

# Divergent selection and heterogeneous migration rates across the range of Sitka spruce (*Picea sitchensis*)

Jason A. Holliday<sup>1,\*</sup>, Haktan Suren<sup>1</sup> and Sally N. Aitken<sup>2</sup>

<sup>1</sup>Department of Forest Resources and Environmental Conservation, Virginia Polytechnic Institute and State University, 304 Cheatham Hall, Blacksburg, VA 24061, USA

<sup>2</sup>Department of Forest Sciences, University of British Columbia, Vancouver, British Columbia, Canada, V6T 1Z4

Gene flow and effective population size ( $N_e$ ) should depend on a population's position within its range: those near the edges are expected to have smaller  $N_e$  and lower relative emigration rates, whereas those nearer the centre should have larger  $N_e$  and higher relative emigration rates. In species with continuous ranges, this phenomenon may limit the ability of peripheral populations to respond to divergent selection. Here, we employ Sitka spruce as a model to test these predictions. We previously genotyped 339 single nucleotide polymorphisms (SNPs) in 410 individuals from 13 populations, and used these data to identify putative targets of divergent selection, as well as to explore the extent to which central–peripheral structure may impede adaptation. Fourteen SNPs had outlier  $F_{ST}$  estimates suggestive of divergent selection, of which nine were previously associated with phenotypic variation in adaptive traits (timing of autumn budset and cold hardiness). Using coalescent simulations, we show that populations from near the centre of the range have higher effective population sizes than those from the edges, and that central populations contribute more migrants to marginal populations than the reverse. Our results suggest that while divergent selection appears to have shaped allele frequencies among populations, asymmetrical movement of alleles from the centre to the edges of the species range may affect the adaptive capacity of peripheral populations. In southern peripheral populations, the movement of cold-adapted alleles from the north represents a significant impediment to adaptation under climate change, while in the north, movement of warm-adapted alleles from the south may enhance adaptation.

**Keywords:** abundant centre; divergent selection; *Picea sitchensis*; migration; effective population size; adaptation

## 1. INTRODUCTION

Gene flow and selection are opposing forces in species that occupy heterogeneous environments. Whereas gene flow tends to homogenize allele frequencies among populations, divergent selection may lead to variation in the frequencies of particular alleles related to local adaptation. As the centre of a range is expected to contain the most suitable habitat for a given species, and therefore, support higher effective population sizes ( $N_e$ ), populations at the edges may be susceptible to gene flow counteracting the effects of natural selection, owing to high relative immigration rates from differentially adapted populations closer to the centre [1]. In addition, dispersal load and stochastic effects associated with the founding of new populations may further limit adaptation in range peripheries [2,3]. There has been substantial interest in recent years to separate neutral evolutionary processes, such as drift and migration that affect the entire genome, from the molecular adaptations that enable species to inhabit diverse environments [4–7]. Contemporary coalescent-based methods enable estimation of parameters such as  $N_e$  and migration rates, and  $F_{ST}$ -outlier approaches facilitate the detection of locally adaptive

polymorphisms by comparing  $F_{ST}$  values for individual single nucleotide polymorphisms (SNPs) to the genome-wide average [8,9].

Since the last glacial maxima approximately 13 000 years ago, most temperate and boreal plant species have expanded or shifted from refugia to fill large geographical ranges, exceeding 20° of latitude in extreme cases [10]. These ranges often contain strong clinal variation in climate-related traits, such as flowering time and vegetative bud phenology, which is interpreted as evidence of local adaptation [7,10,11]. For woody perennials, survival during the winter months requires cycling between periods of growth in summer and dormancy in winter, as substantial cold hardiness and active growth are incompatible. Local populations integrate environmental cues, primarily daylength and temperature, in order to correctly time these transitions [12]. In the context of climate change, maintenance of natural genetic variation related to local climatic adaptation is of increasing concern, particularly, since it has become apparent that neutral genetic markers are not good proxies for conservation of the molecular genetic determinants of quantitative traits [13].

Can specific genes be identified that contribute to this adaptive differentiation? One approach to this problem is to compare  $F_{ST}$  estimates at individual loci with the genome-wide average to identify polymorphisms with excess differentiation, a signature of divergent selection.

\*Author for correspondence (jah1@vt.edu).

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rspb.2011.1805> or via <http://rspb.royalsocietypublishing.org>.

This method was first suggested by Lewontin & Krakauer [14], and has been refined in recent years to account for the dependence of  $F_{ST}$  on heterozygosity [9] and for variation in  $N_e$  among studied populations [8]. Ideally, one would saturate the genome with markers and scan for excess differentiation, an approach that has been carried out in humans [15,16]. In ecologically important temperate and boreal conifers, which typically have very large genomes (10–30 Gb), an alternative approach is to study a suite of candidate genes with plausible functional roles in local adaptation. For example, Eveno *et al.* [17] assessed genetic differentiation across the range of maritime pine (*Pinus pinaster*) and found evidence for divergent selection in five out of 11 tested genes. More recently, studies of large cohorts of mapped SNPs in white and black spruce (*Picea glauca* and *Picea mariana*, respectively) suggest that loci relevant to local adaptation are widely distributed across the genomes of each of these species [18,19].

The match between multi-locus allele frequencies and local climate is unlikely to be optimal, in part because of adaptational lag, and in part owing to the disruptive effects of gene flow from adjacent populations that occupy different climatic niches [11]. Peripheral populations (both at the leading and lagging edge of the range) are often assumed to be of critical importance to range shifts predicted by climate-based species distribution models [20,21], but central–peripheral structure may impinge on the ability of these populations to respond to climate-related selection [1]. It is unclear whether the central–peripheral hypothesis is well supported in natural populations, and an understanding of contemporary gene flow is therefore an important component of predicting how species ranges and individual populations will respond to climate change. The conventional way to evaluate gene flow is using anonymous markers to estimate  $F_{ST}$ , but converting  $F_{ST}$  estimates into migration rates is problematic when subpopulation sizes are unequal or migration rates are not symmetric [22]. Coalescent samplers, such as the program MIGRATE [23,24], provide an alternative to summary statistics susceptible to multiple evolutionary forces [25], and allow for the joint estimation of migration rates that may be asymmetric, and effective population sizes that may vary among populations.

Sitka spruce (*Picea sitchensis* (Bong.) Carr.) is an amenable model with which to address these questions, owing to strong clinal variation in climate-related quantitative traits [26] and evidence for pronounced central–peripheral structure [27–29]. In addition to having ecological characteristics useful to molecular studies of local adaptation, one of the largest expressed sequence tag (EST) resources for a non-model plant has been developed in recent years for Sitka and related spruce species [30]. We previously employed a 21.8 k cDNA microarray to select approximately 300 local adaptation-related candidate genes [31], and subsequently identified numerous SNPs associated with adaptive traits in a subset of these genes [32]. In this paper, we use these data to identify putative targets of divergent selection, compare these results with data from the association study, and explore the role variable migration rates and effective population sizes may play in shaping adaptive divergence.

## 2. MATERIAL AND METHODS

### (a) *Plant material, candidate genes and sequencing*

We previously reported re-sequencing of local adaptation-related candidate genes, and subsequent genotyping a subset of SNPs within these genes in a large population of Sitka spruce for the purposes of association mapping [32,33]. For these studies, foliage was obtained from 6 year old Sitka spruce grown in a raised bed outdoor common garden at Vancouver, British Columbia, Canada (49° N, 123° W), from seed collected from natural populations spanning the species range [26] (table 1). For SNP discovery, 24 individuals from across the species range were sampled and re-sequenced for approximately 300 candidate genes. These candidate genes were selected on the basis of their previously characterized autumn expression profiles [31], or their putative function based on their nearest *Arabidopsis* homologue. Contigs corresponding to ESTs of interest were selected for primer design, and were subsequently sequenced in the forward and reverse direction, aligned and scored for polymorphisms [33]. Following SNP scoring, a subset of 670 SNPs with minor allele frequencies greater than 10 per cent in the re-sequencing panel were genotyped in a larger cohort of 410 individuals using an Illumina GoldenGate array [32]. An additional 98 SNPs were also included on this array for a separate study, for a total of 768. Among these, 470 loci distributed across 202 genes had high call rates (greater than 90%) and GenTrain scores (greater than 0.25; excluding the 98 mentioned earlier) [32], and 339 were polymorphic.

### (b) *Tests for selection*

High  $F_{ST}$  for a particular candidate gene or SNP can be indicative of divergent selection, and we therefore sought to characterize the patterns of population subdivision in Sitka spruce for the SNPs described earlier. We used the approach of Beaumont & Balding [8], implemented in the program BAYESFST, which employs Markov chain Monte Carlo to regress  $\log(F_{STij}/1 - F_{STij})$  onto a locus effect ( $\alpha_i$ ), a population effect ( $\beta_j$ ) and a locus-by-population interaction term ( $\gamma_{ij}$ ). This model separates locus-specific effects (mutation, selection) from population-specific effects ( $N_e$ , migration), and the interaction term accounts for selection that may be restricted to a single population. Following Beaumont & Balding [8], Gaussian priors with mean 0 and standard deviations of 1 and 0.5 were placed on  $\alpha_i$  and  $\gamma_{ij}$ , respectively, and  $\beta_j$  was assigned a prior mean of  $-2$  with standard deviation of 1.8. Only the 339 polymorphic SNPs were included in this analysis. Multiple runs were performed to ensure consistency, with 50 000 iterations of the Markov chain discarded as burn-in, followed by 500 000 iterations post-burn-in, of which 2000 uncorrelated outputs were sampled to estimate the posterior distribution for each parameter. Locus effects were considered significant if the two-tailed  $100(1 - P)\%$  posterior interval excluded the prior mean.

### (c) *Estimating migration rates and effective population sizes*

We used the program MIGRATE to estimate mutation-scaled effective population sizes ( $\theta_w = 4N_e\mu$ ) and mutation-scaled migration rates ( $M = m/\mu$ ) describing unidirectional number of immigrants received per generation from neighbouring populations [23,24]. For this analysis, we removed populations so that those remaining were approximately equally distributed across the species range, as coalescent

Table 1. Origins of study populations. (Two-letter codes are given for populations included in MIGRATE analysis and shown in figure 3. Dist. indicates distance from the geographical centre of the species range.)

population	state/province	latitude	longitude	dist.
Redwood (RW)	California (CA)	42° N	124° W	1496
Columbia River	Oregon (OR)	47° N	124° W	1047
Vancouver (VA)	British Columbia (BC)	49° N	123° W	689
Vancouver Island	British Columbia (BC)	49.5° N	125° W	548
Ocean Falls (OF)	British Columbia (BC)	52.5° N	128° W	203
Queen Charlotte Islands	British Columbia (BC)	53° N	132° W	5
Prince Rupert (PR)	British Columbia (BC)	53.5° N	130° W	126
Icy Bay (IB)	Alaska (AK)	60.5° N	141° W	1024
Valdez	Alaska (AK)	62° N	146° W	1341
Montague (MI)	Alaska (AK)	61° N	147° W	1369
Rocky Bay	Alaska (AK)	58° N	151° W	1590
Iniskin (IN)	Alaska (AK)	60° N	153° W	1724
Kodiak Island (KD)	Alaska (AK)	57° N	153° W	1787

samplers such as MIGRATE do not perform well in the presence of very high gene flow. The only exception to this population thinning was the inclusion of the Kodiak Island population, which was of interest owing to its geographical isolation from the contiguous mainland species range. SNPs with evidence of divergent selection on the basis of the outlier test earlier described were excluded from the MIGRATE analysis. However, to take advantage of all the information present in the data, we included the putative SNPs (based on re-sequencing) that were monomorphic on the genotyping array. This left 456 loci, which were grouped according to their linkage status within genes. To improve computational efficiency, one diploid individual per half-sibling family was chosen randomly from each population, for a total of 10 individuals, or 20 haplotypes, per population, and migration rates were only estimated between adjacent populations. The program was run three times to ensure consistent results, with 10 short chains of 50 000 steps, followed by one long chain of 500 000 steps. In each case, the first 10 000 trees were discarded as burn-in.

It should be noted that some ascertainment bias exists because the SNPs used in this analysis were chosen on the basis of being polymorphic in a range-wide discovery sample (although the genotyping array revealed that some of these were in fact monomorphic). As a result, parameter estimates from MIGRATE may be overestimated [34]. More generally, Kuhner *et al.* [34] found that maximum-likelihood coalescent approaches using SNPs tend to overestimate population parameters when diversity is low ( $\theta_w < 0.01$ ), which is the case for most gymnosperm trees [35]. However, as our purpose here was to understand relative  $N_e$  and migration rates across the range, rather than absolute values for individual populations, this bias should not significantly affect our inferences.

### 3. RESULTS

#### (a) Detection of $F_{ST}$ outlier loci

Nine SNPs within seven candidate genes had two-tailed 99 per cent posterior intervals that were positively shifted from the prior mean (corresponding to Bayesian  $p$ -values  $< 0.01$ ), and are therefore candidates for divergent selection (table 2 and figure 1). As the distribution of  $F_{ST}$  depends on heterozygosity, we plotted  $F_{ST}$  against expected heterozygosity ( $H_e$ ) for each locus, and found

that the outlier loci exhibited a range of  $H_e$  values, mostly between 0.2 and 0.5 (electronic supplementary material, figure S1). This suggests that the outlier status of these loci is not being driven by, for example, very low  $H_e$ . Among the genes, harbouring significant outliers was a phytochrome-like gene (*phyA*), an endochitinase-like gene (*chib4*), a peroxidase-like gene (*per3*) and several genes putatively involved in carbohydrate metabolism (*gh2*, *gal2* and *xth1*; table 2).  $F_{ST}$  values for these SNPs varied between 0.117 and 0.277, whereas average  $F_{ST}$  across all 339 SNPs was 0.06. Five additional SNPs reached the 95 per cent threshold, including one in each of the *xth1* and *per3* genes, as well as a putative auxin efflux carrier (*aec1*), isoflavone reductase (*ifr6*) and phosphatidylinositol kinase (*pip5k*). Among SNPs with unusually low  $F_{ST}$  suggestive of homogenizing selection, seven reached the 99 per cent threshold, and an additional 15 were significant at an  $\alpha$ -level of 0.05 (figure 1). No significant locus-by-population ( $\gamma_{ij}$ ) effects were detected.

#### (b) Effective population sizes and migration rates

The Bayesian regression method implemented in BAYESFST accounts for variable migration rates by explicitly modelling those population effects in the  $\beta_j$  term. Positive shifts from the prior mean suggest that a particular population has either a smaller  $N_e$  or restricted migration (or both), whereas negative shifts imply higher migration or a larger  $N_e$  [17]. The abundant centre model predicts that populations nearer the centre of a species range should have higher relative effective population sizes and greater connectivity, whereas populations towards the peripheries should have lower relative  $N_e$  and less connectivity. We calculated the posterior mode of  $\beta_j$  for each of the 13 populations and plotted this against distance of each population from the centre of the range, which was estimated by dividing by two the distance between the two extremes of the range at Fort Bragg, CA, USA in the south, and Kodiak Island, AK, USA in the north. Note that although Kodiak island is further south than adjacent populations on the mainland of Alaska, it is the farthest from the southern limit of the Sitka spruce range in terms of gene flow owing to the orientation of the coastline in this area, and is the most recently colonized population [36]. The Fort Bragg population is the southernmost

Table 2. Significant SNP outliers detected using BAYESFST (two-tailed,  $p$ -value  $< 0.05$ ). ('Position' and 'type' indicated location within amplicon (bp) and substitution type (synonymous; S or non-synonymous; NS), respectively. Significant phenotypic associations with either autumn cold hardiness or budset timing [22] are indicated.)

locus	BLASTX versus <i>Arabidopsis</i>	position	type	$F_{ST}$	$p$	phenotypic association?
<i>gal2</i>	galactosyltransferase	316	S	0.277	0.0005	no
<i>per3</i>	peroxidase	352	NS	0.186	0.0010	yes
<i>xth1</i>	xyloglucan:xyloglucosyl transferase	289	S	0.185	0.0005	yes
<i>xth1</i>	xyloglucan:xyloglucosyl transferase	350	S	0.169	0.0005	yes
<i>chib4</i>	basic endochitinase	189	S	0.158	0.0010	no
<i>phyA</i>	phytochrome A	441	S	0.157	0.0020	yes
<i>hta3</i>	histone	523	S	0.155	0.0005	yes
<i>per3</i>	peroxidase	570	NS	0.153	0.0030	yes
<i>xth1</i>	xyloglucan:xyloglucosyl transferase	39	S	0.136	0.0340	yes
<i>ifr6</i>	isoflavone reductase	358	NS	0.128	0.0350	no
<i>gh2</i>	glycosyl hydrolase	545	S	0.117	0.0020	no
<i>aec1</i>	auxin efflux carrier	36	S	0.111	0.0350	no
<i>per3</i>	peroxidase	619	NS	0.107	0.0130	yes
<i>pip5k</i>	phosphatidylinositol kinase	256	S	0.095	0.0430	yes

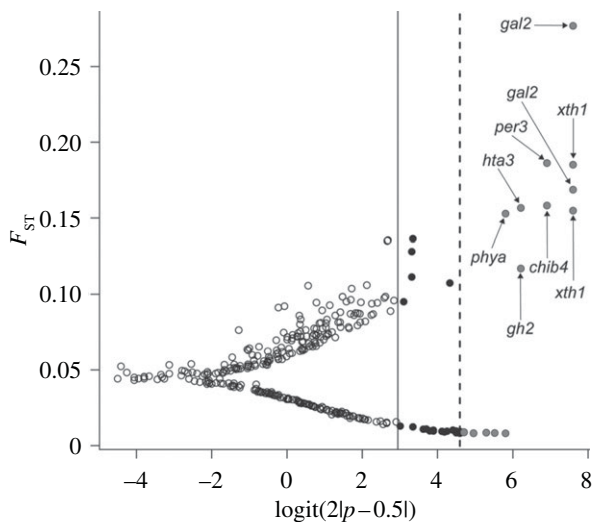


Figure 1. Results of Bayesian outlier detection.  $F_{ST}$  is plotted against transformed  $p$ -values such that SNPs exhibiting evidence of divergent selection (high  $F_{ST}$ ) and balancing selection (low  $F_{ST}$ ) both fall on the right side of the plot. The dashed and solid vertical lines indicate the two-tailed 5 and 1% Bayesian probability levels, respectively. Hence, points highlighted in filled black were significant at the 5% level, while those highlighted in filled grey were significant at the 1% level. Abbreviations are given for genes with SNPs exhibiting evidence of divergent selection ( $p < 0.01$ ).

stand of Sitka spruce, but we did not have a sufficient number of samples from this population to include it in the study. We found a strong and significant relationship between the distance from the centre of the species range and the log-transformed posterior mode of  $\beta_j$  ( $r^2 = 0.82$ ,  $p < 0.001$ ; figure 2), which suggests that there may be differences in  $N_e$  or migration rates related to a population's position in the range.

To further evaluate the fit of the abundant centre model to the case of Sitka spruce, we used the program MIGRATE to estimate effective population sizes and bidirectional migration rates. This approach revealed that scaled  $N_e$  in Sitka spruce is related to proximity to the centre of the species range ( $r^2 = 0.43$ ,  $p = 0.04$ ; electronic supplementary material, figure S2a). The largest

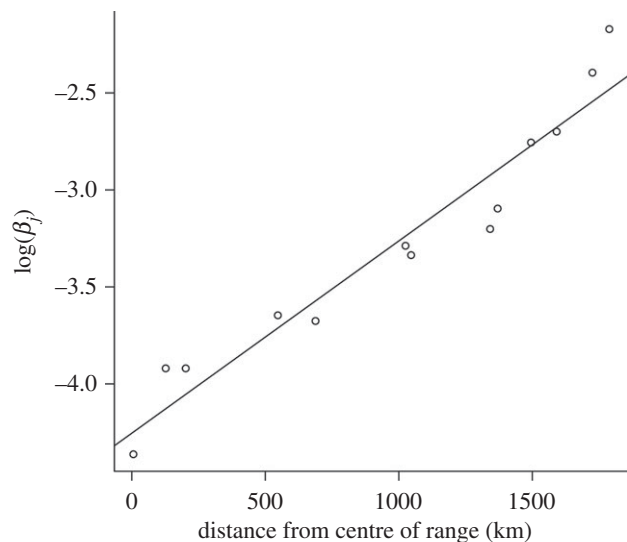


Figure 2. Relationship between log-transformed posterior mode for the population parameters ( $\beta_j$ ) from BAYESFST and the distance of each population from the geographical centre of the range of Sitka spruce ( $r^2 = 0.92$ ,  $p < 0.001$ ).

estimated scaled  $N_e$  value was for the Ocean Falls population, which is just south of the range centre at Prince Rupert, whereas the smallest population was Kodiak Island, which is the farthest population from the centre of the range (figure 3 and electronic supplementary material, table S1). Scaled  $N_e$  for Kodiak Island was 25 per cent smaller than the adjacent mainland population at Iniskin, AK, USA and the Ocean Falls population was about double that of Kodiak. Scaled  $N_e$  decreased in both directions from Ocean Falls, although not to the same extent towards the south as towards the north. In spite of being roughly equidistant from the range centre, the northern California population (Redwood) had a larger estimated scaled  $N_e$  than mainland Alaska populations by approximately 10 per cent.

The ratio of emigration to immigration was higher for the central populations than those closer to the edges of the range ( $r^2 = 0.58$ ,  $p = 0.02$ ; electronic supplementary material, figure S2b). The highest estimated migration rate was from Ocean Falls to Vancouver, BC, Canada,

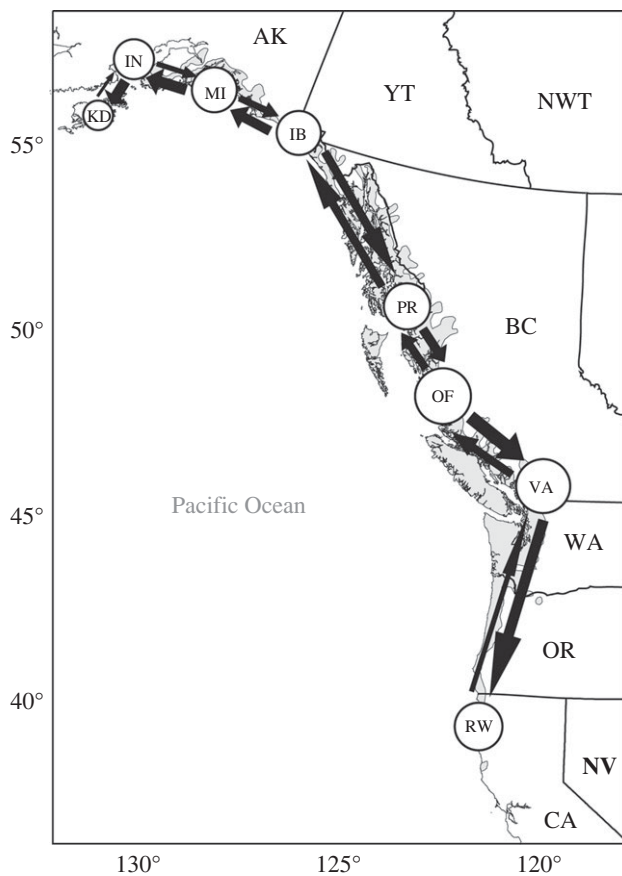


Figure 3. Results of MIGRATE for a subset of the sampled populations. Circle diameter indicates scaled effective population size ( $\theta_w = 4N_e\mu$ ; standardized to the population with the largest  $N_e$ —Ocean Falls (OF)) and arrow width indicates relative migration rates ( $M = m/\mu$ ; standardized to the largest migration rate—from Ocean Falls to Vancouver (VA)). A key to population abbreviations can be found in table 1, and the estimated values for  $4N_e\mu$  and  $m/\mu$  are given in the electronic supplementary material, table S1.

while the lowest was from Kodiak Island to Iniskin, AK, USA (figure 3 and electronic supplementary material, table S1). Although Ocean Falls had the largest scaled  $N_e$  estimate, the inflection point for migration rates was at Icy Bay. That is, migration rates were higher from north to south (i.e. from the centre to the margin) than vice versa for all populations south of Icy Bay. From Icy Bay northwards, migration rates were higher from south to north (again, from the centre to the margin) than vice versa. It should be noted, however, that although the estimated migration rate was slightly higher from Icy Bay to Prince Rupert, BC, the value for the reverse migration rate was very similar and the confidence limits on these estimates overlapped (electronic supplementary material, table S1).

#### 4. DISCUSSION

##### (a) $F_{ST}$ outliers

High population differentiation for a particular polymorphism relative to the genome-wide average provides evidence for divergent selection on that locus. Simulation studies have revealed that the distribution of  $F_{ST}$  is not as susceptible to demographic history and patterns of migration (e.g. island versus stepping-stone models) as

originally thought [9], particularly as the number of sampled populations increases [37]. Potential distortions owing to variation in migration rates or  $N_e$  among sampled populations can be accounted for by explicitly modelling these effects, which is the approach taken in BAYESFST. Previous studies of population structure in Sitka spruce suggest moderate neutral differentiation, mostly driven by isolated peripheral populations. Mimura & Aitken [26] used five microsatellite markers to estimate a mean  $F_{ST}$  of 0.11 for a subset of the populations included in this study. Although most pairwise  $F_{ST}$  estimates were less than 0.1, contrasts involving the Kodiak Island population were higher, which reflects its recent establishment and concomitant founder effects, as well as local inbreeding [26–28]. We identified 14 SNPs within 11 candidate genes that have unusually high  $F_{ST}$  estimates suggestive of divergent selection ( $p < 0.05$ ). Of these 14 SNPs, nine were previously shown to be associated with autumn budset timing, cold hardiness, or both [32]. This relationship provides additional evidence for the adaptive role of these SNPs. On the other hand, five of the putatively adaptive SNPs reported here did not show significant phenotypic associations. This may be because the selection pressure operating on these SNPs is not related to the measured traits, or because in adjusting for the confounding effects of population structure in the association study (reducing false positives), we lost the statistical power to detect the relationships (increasing false negatives). The latter situation is a particular problem in Sitka spruce, for which population differentiation for climate-related traits covaries with population structure along the essentially one-dimensional (coastal) range of the species.

No significant locus-by-population ( $\gamma_{ij}$ ) effects were detected, which may be owing to our relatively small within-population sample sizes, or because this model term accounts for divergent selection-specific to individual populations [8]. As Sitka spruce is continuously distributed throughout most of its range, we would not expect such geographically restricted selection pressures as our candidate genes were chosen owing to their putative involvement in abiotic stress tolerance, for which the relevant selection pressures—e.g. temperature and precipitation—are also continuously distributed. Although climatic thresholds (e.g. temperature above or below some physiologically relevant limit) may be important to adaptation, it is unlikely that such a threshold uniquely applies to any particular population in this study.

The putative functions of genes with  $F_{ST}$  outliers suggest their possible roles in adaptation to climatic variation across the range of Sitka spruce. The strongest differentiation was for a putative galactosyltransferase ( $F_{ST} = 0.277$ ), which falls within a gene family involved in the synthesis of disaccharides, oligosaccharides and polysaccharides. Among the multi-faceted response of plants to freezing stress, carbohydrate remodelling has been well characterized. Typically, an increase in di- and oligosaccharides is observed in conjunction with starch breakdown, which is thought to facilitate water deficit resistance, among other possible roles [38]. Two candidate genes had annotations suggesting their involvement in response to oxidative stress—a peroxidase and isoflavone reductase. This result is intriguing as radical

oxygen species present one of the primary stresses associated with freezing temperatures [39]. Although genes involved in mitigating oxidative stress are often observed to be upregulated during cold acclimation [31], their role in freezing tolerance is not well understood. For the peroxidase (*per3*), three outlier SNPs were observed, which may be the result of hitchhiking of polymorphisms adjacent to the locally adaptive SNP.

Apoplasmic antifreeze proteins, which are homologous to pathogenesis-related genes, are well-studied cold hardiness-related gene products. These secreted proteins adhere to the surface of ice, preventing its propagation and the concomitant dehydrative and mechanical stress [40]. One such candidate antifreeze gene, a basic endochitinase (*chib4*), contained an outlier SNP, and although we cannot rule out the possibility that this gene may be involved in local adaptation to fungal pathogens, it is upregulated (sevenfold between August and December) during the cold acclimation period in Sitka spruce [31], which supports its role in climatic adaptation.

Finally, perhaps the most expected outlier SNP was within a gene similar to PHYTOCHROME A (*phya*). Phytochromes are photoreceptors involved in night-length-mediated growth cessation in perennials [41], and several lines of population genetic evidence suggest the role of phytochromes in adaptation along a latitudinal cline in European aspen (*Populus tremula*) [42,43]. Our results suggest that the putative role of phytochromes in local adaptation in angiosperm trees may be conserved in conifers, as this SNP was also significantly associated with autumn budset in the association study.

#### (b) *An abundant range centre in Sitka spruce*

While several studies have investigated the abundant centre hypothesis using species abundance data, few have examined effective population size and gene flow. Results from the former class of studies have been mixed, possibly owing to inadequate sampling of species margins [44]. Among studies that employed molecular markers, most have focused on allele diversity at anonymous loci, and many have found evidence of a decline in diversity towards the species margins, though this is by no means a universal result [45–49]. Eckert *et al.* [50] conducted a meta-analysis in an effort to determine the extent to which the literature supports the central–peripheral hypothesis. Among 134 studies that explicitly set out to test whether genetic diversity declines towards the range margin, 64.2 per cent found support for this hypothesis. Many of these studies focused only on the northern range boundary of a Northern Hemisphere species, making it difficult to disentangle whether reduced diversity at the range margin was owing to colonization history or central–peripheral structure. However, there was no difference in the frequency of support for the central–peripheral hypothesis between studies that focused only on the northern range limit and those that considered more than one range boundary, suggesting that observed declines in diversity towards range margins may be at least, in part, owing to central–peripheral structure where they were found.

Though no range-wide demographic surveys have been carried out in Sitka spruce, previous studies using molecular markers suggest strong central–peripheral population

structure. Gapare *et al.* [28] and Gapare & Aitken [27] found higher inbreeding, and greater within-population spatial genetic structure in peripheral populations, respectively, and Mimura & Aitken [29] found greater selfing and biparental inbreeding, and decreased number of effective pollen donors in peripheral populations relative to central populations. In addition, Mimura & Aitken [26] found lower diversity for microsatellite markers in peripheral populations, particularly in those that were isolated from the continuous portion of the range.

We took a complementary approach to more fully characterize central–peripheral structure in this species. First, we estimated the posterior mode for the population parameter from BAYESFST ( $\beta_j$ ), and found a strong and significant relationship with the distance of each population from the centre of the species range. This suggests that populations further from the centre have smaller effective population sizes or reduced gene flow [8,17]. To explore this relationship more fully, we used coalescent simulations to estimate scaled  $N_e$  and bidirectional migration rates for a subset of the populations. This analysis revealed that populations near the centre of the range had larger  $N_e$  than peripheral populations. Interestingly, the populations with the largest relative effective sizes, Ocean Falls and Vancouver, BC, Canada, were slightly south of the geographical centre of the range at Prince Rupert, BC, Canada. The size of the Prince Rupert population was roughly equivalent to the most southerly population (Redwood, CA, USA). This result may reflect the fact that the Redwood population is relatively close to a hypothesized refugia in coastal northern California [17], and would therefore have not been affected by founder effects associated with post-glacial migration. Indeed, we recently applied the Approximate Bayesian Computation framework to assess the demographic history of these populations, and found that those in the south, i.e. Redwood, CA, USA through Vancouver, BC, Canada, were much closer to equilibrium than those north of Vancouver [33]. Nevertheless, the present-day climate along the California coast is marginal habitat for Sitka spruce—the species can only survive within a few kilometres of the ocean where persistent fog banks provide moisture and a cooler climate in the absence of precipitation during the summer months. It is therefore not surprising that the Redwood population had a smaller  $N_e$  than the more recently founded populations in southern British Columbia, which have a wider eastwest distribution of habitat owing to the presence of Vancouver Island as well as the mainland. In addition, given the long generation times of northern tree species, it is possible that our  $N_e$  estimates also represent somewhat non-equilibrium values, and that after many more years, drift in the contracted population in the south will further reduce  $N_e$ .

Bidirectional migration rates estimated using MIGRATE also support an abundant range centre in Sitka spruce. Migration rates were generally higher from the centre to the range peripheries, with the highest rate being from the largest population at Ocean Falls, to its southern neighbour at Vancouver. Conversely, the lowest migration rate was from the smallest population, Kodiak Island, AK, USA to its mainland neighbour at Iniskin, AK, USA. Interestingly, this effect was even evident for the three mainland Alaska populations, where populations closer to the centre of the range contributed more

migrants to their more peripheral neighbours than vice versa. One possible confounding factor in our interpretation of these results is the effects of post-glacial migration. However, if the estimated migration rates were the effect of successive colonization of more northern populations by those in the south following the retreat of the ice sheet, we would expect that migration rates would be uniformly higher from populations in the south to their respective neighbours in the north. That this is not the case lends support to our interpretation that our results are owing to an abundant centre in this species.

## 5. CONCLUSIONS

Our results suggest that divergent selection for SNPs relevant to local climatic adaptation has led to among-population divergence in allele frequencies across the range of Sitka spruce. This conclusion is supported by the phenotypic associations we previously described for the majority of these SNPs [32]. Future work to place these SNPs onto a linkage map for Sitka spruce will enhance our understanding of the genomic architecture of local adaptation. Namroud *et al.* [18] and Prunier *et al.* [19] found that outlier loci related to local adaptation in white and black spruce, respectively, were randomly spread across the genome. Understanding the generality of this phenomenon in outcrossing species with high gene flow, such as Sitka and white spruce, will enhance our ability to predict the likelihood of adaptation under climate change. If adaptive loci were instead clustered in the genome, the speed of adaptation would be expected to increase. However, if the results of Namroud *et al.* are the norm, we would expect the rate of adaptation to a changing climate to be substantially slower.

Although we found evidence for divergent selection across the climatically heterogeneous range of Sitka spruce, our results demonstrate a pronounced relationship between position in the range and both effective population size and the ratio of emigration to immigration in Sitka spruce. This relationship may in part be owing to the wide latitudinal but narrow longitudinal nature of the species' range. As the climate warms, the adaptive capacity of local populations may be affected by variation in  $N_e$  related to position in the range, and concomitant differences in bidirectional migration. For northern populations, the movement of pre-adapted alleles from the south may facilitate adaptation to warmer conditions, but the movement of these alleles may inhibit colonization of newly available but still marginal habitat further north. The effects of these asymmetrical migration rates can be expected to be most severe for lagging-edge populations in the south, where adaptation to warmer temperatures may be hampered by the influx of more cold-adapted alleles from the north. One possible exception to this scenario is in isolated peripheral populations, for which geographical barriers may reduce gene flow and thereby enhance adaptation. Kodiak Island is the only population in this study with this property, but we found that immigration from the adjacent mainland to be similar in magnitude to that of other population pairs. However, it is possible that this result partly reflects founder effects, as Sitka spruce reached Kodiak Island only about

400 years ago [36]. It should be noted that the abundant centre model developed by Kirkpatrick & Barton [1] was found to be particularly sensitive to changes in its parameters—for example, a change in environment. The pace of anthropogenic climate change may therefore provide the impetus for more rapid adaptation in spite of the asymmetrical effects of gene flow across the species range.

We would like to thank Dr Carol Ritland, Dylan Thomas, Michelle Tang and Leyla Tabanfar for assistance with sequencing, Dr Makiko Mimura for establishing the Sitka spruce population collection and common garden, and Dr Pia Smets and Joanne Tuytel for maintenance of the common garden. We would also like to thank two anonymous reviewers for helpful comments. This work was supported by Genome British Columbia, Genome Canada and the Province of British Columbia (grant to S.N.A.), by the Natural Science and Engineering Research Council of Canada (NSERC; grant to S.N.A.), by an NSERC Post-graduate Scholarship to J.A.H., and by Virginia Tech startup funds to J.A.H.

## REFERENCES

- Kirkpatrick, M. & Barton, N. H. 1997 Evolution of a species' range. *Am. Nat.* **150**, 1–23. (doi:10.1086/286054)
- Bridle, J. R., Polechova, J., Kawata, M. & Butlin, R. K. 2010 Why is adaptation prevented at ecological margins? New insights from individual-based simulations. *Ecol. Lett.* **13**, 485–494. (doi:10.1111/j.1461-0248.2010.01442.x)
- Alleaume-Benharira, M., Pen, I. R. & Ronce, O. 2006 Geographical patterns of adaptation within a species' range: interactions between drift and gene flow. *J. Evol. Biol.* **19**, 203–215. (doi:10.1111/j.1420-9101.2005.00976.x)
- Siol, M., Wright, S. I. & Barrett, S. C. H. 2010 The population genomics of plant adaptation. *New Phytol.* **188**, 313–332. (doi:10.1111/j.1469-8137.2010.03401.x)
- Neale, D. B. & Ingvarsson, P. K. 2008 Population, quantitative and comparative genomics of adaptation in forest trees. *Curr. Opin. Plant Biol.* **11**, 149–155. (doi:10.1016/j.pbi.2007.12.004)
- Barrett, R. D. H. & Schluter, D. 2008 Adaptation from standing genetic variation. *Trends Ecol. Evol.* **23**, 38–44. (doi:10.1016/j.tree.2007.09.008)
- Savolainen, O., Pyhajarvi, T. & Knurr, T. 2007 Gene flow and local adaptation in trees. *Annu. Rev. Ecol. Syst.* **38**, 595–619. (doi:10.1146/annurev.ecolsys.38.091206.095646)
- Beaumont, M. A. & Balding, D. J. 2004 Identifying adaptive genetic divergence among populations from genome scans. *Mol. Ecol.* **13**, 969–980. (doi:10.1111/j.1365-294X.2004.02125.x)
- Beaumont, M. A. & Nichols, R. A. 1996 Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc. Lond. B* **263**, 1619–1626. (doi:10.1098/rspb.1996.0237)
- Morgenstern, E. K. 1996 *Geographic variation in forest trees: genetic basis and application of knowledge in silviculture*. Vancouver, Canada: University of British Columbia Press.
- Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T. & Curtis-McLane, S. 2008 Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evol. Appl.* **1**, 95–111. (doi:10.1111/j.1752-4571.2007.00013.x)
- Weiser, C. J. 1970 Cold resistance and injury in woody plants. *Science* **169**, 1269–1278. (doi:10.1126/science.169.3952.1269)
- Reed, D. H. & Frankham, R. 2001 How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution* **55**, 1095–1103.

- 14 Lewontin, R. C. & Krakauer, J. 1973 Distribution of gene frequency as a test of theory of selective neutrality of polymorphisms. *Genetics* **74**, 175–195.
- 15 Akey, J. M., Zhang, G., Zhang, K., Jin, L. & Shriver, M. D. 2002 Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.* **12**, 1805–1814. (doi:10.1101/gr.631202)
- 16 Oleksyk, T. K., Zhao, K., De La Vega, F. M., Gilbert, D. A., O'Brien, S. J. & Smith, M. W. 2008 Identifying selected regions from heterozygosity and divergence using a light-coverage genomic dataset from two human populations. *PLoS ONE* **3**, e1712. (doi:10.1371/journal.pone.0001712)
- 17 Eveno, E. *et al.* 2008 Contrasting patterns of selection at *Pinus pinaster* Ait. drought stress candidate genes as revealed by genetic differentiation analyses. *Mol. Biol. Evol.* **25**, 417–437. (doi:10.1093/molbev/msm272)
- 18 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. *Mol. Ecol.* **17**, 3599–3613. (doi:10.1111/j.1365-294X.2008.03840.x)
- 19 Prunier, J., Laroche, J., Beaulieu, J. & Bousquet, J. 2011 Scanning the genome for gene SNPs related to climate adaptation and estimating selection at the molecular level in boreal black spruce. *Mol. Ecol.* **20**, 1702–1716. (doi:10.1111/j.1365-294X.2011.05045.x)
- 20 Gibson, S. Y., Van der Marel, R. C. & Starzomski, B. M. 2009 Climate change and conservation of leading-edge peripheral populations. *Conserv. Biol.* **23**, 1369–1373. (doi:10.1111/j.1523-1739.2009.01375.x)
- 21 Hamann, A. & Wang, T. L. 2006 Potential effects of climate change on ecosystem and tree species distribution in British Columbia. *Ecology* **87**, 2773–2786. (doi:10.1890/0012-9658(2006)87[2773:PEOCCO]2.0.CO;2)
- 22 Whitlock, M. C. & McCauley, D. E. 1999 Indirect measures of gene flow and migration:  $F_{ST} \neq 1/(4Nm+1)$ . *Heredity* **82**, 117–125. (doi:10.1038/sj.hdy.6884960)
- 23 Beerli, P. & Felsenstein, J. 1999 Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* **152**, 763–773.
- 24 Beerli, P. & Felsenstein, J. 2001 Maximum likelihood estimation of a migration matrix and effective population sizes in  $n$  subpopulations by using a coalescent approach. *Proc. Natl Acad. Sci. USA* **98**, 4563–4568. (doi:10.1073/pnas.081068098)
- 25 Kuhner, M. K. 2009 Coalescent genealogy samplers: windows into population history. *Trends Ecol. Evol.* **24**, 86–93. (doi:10.1016/j.tree.2008.09.007)
- 26 Mimura, M. & Aitken, S. N. 2007 Adaptive gradients and isolation-by-distance with postglacial migration in *Picea sitchensis*. *Heredity* **99**, 224–232. (doi:10.1038/sj.hdy.6800987)
- 27 Gapare, W. J. & Aitken, S. N. 2005 Strong spatial genetic structure in peripheral but not core populations of Sitka spruce *Picea sitchensis* (Bong.) Carr. *Mol. Ecol.* **14**, 2659–2667. (doi:10.1111/j.1365-294X.2005.02633.x)
- 28 Gapare, W. J., Aitken, S. N. & Ritland, C. E. 2005 Genetic diversity of core and peripheral Sitka spruce (*Picea sitchensis* (Bong.) Carr) populations: implications for conservation of widespread species. *Biol. Conserv.* **123**, 113–123. (doi:10.1016/j.biocon.2004.11.002)
- 29 Mimura, M. & Aitken, S. N. 2007 Increased selfing and decreased effective pollen donor number in peripheral relative to central populations in *Picea sitchensis* (Pinaceae). *Am. J. Bot.* **94**, 991–998. (doi:10.3732/ajb.94.6.991)
- 30 Ralph, S. G. *et al.* 2008 A conifer genomics resource of 200,000 spruce (*Picea* spp.) ESTs and 6,464 high-quality, sequence-finished full-length cDNAs for Sitka spruce (*Picea sitchensis*). *BMC Genomics* **9**, 484. (doi:10.1186/1471-2164-9-484)
- 31 Holliday, J. A., Ralph, S. G., White, R., Bohlmann, J. & Aitken, S. N. 2008 Global monitoring of autumn gene expression within and among phenotypically divergent populations of Sitka spruce (*Picea sitchensis*). *New Phytol.* **178**, 103–122. (doi:10.1111/j.1469-8137.2007.02346.x)
- 32 Holliday, J. A., Ritland, K. & Aitken, S. N. 2010 Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (*Picea sitchensis*). *New Phytol.* **188**, 501–514. (doi:10.1111/j.1469-8137.2010.03380.x)
- 33 Holliday, J. A., Yuen, M., Ritland, K. & Aitken, S. N. 2010 Postglacial history of a widespread conifer produces inverse clines in selective neutrality tests. *Mol. Ecol.* **19**, 3857–3864. (doi:10.1111/j.1365-294X.2010.04767.x)
- 34 Kuhner, M. K., Beerli, P., Yamato, J. & Felsenstein, J. 2000 Usefulness of single nucleotide polymorphism data for estimating population parameters. *Genetics* **156**, 439–447.
- 35 Savolainen, O. & Pyhajarvi, T. 2007 Genomic diversity in forest trees. *Curr. Opin. Plant Biol.* **10**, 162–167. (doi:10.1016/j.pbi.2007.01.011)
- 36 Griggs, R. F. 1937 Timberlines as indicators of climatic trends. *Science* **85**, 251–255. (doi:10.1126/science.85.2202.251)
- 37 Whitlock, M. C. 2008 Evolutionary inference from Q(ST). *Mol. Ecol.* **17**, 1885–1896. (doi:10.1111/j.1365-294X.2008.03712.x)
- 38 Zwiazek, J. J., Renault, S., Croser, C., Hansen, J. & Beck, E. 2001 Biochemical and biophysical changes in relation to cold hardiness. In *Conifer cold hardiness* (eds F. J. Bigras & S. J. Colombo), pp. 165–186. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- 39 McKersie, B. D. 1991 The role of oxygen free radicals in mediating freezing and desiccation stress in plants. In *Active oxygen/oxidative stress and plant metabolism* (eds E. Pell & K. Steffen), pp. 107–118. Rockville, MD: American Society of Plant Physiologists.
- 40 Griffith, M. & Yaish, M. W. 2004 Antifreeze proteins in overwintering plants: a tale of two activities. *Trends Plant Sci.* **9**, 399–405. (doi:10.1016/j.tplants.2004.06.007)
- 41 Howe, G. T., Gardner, G., Hackett, W. P. & Furnier, G. R. 1996 Phytochrome control of short-day-induced bud set in black cottonwood. *Physiol. Plantarum* **97**, 95–103. (doi:10.1111/j.1399-3054.1996.tb00484.x)
- 42 Ingvarsson, P. K., Garcia, M. V., Hall, D., Luquez, V. & Jansson, S. 2006 Clinal variation in *phyB2*, a candidate gene for day-length-induced growth cessation and bud set, across a latitudinal gradient in European aspen (*Populus tremula*). *Genetics* **172**, 1845–1853. (doi:10.1534/genetics.105.047522)
- 43 Ingvarsson, P. K., Garcia, M. V., Luquez, V., Hall, D. & Jansson, S. 2008 Nucleotide polymorphism and phenotypic associations within and around the phytochrome B2 locus in European aspen (*Populus tremula*, Salicaceae). *Genetics* **178**, 2217–2226. (doi:10.1534/genetics.107.082354)
- 44 Sagarin, R. D. & Gaines, S. D. 2002 The 'abundant centre' distribution: to what extent is it a biogeographical rule? *Ecol. Lett.* **5**, 137–147. (doi:10.1046/j.1461-0248.2002.00297.x)
- 45 Munwes, I., Geffen, E., Roll, U., Friedmann, A., Daya, A., Tikochinski, Y. & Gafny, S. 2010 The change in genetic diversity down the core-edge gradient in the eastern spadefoot toad (*Pelobates syriacus*). *Mol. Ecol.* **19**, 2675–2689. (doi:10.1111/j.1365-294X.2010.04712.x)



- 46 Schwartz, M. K., Mills, L. S., Ortega, Y., Ruggiero, L. F. & Allendorf, F. W. 2003 Landscape location affects genetic variation of Canada lynx (*Lynx canadensis*). *Mol. Ecol.* **12**, 1807–1816. (doi:10.1046/j.1365-294X.2003.01878.x)
- 47 Yakimowski, S. B. & Eckert, C. G. 2008 Populations do not become less genetically diverse or more differentiated towards the northern limit of the geographical range in clonal *Vaccinium stamineum* (Ericaceae). *New Phytol.* **180**, 534–544. (doi:10.1111/j.1469-8137.2008.02582.x)
- 48 Yeaman, S. & Jarvis, A. 2006 Regional heterogeneity and gene flow maintain variance in a quantitative trait within populations of lodgepole pine. *Proc. R. Soc. B* **273**, 1587–1593. (doi:10.1098/rspb.2006.3498)
- 49 Jump, A. S., Woodward, F. I. & Burke, T. 2003 *Cirsium* species show disparity in patterns of genetic variation at their range-edge, despite similar patterns of reproduction and isolation. *New Phytol.* **160**, 359–370. (doi:10.1046/j.1469-8137.2003.00874.x)
- 50 Eckert, C. G., Samis, K. E. & Loughheed, S. C. 2008 Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Mol. Ecol.* **17**, 1170–1188. (doi:10.1111/j.1365-294X.2007.03659.x)