousquet & N. Isabel, unpublished).

**Geographic clines**

While patterns of interspecific differentiation may reveal longterm influences of selection, direct examination of the introgression zone using clinal analysis may provide additional power to detect targets of selection under nonequilibrium conditions. Geographic cline parameters, slope, and the ‘lag phase’ (lambda) were estimated for each SNP (Table 3, Fig. S2). Cline parameters reflect the change in frequency of the minor ‘Sitka spruce’ allele (the less frequent allele in Sitka spruce corresponding to the more frequent allele in white spruce) with increasing drainage distance from the ocean. Maximum slope of regressions, describing changes in allele frequency of drainage distance, averaged 0.003 km⁻¹ and ranged from −0.001 to 0.012 (Table 3). Lambda 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 of lambda ranged from 0 to 236 km, with an average of 119 km across all loci. Negative drainage distance values were set to zero for 22 loci, and lambda was not estimated for 45 loci where the slope was estimated as zero. Few loci exhibited maximum cline slopes >0.011 km⁻¹, suggesting the absence of strong or complex barriers to gene flow between the two species.

For loci where patterns were comparable between geographic and genomic clines, the distribution of genotypes across the hybrid zone was strongly associated with MAP (Fig. 3). Analysis of variance indicated that locus-specific genotypes were significantly differentiated based on precipitation ( ). Sitka spruce genotypes spanned a range above 1000 mm, white spruce genotypes were distributed below 1000 mm, and heterozygous genotypes had a small range in precipitation centered at c. 1000 mm, with little overlap with either homozygous genotypes (Fig. 3).

**Discussion**

Differential patterns of introgression exhibited within this natural hybrid zone suggest that targets of selection may vary strongly geographically across the landscape and between divergent genomic backgrounds. Substantial heterogeneity, both across the geographic landscape and within the genome itself, indicate the general permeability of these spruce genomes to introgression, yet
identify specific genomic regions and geographic barriers where interspecific recombination may be limited. Moreover, fine-scale variability amongst loci indicates that both extrinsic and intrinsic selective factors likely influence introgression. Correspondence between geographic and genomic clines for some SNPs indicates a strong role for extrinsic selection, particularly in association with precipitation gradients. In addition, patterns of differentiation identified through outlier analysis point toward candidate gene loci that may be linked to genes important to local adaptation or the maintenance of species barriers over a longer evolutionary period. These results provide a fine-scale picture of the movement of genetic material across this ecological transition zone, identifying regions of the genome that warrant further exploration.

**Interspecific patterns of genetic differentiation**

Selection can be inferred from the distribution of genetic differentiation among species or populations 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 either reproductive isolation or adaptation (Beaumont & Nichols, 1996; Beaumont & Balding, 2004). We observed few candidate genes for divergent selection (Fig. 2); although some of these candidates may be linked to loci under divergent selection (Via, 2012).

The use of SNPs preselected from putatively adaptive genes developed from intraspecific studies requires a caveat for development of a neutral model of no-locus effect. Few outlier loci, combined with evidence from Arabidopsis thaliana that suggest different genes may be targets of selection in different environments (Hancock et al., 2011), indicate that these loci are likely reasonable for development of a neutral distribution. While the interpretation of posterior probabilities provides evidence for selection, other neutral processes may contribute to the patterns of variation observed. However, in a continuously distributed, wind-pollinated study system with large effective population sizes, these processes are likely less important. Consequently, while these test results should be interpreted with some caution, they offer a useful first step in pointing towards some candidate genes that may be targets of or linked to targets of selection. Indeed, the reported number of outliers may be conservative, owing to the use of candidate genes associated with putative function from gene expression studies. As these candidate loci have a higher probability of being under selection, shifting the distribution of $F_{ST}$ estimates upwards compared with purely neutral loci, some loci under weak selection may not be detected.

Our results for the Sitka-white spruce hybrid zone across an east–west precipitation and temperature gradient identified different candidates for selection than those observed along a north–south temperature gradient in Sitka spruce by Holliday et al. (2010). SNPs 53 and 55 are found in a gene homologous to sphingolipid desaturase (SLD) identified in Arabidopsis (At2g46210.1), and are c. 100 bp apart. The strong LD between SNPs 53 and 55 ($D' = 0.99, P < 0.0001$) implies that one of these SNPs or a locus linked to both is likely the target of selection. These SNPs have been assayed in previous studies examining intraspecific variation and SNP–phenotype associations in the development of autumnal cold-hardiness and bud set timing in Sitka spruce (Holliday et al., 2008, 2010). Holliday et al. (2010) studied populations spanning gradients in Sitka spruce across the species’ range of mean annual temperature and degree-days associated with phenological cues and development of cold hardiness. This gene, believed to be a fatty acid/sphingolipid desaturase, produces divergent phenotypes in Arabidopsis in response to

<table>
<thead>
<tr>
<th>Homozygous genomic clines</th>
<th>Heterozygous genomic clines</th>
<th>$N$</th>
<th>$F_{ST}$</th>
<th>Slope (per km)</th>
<th>Lambda (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WW+ SS−</td>
<td>WS+</td>
<td>21</td>
<td>0.20 (0.07–0.36)</td>
<td>0.004 (0.001–0.008)</td>
<td>116 (0–190)</td>
</tr>
<tr>
<td>WW− SS+</td>
<td>WS+</td>
<td>14</td>
<td>0.20 (0.05–0.39)</td>
<td>0.004 (0.000–0.008)</td>
<td>115 (0–201)</td>
</tr>
<tr>
<td>WW− SS−</td>
<td>WS+</td>
<td>8</td>
<td>0.12 (0.03–0.44)</td>
<td>0.002 (0.000–0.009)</td>
<td>41 (0–163)</td>
</tr>
<tr>
<td>WW− SS+</td>
<td>WS+</td>
<td>8</td>
<td>0.23 (0.10–0.43)</td>
<td>0.003 (0.000–0.005)</td>
<td>137 (0–221)</td>
</tr>
<tr>
<td>WW− SS−</td>
<td>WS+</td>
<td>16</td>
<td>0.15 (0.04–0.47)</td>
<td>0.001 (−0.001 to 0.003)</td>
<td>57 (0–198)</td>
</tr>
<tr>
<td>WW− SS+</td>
<td>WS+</td>
<td>28</td>
<td>0.26 (0.10–0.47)</td>
<td>0.003 (0.000–0.009)</td>
<td>144 (34–220)</td>
</tr>
<tr>
<td>WW− SS−</td>
<td>WS+</td>
<td>4</td>
<td>0.10 (0.08–0.12)</td>
<td>0.002 (0.001–0.003)</td>
<td>117 (77–146)</td>
</tr>
<tr>
<td>WW− SS+</td>
<td>WS+</td>
<td>14</td>
<td>0.21 (0.09–0.34)</td>
<td>0.006 (0.000–0.012)</td>
<td>167 (96–231)</td>
</tr>
<tr>
<td>WW− SS−</td>
<td>WS+</td>
<td>3</td>
<td>0.10 (0.04–0.16)</td>
<td>0.002 (0.000–0.003)</td>
<td>73 (0–146)</td>
</tr>
<tr>
<td>WW− SS+</td>
<td>WS+</td>
<td>2</td>
<td>0.17 (0.05–0.29)</td>
<td>0.002 (0.000–0.003)</td>
<td>127 (127–127)</td>
</tr>
<tr>
<td>WW− SS+</td>
<td>WS+</td>
<td>2</td>
<td>0.26 (0.13–0.38)</td>
<td>0.004 (0.001–0.006)</td>
<td>70 (0–140)</td>
</tr>
<tr>
<td>WW− SS−</td>
<td>WS+</td>
<td>1</td>
<td>0.19 (0.19–0.19)</td>
<td>0.003 (0.003–0.003)</td>
<td>59 (50–50)</td>
</tr>
<tr>
<td>WW− SS+</td>
<td>WS+</td>
<td>12</td>
<td>0.23 (0.10–0.47)</td>
<td>0.003 (0.001–0.005)</td>
<td>129 (0–205)</td>
</tr>
<tr>
<td>WW− SS−</td>
<td>WS+</td>
<td>5</td>
<td>0.19 (0.07–0.25)</td>
<td>0.003 (0.001–0.004)</td>
<td>77 (1–155)</td>
</tr>
<tr>
<td>WW− SS+</td>
<td>WS+</td>
<td>7</td>
<td>0.16 (0.03–0.25)</td>
<td>0.003 (0.000–0.005)</td>
<td>100 (46–119)</td>
</tr>
<tr>
<td>WW− SS−</td>
<td>WS+</td>
<td>9</td>
<td>0.16 (0.03–0.29)</td>
<td>0.002 (0.000–0.006)</td>
<td>97 (18–174)</td>
</tr>
<tr>
<td>WW− SS+</td>
<td>WS+</td>
<td>94</td>
<td>0.19 (0.02–0.43)</td>
<td>0.003 (0.000–0.011)</td>
<td>118 (0–236)</td>
</tr>
<tr>
<td>WW− SS−</td>
<td>WS+</td>
<td>16</td>
<td>0.21 (0.04–0.45)</td>
<td>0.003 (0.000–0.011)</td>
<td>159 (84–215)</td>
</tr>
</tbody>
</table>

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 genome-wide ancestry, or neutral. Mean and range for each class are provided for estimates of genetic differentiation ($F_{ST}$), slope (per km), and lambda (km).
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). This candidate gene was not associated with bud set or cold injury traits within Sitka spruce by Holliday et al. (2010). These results may provide support for different genes being targets of selection in different environments (Barrett & Hoekstra, 2011; Hancock et al., 2011).

SNP106 (GAI, Gibberellic Acid Insensitive), with putative function in Arabidopsis in gibberellic acid-mediated signaling (At1g14920.1), also exhibited excess interspecific differentiation. Interestingly, this was one of a number of candidate genes identified from expression studies of resistance in Sitka spruce to the white pine shoot tip weevil (Pissodes strobi, Ralph et al., 2006). Given gibberellins' role 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 involved in defense (Richards et al., 2001). Previous studies have observed significant differentiation between parental species in the formation of resin canals (O’Neill et al., 2002). Association tests within a phenotyped population of advanced-generation introgressed individuals are needed to further test these loci.

Genomic clines

Using genomic clines, we compared deviations from a model of neutral introgression for individual loci indicated by an excess or deficit of specific genotypic classes given expectations based on genome-wide ancestry across the hybrid zone. Loci exhibited a diversity of patterns ranging from under- to over-representation of certain genotypes. The single most common category across homozygous and heterozygous genotypes fitted the neutral model of introgression (WW WS SS, 94/264 loci) although a majority of loci displayed a diversity of patterns inconsistent with neutral introgression. The most common categories deviating from a pattern of neutral introgression were a deficit (WW/C0 WS/C0 SS/C0, 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: white spruce (WW, top), heterozygous (WS, middle) or Sitka spruce (SS, bottom). The frequency of observed genotypes is indicated on the right of the panel. (b) Geographic clines indicate the relationship between the frequency of the minor Sitka spruce allele and drainage distance (km). Geographic cline parameters indicate the proportion of variance accounted for using a linear or nonlinear (logistic or exponential) regression and maximum slope and intercept (lambda, dashed line) for fitted values (solid line). (c) Distribution of genotypes with respect to mean annual precipitation (mm) across the hybrid zone.
and an excess of white spruce genotypes (WW+ WS+ SS−, 21/264). These loci may point towards candidate regions of the genome involved in either reproductive isolation or adaptive introgression, respectively, although we cannot exclude the influence of neutral evolution. Surprisingly, all three previously described FST-outlier loci exhibit neutral patterns of introgression in the genomic cline analysis. However, the outlier approach analyzes the long-term accumulation of differentiation between allopatric populations, while the genomic cline analysis reveals the more contemporary influence of locus-specific selection. Furthermore, while FST-outlier tests may reveal loci under selection, there are many conditions where these tests may fail to identify loci involved in adaptation (Le Corre & Kremer, 2012), particularly for polygenic traits, pointing towards the importance of a combined analytical approach.

While a false discovery rate of 0.007 at α = 0.05 indicates few false positives, sampling design may have influenced the patterns we detected (Storey, 2002; Teeter et al., 2010). Few individuals were sampled from the eastern end of the introgression zone, and this likely impacted the observed frequency of white spruce genotypes. However, because genomic clines reflect locus-specific introgression with respect to genome-wide ancestry across all loci, the patterns reflect the rate of introgression given present ancestry. In addition, asymmetric introgression toward Sitka spruce (Hamilton et al., 2012) may result in selection against white spruce genotypes where strong pre- or postzygotic barriers exist. A deficit of heterozygotes (WW WS− SS) was observed in 16 loci. These loci may be within or linked to genes that result in decreased fitness of heterozygotes, although whether this is a result of intrinsic or extrinsic genetic factors remains unclear. Selection against heterozygotes may indicate either simple underdominance or intrinsic Dobzhansky–Müller incompatibilities resulting from heterozygote incompatibility (Teeter et al., 2010). SNP16 had a very steep geographic cline, exhibiting a pattern of underdominance, further supporting comparisons between geographic and genomic clines. This SNP has putative function within a Scarcecrow-Like protein (SCL3, At1g50420.1), which interacts with the gibberellin (GA) pathway, regulating growth in Arabidopsis (Zhang et al., 2011).

Few markers (9/264) exhibited an excess of heterozygotes relative to neutral expectation. These loci, however, may indicate an adaptive advantage for hybrid genotypes (Nolte et al., 2009). Transgressive patterns exhibited in hybrids for cold tolerance at moderately cold temperatures in a previous study suggest higher fitness of hybrid genotypes within the hybrid zone (Hamilton et al., 2012). This lends support to the role of hybrid zones in the transfer of adaptations between species (Rieseberg & 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, although some caveats are necessary. While the majority of loci exhibit deviations from a neutral model of introgression, some of these could result from genetic drift alone, particularly if drift occurred independently in different populations. However, it is unlikely that drift would overwhelm selection, particularly given the large effective 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 overestimated as a result of LD among loci, which is expected to be substantial within hybrid zones ( Lexer et al., 2007). Estimates of LD between SNPs on the same gene (two to four SNPs per gene) indicate that those SNPs are significantly linked (average D 2 = 0.89 across 18 genes, P < 0.0001) and so signatures of selection may reflect nearby polymorphisms rather than specific SNPs. In future studies, estimation of linkage relationships and the degree of recombination following first contact will be possible as genetic maps and a whole genome sequence become available for spruce. Finally, distinguishing loci under neutral introgression from those loci under positive or negative selection is integral to the genomic clines method. With interspecific crosses we expect more loci to have negative fitness consequences than positive ones (Barton, 2001). There may be a slight bias toward overestimating positively selected loci in the hybrid zone and underestimating negatively selected ones if rates of neutral introgression are underestimated (Rieseberg et al., 1999). This suggests our estimate of the frequency of loci exhibiting negative selection is conservative.

Geographic clines

The ecological transition from maritime to continental climates is reflected at a finer spatial scale in the substitution of largely Sitka spruce alleles for white spruce alleles with increasing drainage distance up river valleys. While there is variation among locus-specific clines, on average the maximum rate of change in allele frequencies (0.003 km−1) occurs at c. 120 km from the coast. This distance is concordant with the beginning of a sharp decline in MAP, suggesting that moisture may be a strong selective factor across this ecological gradient. Our previous research indicates that ancestry is influenced by precipitation and temperature gradients spanning the ecological transition zone (Hamilton et al., 2012). Particularly sharp clines in genome-wide ancestry associated with precipitation indicate that the intensity and direction of selection may change below a threshold of c. 1000 mm of precipitation annually (Hamilton et al., 2012).

We have estimated proxies for the most commonly used cline parameters: maximum slope, indicating the maximum rate of change in allele frequency km−1 drainage distance; and lambda, the geographic point on the landscape where we observe the greatest rate of change. Loci that exhibit shallower clines may be candidates for neutral introgression, while steep clines are more likely to be attributed to selection against hybridization, indicating intrinsic Dobzhansky–Mueller incompatibilities or extrinsic (environmental) incompatibilities limiting introgression. Teasing apart the influence of intrinsic or extrinsic incompatibilities would require testing the fitness of progeny from controlled crosses in multiple environments across the hybrid zone.

Relating geographic and genomic clines

The five candidate gene loci (16, 137, 142, 153, and 255, Fig. 3) that exhibited steepest clines (maximum clines slopes of
0.011–0.012 kmâ€’1) are candidates for divergent selection or reproductive isolation or are linked to such genes. Only SNP16, associated with regulating gibberellin signaling and plant development (Zhang et al., 2011), exhibited a pattern of selection against heterozygotes (WSâ€’), while SNPs 137, 142, and 153 exhibited positive selection for the white spruce genotypes (WW+) and SNP255 was neutral (Fig. 3). This may indicate that SNP16 is a candidate for or is linked to a locus involved in intrinsic Dobzhansky–Mueller incompatibilities resulting in hybrid inferiority or breakdown (Carling & Brumfield, 2008; Schluter & Conte, 2009; Teeter et al., 2010). SNPs 137, 142, and 153, on the other hand, all exhibited the greatest rate of change in allele frequency between 213 and 220 km, along with an excess of white spruce genotypes and a lack of heterozygotes in a white spruce background (Fig. 3). This geographic region again corresponds to the drier end of the steep precipitation gradient that begins c. 120 km inland from the ocean (Hamilton et al., 2012). At 200 km drainage distance from the coast, all MAP values fall below 1000 mm of annual precipitation. An ANOVA comparing homozygous and heterozygous genotypes with MAP suggests that genotypes across the hybrid zone are distributed based on precipitation (P<0.0001, Fig. 3). Both SNP137 (Plasmamembrane intrinsic protein (PIP2E), At2g39010.1) with putative function in water channel activity and SNP255 (early-response-to-drought-6 (erd6), At1g08930.2) are homologous to genes important to drought response in Arabidopsis. Signatures of selection in candidate genes related to drought stress in Pinus taeda identified a related early-response-to-drought-3 (erd3) gene as potentially adaptive (Gonzalez-Martinez et al., 2006). This suggests these genes may be important in terms of drought response across species.

While Sitka spruce genotypes and heterozygous genotypes are common across a range of precipitation values exceeding 1000 mm, there appears to be a threshold below which the white spruce is the sole genotype observed. Similar trends were observed for mean summer precipitation and precipitation as snow; however, the relationship with these climatic variables was weaker and did not reflect the abrupt transition seen for MAP. These results suggest the distribution of some alleles across the landscape may be strongly influenced by extrinsic selection in drier environments, but not in wetter areas, where they may be neutral.

Studies of Arabidopsis thaliana populations sampled across steep environmental gradients show some SNPs are neutral in one test environment, but adaptive in another (Fournier-Level et al., 2011; Hancock et al., 2011). Our results point towards a similar dynamic within this 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. 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 (Fig. 3).

Examining differential introgression using interspecific differentiation combined with geographic and genomic clines identifies heterogeneous patterns among loci across the genome, suggesting porous spruce genomes. Not unexpectedly, these tests identify different subsets of candidate genes under selection, depending on gene action, strength of selection, and distribution of underlying environmental variation. This points towards the importance of using multiple approaches to investigate SNPs that may be under selection. 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 across the landscape validate patterns of introgression observed into varying genomic backgrounds, identifying an important role for environmental selection, particularly along precipitation gradients. These approaches identify a suite of candidate gene SNPs for future inclusion in association and functional assays. This fine-scale approach provides the context for broader questions regarding the maintenance of species barriers in these long-lived species, distinguishing regions of the genome where selection may prevent introgression. Identifying the genomic regions where differences accumulate in this and other conifer hybrid zones will provide an opportunity to identify loci and genomic regions that maintain species.

Acknowledgements

We would like to thank John King for establishing the common garden; Christine Chourmouzis, Lisa Erdle, Nina Lobo, and Jon Sweetman for field assistance; Dorothea Lindtke for advice with analysis; and Loren Rieseberg and Kermit Ritland for helpful manuscript comments. The work was supported by Genome British Columbia, Genome Canada and the Forest Genetics Council of British Columbia (funding to S.A.) by the Natural Sciences and Engineering Research Council of Canada (NSERC; grant to S.A.), and by an NSERC Canada Graduate Scholarship and University of British Columbia Doctoral Fellowship to J.H.

References


R Development Core Team. 2007. *R: a language and environment for statistical computer*. Vienna, Austria: Computing RFS.


**Supporting Information**

**Fig. S1** Genomic clines for all loci indicating locus-specific patterns of introgression using the genome-wide estimate of admixture to estimate the probability of observing a particular genotype at that locus.

**Fig. S2** Locus-specific geographic clines for all loci indicating the relationship between minor Sitka spruce allele frequency and drainage distance.

**Table S1** Summary of locus-specific patterns across the Sitka–white spruce zone of introgression, including genetic data analysis, and geographic and genomic cline analysis for 268 candidate gene SNPs.