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

Jill A. Hamilton[1], Christian Lexer[2] and Sally N. Aitken[1]

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

Summary

Author for correspondence:
Jill Hamilton
Tel: +1 604 822 9102
Email: jillahamilton@gmail.com

Key words: adaptation, ecological gradient, genomic cline, geographic cline, hybrid zone, selection, single nucleotide polymorphisms (SNPs), spruce.

- 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 counteracts 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

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

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

MAT, mean annual temperature; MAP, mean annual precipitation. NA, not applicable.
Reproduced, with permission, from Hamilton *et al.* (2012).
[1]Parental populations used to calculate hybrid index.

parental species 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.*, 2005; Fan *et al.*, 2006). Genotyping quality was evaluated using the GenomeStudio Genotyping Module (v1.0) as described in detail in Hamilton *et al.* (2012). Of the initial 384 SNPs genotyped, 268 met our quality standards for subsequent analysis.

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
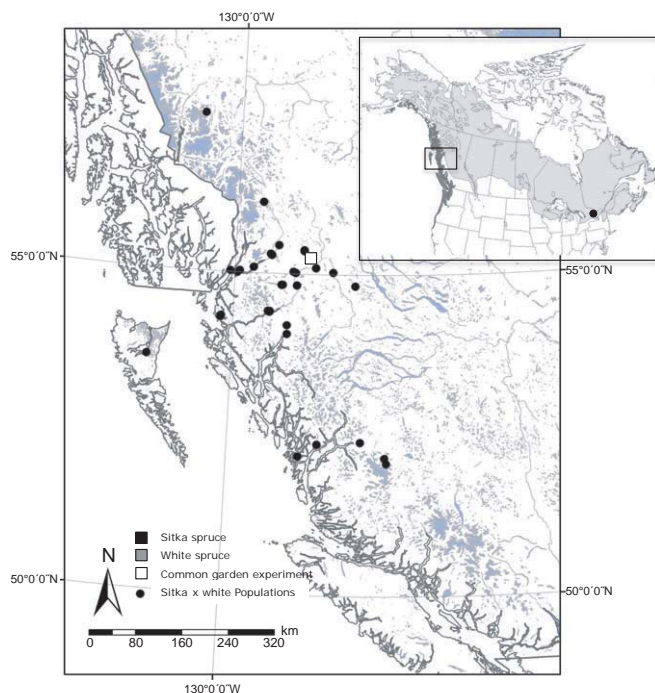
Fig. 1 Sampling across the introgression zone between Sitka (*Picea sitchensis*) and white (*Picea glauca*) spruce in northwestern British Columbia, indicating provenance of origin of samples (circle) collected and planting location of the common garden experiment (square). Inset map provides North American range of *P. sitchensis* (dark grey) and *P. glauca* (light gray), and indicates source of the reference Eastern North American population. Reproduced, with permission, from Hamilton *et al.* (2012).

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 (α_i) and population-specific (β_j) effects using logistic regression to calculate locus-specific estimates of F_{ST} (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 (α_i) indicate that locus i may be more differentiated on average, while negative values indicate less differentiation on average. 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, assuming equal prior probabilities for each model. The ratio of the marginal likelihoods of the two models provides a Bayes' factor (BF; Foll & Gaggiotti, 2008), which may be interpreted in terms of amount of support for loci being subject to selection (very strong support between $\log_{10}(BF) = 1.5$ and 2; decisive support for $\log_{10}(BF) \geq 2$) 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 5000 iterations, we used

50 000 additional iterations (sample size of 5000 and thinning interval of 10) to identify loci that may be candidates for selection. While these tests are informative, they remain subject to other influences of variation, including population demography and hierarchical population structure, and can result in overestimation of selection (Excoffier *et al.*, 2009; Siol *et al.*, 2010).

Genomic cline analysis

Genome-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 (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). 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 R^2 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 (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).

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 (F_{ST}) across all populations sampled was 0.18, ranging from 0.02 to 0.47 for individual loci (Supporting Information, Table S1). Expected heterozygosity (H_e) averaged 0.31 across all loci, ranging from 0.04 to 0.5. The inbreeding coefficient (F_{IS}) averaged 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 indicated that the majority of variation is accounted for between species (Table 2). Overall, the estimate of F_{ST} 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,

Table 2 AMOVA for 268 candidate gene single nucleotide polymorphisms (SNPs) between *Picea sitchensis* and *Picea glauca*

Source of variation	df	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	

55, and 106) are candidates for or are linked to candidates for diversifying selection (Fig. 2). There were no candidates for balancing selection, an expected outcome given that loci were initially selected based on degree of interspecific differentiation in SNP frequencies (F_{ST}).

Genomic cline analyses

Extensive heterogeneity in genomic clines was observed, indicating many loci deviated considerably from the genome-wide average as estimated from all SNPs (Tables 3, S1, Fig. S1). All loci were assessed for deviations from the genome-wide average estimate of ancestry and null model of neutral introgression, excluding loci that lacked allele frequency differences between parent species. Of these, 94 exhibited a neutral pattern of introgression, indicating no deviation from the genome-wide estimate of ancestry. Other patterns included 64 loci exhibiting evidence for excess ancestry of white spruce alleles in a Sitka spruce background (WW+, Table 3, Fig. S1) and 67 loci with an excess of Sitka spruce alleles in a white spruce background (SS+, Table 3, Fig. S1), indicating potential positive selection. Sixty-seven loci had patterns consistent with a deficit of heterozygous genotypes (WS−), while 51 loci showed an excess of heterozygotes (WS+) compared with the model of neutral introgression (Table 3, Fig. S1). Of these 118 loci, two small groups exhibited a pattern of deficiency (16 loci, WS−) and excess (nine loci, WS+) of heterozygotes, respectively, while homozygous genotypes exhibited no deviation. Loci where heterozygotes are overrepresented may be candidates for or are linked to genomic 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. SNPs within the same candidate gene indicate that physical proximity likely contributes to LD (i.e. SNP 53 and 55 – $D' = 0.99$, $P < 0.0001$). 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 long-term 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 '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 km^{-1} 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 ($P < 0.001$). 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
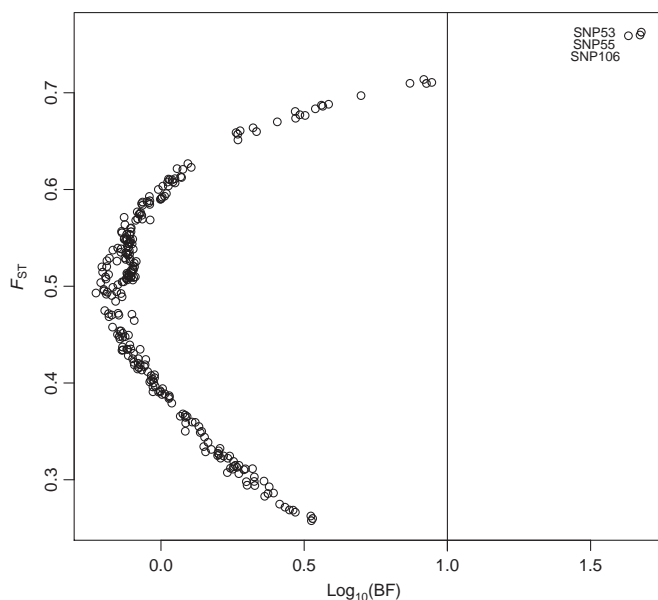

Fig. 2 Log-transformed Bayes factors and locus-specific F_{ST} estimated from Bayescan for 268 candidate gene single nucleotide polymorphisms (SNPs), identifying potential candidates for diversifying (excess differentiation) or balancing (deficient differentiation) selection across the Sitka (*Picea sitchensis*)–white (*Picea glauca*) spruce hybrid zone. The vertical line $log_{10}(BF) = 1$ (solid) corresponds to a posterior probability of 0.91. For points beyond this line there is substantial to very strong evidence in favor of selection..

Table 3 Summary statistics for genomic clines, geographic clines and genetic differentiation (F_{ST}) across all loci

Homozygous genomic clines	Heterozygous genomic clines	N	F_{ST}	Slope (per km)	Lambda (km)
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 to 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)	73 (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)

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).

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* (At2g-46210.1), 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
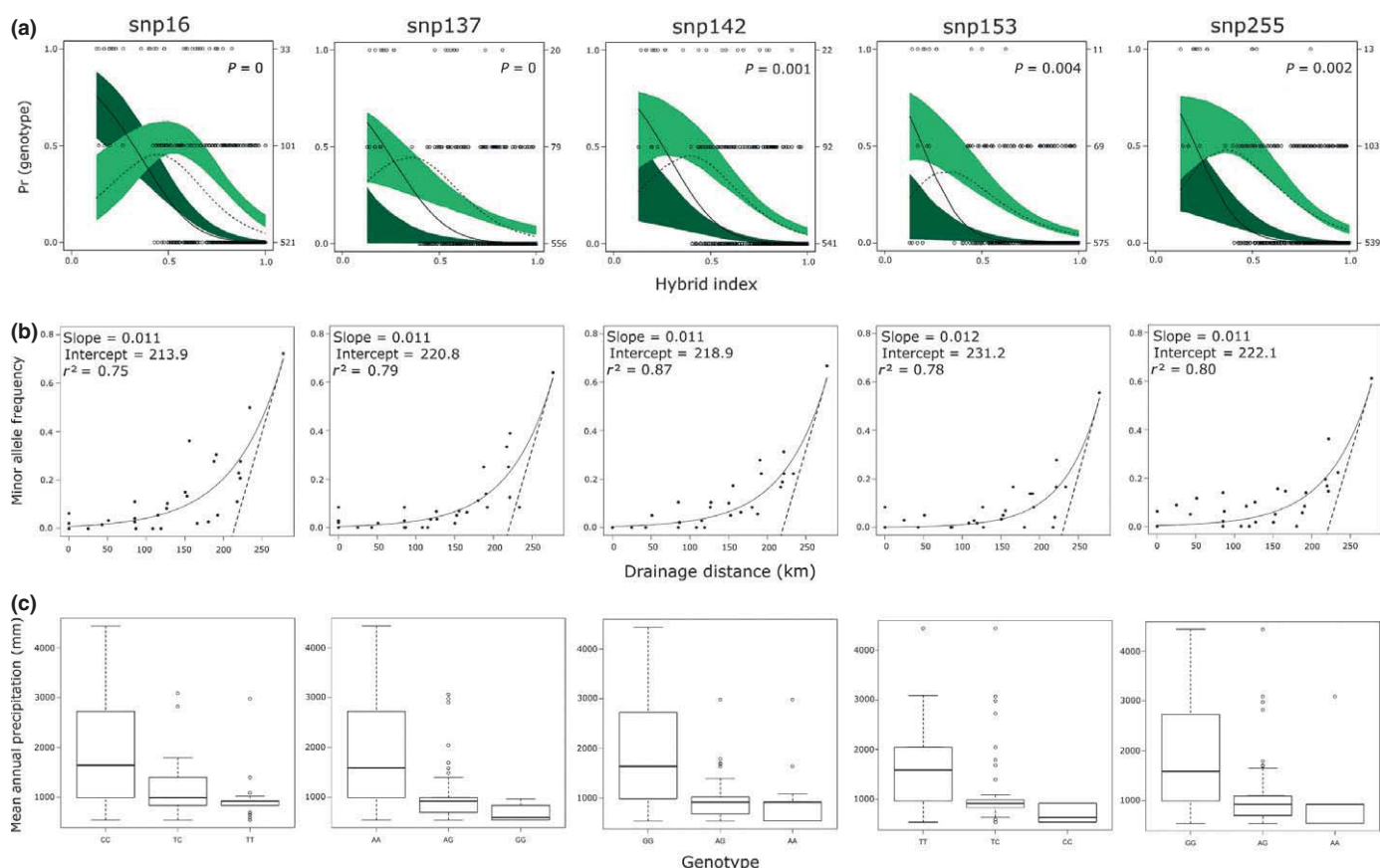
Fig. 3 Comparison of genomic clines (top), geographic clines (middle) and climatic distribution of genotypes for single nucleotide polymorphisms (SNPs) exhibiting a cline width < 100 km. (a) Genomic clines indicate locus-specific patterns of introgression using the genome-wide estimate of admixture (hybrid index, 0; white spruce (*Picea glauca*), 1; Sitka spruce (*P. sitchensis*)) 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: 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.

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− WS− SS+,

28/264) 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 F_{ST}-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 F_{ST}-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 $\alpha = 0.05$ 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–Muller 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 *Scarecrow-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 $D' = 0.89$ 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 ($0.003\,\mathrm{km}^{-1}$) 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–$0.012\ km^{-1}$) 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 (WS−), 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 (*Plasma membrane instrinsic 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 (González-Martínez *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.

References

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

Barrett RDH, Hoekstra HE. 2011. Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics* **12**: 767–780.

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**: 969–980.

Beaumont MA, Nichols RA. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society-B* **263**: 1619–1626.

Bennuah SY, Wang T, Aitken SN. 2004. Genetic analysis of the *Picea sitchensis* × *glauca* introgression zone in British Columbia. *Forest Ecology and Management* **197**: 65–77.

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.

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**: 702–717.

Buerkle C, Lexer C. 2008. Admixture as the basis for genetic mapping. *Trends in Ecology & Evolution* 23: 686–694.

Carling MD, Brumfield RT. 2008. Haldane's rule in an avian system: using cline theory and divergence population genetics to test for differential introgression of mitochondrial, autosomal, and sex-linked loci across the *Passerina* bunting hybrid zone. *Evolution* 62: 2600–2615.

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: 834–847.

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.

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.

Doyle JJ, Doyle JL. 1990. Isolation of plant DNA from fresh tissue. *Focus* 12: 13–15.

Excoffier L, Hofer T, Foll M. 2009. Detecting loci under selection in a hierarchically structured population. *Heredity* 103: 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.

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.

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: 86–89.

Gompert Z, Buerkle CA. 2009. A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology* 18: 1207–1224.

Gompert Z, Buerkle CA. 2010. INTROGRESS: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources* 10: 378–384.

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-Nucleotie Polymorphisms at candidate genes for drought-stress response in *Pinus taeda* L. *Genetics* 172: 1915–1926.

Hamilton JA, Lexer C, Aitken SN. 2012. Genomic and phenotypic architecture of a spruce hybrid zone (*Picea sitchensis* × *P. glauca*). *Molecular Ecology*. doi: 10.1111/mec.12007.

Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F, Bergelson J. 2011. Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* 334: 83–86.

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: 103–122.

Holliday JA, Ritland K, Aitken SN. 2010. Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (*Picea sitchensis*). *New Phytologist* 188: 501–514.

Kahm M, Hasenbrink G, Lichtenberg-Frate H, Ludwig J, Kschischo M. 2010. grofit: fitting biological growth curves with R. *Journal of Statistical Software* 33: 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: 2061–2075.

Ketcheson MV, Braudmandl TF, Meidinger D, Utzig G, Demarchi DA, Wikeem BM. 1991. Interior Cedar-Hemlock Zone. In: Meidinger D, Pojar J, eds. *Ecosystems of British Columbia*. Victoria, BC: Ministry of Forests, 167–181.

Le Corre V, Kremer A. 2012. The genetic differentiation of quantiative trait loci under local adaptation. *Molecular Ecology* 21: 1548–1566.

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: 74–84.

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: 1045–1057.

Lexer C, Henize 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: 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: 699–712.

Martinsen GD, Whitham TG, Turek RJ, Keim P. 2001. Hybrid populations selectively filter gene introgression between species. *Evolution* 55: 1325–1335.

Mimura M, Aitken SN. 2007. Increased selfing and decreased effective pollen donor number in peripheral relative to central populations in *Picea sitchensis* (Pinaceae). *American Journal of Botany* 94: 991–998.

Namroud M-C, 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: 3599–3613.

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.

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: 2615–2627.

Nosil P, Schluter D. 2011. The genes underlying the process of speciation. *Trends in Ecology & Evolution* 26: 160–167.

O'Neill GA, Aitken SN, King JN, Alfaro RI. 2002. Geographic variation in resin canal defenses in seedlings from the Sitka spruce × white spruce introgression zone. *Canadian Journal of Forest Research* 32: 390–400.

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* 3: 245–256.

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: 21.

Payseur BA. 2010. Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular Ecology Resources* 10: 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: 145.

Pojar J, Klinka K, Demarchi DA. 1991. Coastal Western Hemlock Zone. In: Meidinger D, Pojar J, eds. *Ecosystems of British Columbia*. Victoria, BC: Ministry of Forests, 95–111.

Porter AH, Wenger R, Geiger H, Scholl A, Shapiro AM. 1997. The *Pontia daplidice-edusa* hybrid zone in Northwestern Italy. *Evolution* 51: 1561–1573.

Porth I, Hamberger B, White R, Ritland K. 2011. Defense mechanisms against herbivory in Picea: sequence evolution and expression regulation of gene family members in the phenylpropanoid pathway. *BMC Genomics* 12: 608.

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

Ralph SG, Yueh H, Friedmann M, Aeschliman D, Zeznik JA, Nelson CC, Butterfield YSN, Kirkpatrick R, Liu J, Jones SJM *et al.* 2006. Conifer defence against insects: microarray gene expression profiling of Sitka spruce (*Picea sitchensis*) induced by mechanical wounding or feeding by spruce budworms (*Choristoneura occidentalis*) or white pine weevils (*Pissodes strobi*) reveals

large-scale changes of the host transcriptome. *Plant, Cell & Environment* **29**: 1545–1570.

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.

Richards DE, King KE, Ait-ali T, Harberd NP. 2001. How gibberellin regulates plant growth and development: a molecular genetic analysis of gibberellin signaling. *Annual Review of Plant Physiology and Plant Molecular Biology* **52**: 67–88.

Rieseberg LH, Wendel JF. 1993. *Introgression and its consequences in plants. Hybrid zones and the evolutionary process.* New York, NY, USA: Oxford University Press.

Rieseberg LH, Whitton J, Gardner KA. 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* **152**: 713–727.

Roche L. 1969. A genecological study of the genus *Picea* in British Columbia. *New Phytologist* **68**: 505–554.

Schluter D, Conte GL. 2009. Genetics and ecological speciation. *Proceedings of the National Academy of Sciences, USA* **106**: 9955–9962.

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**: 1615–1626.

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**: 70–82.

Siol M, Wright SI, Barrett SCH. 2010. The population genomics of plant adaptation. *New Phytologist* **188**: 313–332.

Steponkus PL. 1984. Role of the plasma membrane in freezing injury and cold acclimation. *Annual Review of Plant Physiology* **35**: 543–584.

Storey JD. 2002. A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B* **64**: 479–498.

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.

Streiff R, Ducousso A, Lexer C, Steinkellner H, Gloessl J, Kremer A. 1999. Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. and *Q. petraea* (Matt.) Liebl. *Molecular Ecology* **8**: 831–841.

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**: 472–485.

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.

Whitlock MC. 1992. Temporal fluctuations in demographic parameters and the genetic variance among populations. *Evolution* **46**: 608–615.

Whitlock MC. 2008. Evolutionary inference from QST. *Molecular Ecology* **17**: 1885–1896.

Wu C. 2001. The genic view of the process of speciation. *Journal of Evolutionary Biology* **14**: 851–865.

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**: 1883–1893.

Zhang Z, Ogawa M, Fleet CM, Zentella R, Hu J, Heo J, Lim J, Kamiya Y, Yamaguchi S, Sun T. 2011. SCARECROW-LIKE 3 promotes gibberellin signaling by antagonizing master growth repressor DELLA in Arabidopsis. *Proceedings of the National Academy of Sciences, USA* **108**: 2160–2165.

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