

## ORIGINAL ARTICLE

# Characterizing the physical and genetic structure of the lodgepole pine × jack pine hybrid zone: mosaic structure and differential introgression

Catherine I. Cullingham,<sup>1\*</sup> Patrick M. A. James,<sup>1,2</sup> Janice E. K. Cooke<sup>1</sup> and David W. Coltman<sup>1</sup>

<sup>1</sup> Department of Biological Sciences, University of Alberta Edmonton, Canada

<sup>2</sup> Sciences Biologique, Université de Montréal Pavillon Marie-Victorin, Montréal, QC, Canada

## Keywords

environmental modeling, hybridization, jack pine, lodgepole pine, *Pinus banksiana*, *Pinus contorta*

## \*Correspondence

Catherine I. Cullingham, Department of Biological Sciences, CW 405 – Biological Sciences Building, University of Alberta, Edmonton, Alberta, Canada T6G 2E9.  
Tel.: +1 780 492 8368;  
fax: +1 780 492 9234;  
e-mail: cathy.cullingham@ualberta.ca

Received: 24 March 2012

Accepted: 8 April 2012

doi:10.1111/j.1752-4571.2012.00266.x

## Abstract

Understanding the physical and genetic structure of hybrid zones can illuminate factors affecting their formation and stability. In north-central Alberta, lodgepole pine (*Pinus contorta* Dougl. ex Loud. var. *latifolia*) and jack pine (*Pinus banksiana* Lamb) form a complex and poorly defined hybrid zone. Better knowledge of this zone is relevant, given the recent host expansion of mountain pine beetle into jack pine. We characterized the zone by genotyping 1998 lodgepole, jack pine, and hybrids from British Columbia, Alberta, Saskatchewan, Ontario, and Minnesota at 11 microsatellites. Using Bayesian algorithms, we calculated genetic ancestry and used this to model the relationship between species occurrence and environment. In addition, we analyzed the ancestry of hybrids to calculate the genetic contribution of lodgepole and jack pine. Finally, we measured the amount of gene flow between the pure species. We found the distribution of the pine classes is explained by environmental variables, and these distributions differ from classic distribution maps. Hybrid ancestry was biased toward lodgepole pine; however, gene flow between the two species was equal. The results of this study suggest that the hybrid zone is complex and influenced by environmental constraints. As a result of this analysis, range limits should be redefined.

## Introduction

Hybridization between closely related species is relevant to spatial evolutionary dynamics (Barton and Hewitt 1985; Seehausen 2004) and also to questions of conservation and wildlife management. For example, the composition of a hybrid zone can be important for invasive species management (Hoban et al. 2009), maintaining high-quality breeding stock (Burgarella et al. 2009) and quantifying potential losses of species diversity in response to climate change (Quintela et al. 2010). The physical structure of a hybrid zone can take many forms depending on the nature of selection on hybrid genotypes. For example, a tension zone occurs when two species hybridize, but the hybrid offspring show reduced viability/fecundity not related to the environment (Key 1968). This results in a relatively narrow, linear region (e.g., Swenson 2006; Carling and Zuckerberg 2011) that is constrained by selection and immigration (Key 1968; Barton and Hewitt 1985, 1989; Barton and Gale

1993; Arnold 1997). A hybrid zone can also develop when there is an environmental selection gradient (May et al. 1975); this occurs when hybridization happens between species with different environmental adaptations. The structure of the latter class of hybrid zone can be quite complex, and environmental heterogeneity can result in a 'mosaic' structure in the hybrid zone (Harrison 1993; Arnold 1997; Bridle et al. 2001; Vines et al. 2003) in which there is often differential selection of genotypes that depends on habitat attributes (May et al. 1975). Therefore, identifying and exploring the factors that influence the physical and genomic structure of a hybrid zone can allow us to predict areas of hybridization and better understand genetic interactions between hybridizing species.

Often, morphological characteristics are not sufficient to distinguish hybrids from parental species. In recent years, molecular methods have been shown to distinguish hybrids more reliably (Thulin et al. 2006; Quintela et al. 2010; Cabria et al. 2011). Molecular information, coupled with

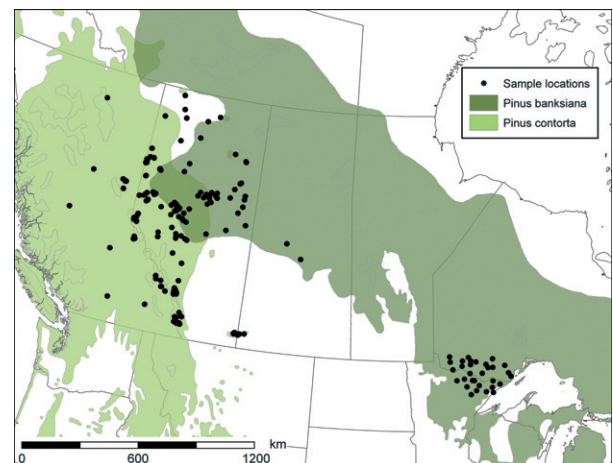
knowledge of an individual's location relative to the parents, can be used to identify associations with environmental features that can, in turn, be used to infer the type of hybrid zone. For instance, one would not expect habitat associations for a tension zone, as this zone is considered to be maintained by selection and migration independent of the environment (Harrison 1993; Arnold 1997). As well, because hybrids in a tension zone have lower survival and/or fecundity, the probability of observing advanced generation hybrids would be drastically reduced, with the consequence that most hybrids should be F1 (Jiggins and Mallet 2000; Gay et al. 2008).

Statistical modeling can be used to investigate the relationship between parents and hybrids with the environment and in doing so, shed light on the nature of the hybrid zone. Habitat association in the hybrid zone is expected when species have adapted to different environments and have subsequently come into secondary contact (Arnold 1997; Gee 2004; Carson et al. 2012). As a result, parents will tend to occur in their adapted habitats and hybrids will be found in intermediate environments largely because they are unable to compete with parents in the contrasting environments. Using climate and geographic variables, hybrid associations have been investigated for a number of diverse species. For example, Vines et al. (2003) found a strong association between aquatic habitat and allele frequencies for a hybrid zone between fire-bellied toads (*Bombina bombina* × *Bombina variegata*), and Dodd and Afzal-Rafii (2004) found that species admixture proportions were correlated with climatic variables for a western North America oak species complex (*Quercus wislizeni*, *Q. parvula*, *Q. agrifolia*, and *Q. kelloggii*). In a study of four avian hybrid zones between Pleistocene diverged species in North America, Swenson (2006) used ecological niche models to determine whether the environment played a significant role in maintaining the position of the hybrid zone among four avian taxa. In all four examples, aspect and mean annual temperature were the two most important determinants of hybrid zone structure and location. Here, we have an opportunity to investigate the environment-hybrid status relationship across a broad geographic region in a hybrid zone between two important forest species, lodgepole pine (*Pinus contorta* Dougl. ex Loud. var. *latifolia*) and jack pine (*Pinus banksiana* Lamb).

Lodgepole pine occurs throughout British Columbia, south into the United States following the Rocky Mountains and into western and central Alberta, and is found at a wide range of elevations (Carlson et al. 1999). It is considered mesophytic and occupies a broad range of soil types, including clay soils and bogs (Carlson et al. 1999; Yang et al. 1999). Jack pine is distributed throughout the boreal forest and extends from the Northwest Territories and Alberta to the east coast of Canada (Fig. 1). This

species is considered xerophytic and typically occurs on well drained, nutrient-poor soils (Kenkel et al. 1997). These two species diverged in allopatry owing to Pleistocene glaciations (Wheeler et al. 1983) and came into secondary contact in north-central Alberta approximately 6000 years ago (MacDonald and Cwynar 1985; McLeod and MacDonald 1997; MacDonald et al. 1998). Here, the two species form a hybrid zone with a complicated and poorly defined spatial structure (Wheeler and Guries 1987; Ye et al. 2002; Rweyongeza et al. 2007; Yang et al. 2007). One of the reasons that the hybrid zone has not been well-delineated is that lodgepole × jack pine hybrids can closely resemble one or the other of the pure species (Wheeler and Guries 1987; Rweyongeza et al. 2007). However, the recent development of molecular markers has made it much easier to reliably identify hybrids (Cullingham et al. 2011).

The area of lodgepole pine and jack pine species overlap in Alberta has recently become one of considerable interest. The mountain pine beetle (MPB; *Dendroctonus ponderosae* Hopkins) is a bark beetle indigenous to western North America that primarily feeds on lodgepole pine. Beetle populations are typically in an endemic phase (Raffa et al. 2008). However, when conditions allow, populations irrupt into large-scale outbreaks (Safranyik and Carroll 2006; Raffa et al. 2008). The most recent outbreak is the largest recorded in the past 125 years affecting millions of hectares of forest in western Canada and the United States (Hicke et al. 2006; Safranyik and Carroll 2006; Raffa et al. 2008; Bentz et al. 2010). In Canada, the range of MPB has recently expanded into northern British Columbia and across the Rocky Mountains into novel territory in Alberta (Robertson et al. 2009; Bentz et al. 2010; Safranyik et al. 2010) where lodge-



**Figure 1** Jack pine (*Pinus banksiana*) and lodgepole pine (*Pinus contorta*) species ranges and sample locations in western North America. Range distributions for *Pinus contorta* and *Pinus banksiana* were obtained from USGS (<http://esp.cr.usgs.gov/data/atlas/little/>, accessed 29 July 2010) and are based on Little (1971).

pole pine and jack pine overlap. There had been some uncertainty regarding the susceptibility of jack pine; however, MPB establishment in this new host was recently documented in natural stands in Alberta (Cullingham et al. 2011). While lodgepole pine is thought to share a long coevolutionary history with MBP (Kelley and Farrell 1998), jack pine is a novel host and thus is considered 'naïve'. Coevolution of lodgepole pine with the MPB suggests that lodgepole pine defenses would be adapted to protect against MPB, whereas one might speculate that this would not be the case for jack pine. Indeed, the two species exhibit notable differences in their monoterpenoid profiles, a key component of the resin defense system (Lusbrink et al. 2011). These and other potential differences between lodgepole, jack, and hybrid pines in susceptibility to MPB attack or in MPB fitness, as well as differences in species densities could result in different host-beetle dynamics with important consequences for MPB outbreak dynamics and spread. Accurate species and hybrid distribution mapping is thus an important component of improved risk assessment and prediction of MPB spread.

In this study, we build upon the aforementioned analysis by Cullingham et al. (2011) through the addition of 1320 samples genotyped at the same 11 microsatellite loci. Our objective was to delineate the area of jack–lodgepole pine hybridization in Alberta and estimate the proportion of ancestry from each species into the hybrid zone. Understanding the distribution of pure parentals and their hybrids – and the factors influencing this distribution – is important information for forest management and can be used to develop improved models of MPB spread risk. To understand the factors affecting the distribution of hybrids, we used logistic regression to assess how environmental and climate variables explain the distribution of genetic ancestry. Our specific objectives were to (i) describe the hybrid zone between lodgepole and jack pine, including patterns of ancestry and gene flow in Alberta and British Columbia, (ii) develop a species distribution model using genetic species assignment as a function of environmental and climate variables to determine the extent to which the environment predicts species distribution, and (iii) using this model, spatially delineate the current hybrid zone and compare this with the historical, morphologically based hybrid zone.

## Methods

### Sample collection and microsatellite genotyping

We used 662 of the 678 lodgepole pine, jack pine, and hybrid samples collected in British Columbia, Alberta, Saskatchewan, Ontario, and Minnesota that were previously genotyped by Cullingham et al. (2011, Dryad data doi: 10.5061/dryad.8677) for which detailed geographic information was available. Additional foliage samples were

collected from 17 locations in British Columbia and Alberta ( $N = 901$ ). The majority of the foliage samples were collected from January 2007–April 2008 from the crown using pole pruners or a shotgun. Detailed field sampling procedures can be found in the study by Cullingham et al. (2011). Additional samples were obtained from nursery stock originating from five distinct locations in British Columbia ( $N = 18$ ) and seeds from the Alberta Tree Improvement and Seed Centre seed bank ( $N = 426$ , seven from a provenance in British Columbia and the remainder from 40 locations in Alberta; Fig. 1).

Seeds were germinated to obtain seedlings for genotyping. Prior to germination, seeds were surface sterilized according to Groome et al. (1991) to remove potential fungal contamination and then stratified to improve germination rates. For stratification, seeds were placed on a moist, autoclaved Kimpad (Kimberly-Clark, Irving, TX, USA) in closed containers and stored at 4°C in the dark for 2 weeks. To germinate, seeds were transferred to fresh, autoclaved seed germination trays lined with moist Kimpad, and placed in a growth chamber (25°C; 75% humidity; 12-h dark/12-h light; 250  $\mu\text{mol m}^{-3}$  light intensity). Seeds were checked daily and watered when necessary. Seedlings were harvested when they were 2–3 cm long and stored at –20°C until DNA extraction.

Genomic DNA was isolated from ground needle and seedling tissue as described in Cullingham et al. (2011). Genotyping was completed for all individuals at 11 microsatellite loci; details of amplification and data collection can be found in Cullingham et al. (2011).

### Diversity measures

The likelihood of microsatellite scoring errors, including stutter errors, large-allele dropout, and null alleles, was estimated in MICROCHECKER (Oosterhout et al. 2004). We assessed Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) across loci separately for lodgepole and jack pine, once defined, in GENEPOP 4.0 (Raymond and Rousset 1995; web version, <http://genepop.curtin.edu.au/>, accessed 27 April 2012). Significance was assessed using Bonferroni corrected alpha-values (Rice 1989). The following estimates of allelic diversity were calculated in GenAEx 6.0 (Peakall and Smouse 2006): number of alleles, effective number of alleles, observed heterozygosity ( $H_O$ ), expected heterozygosity ( $H_E$ ), and the fixation index ( $F$ ). Again, these were calculated separately for lodgepole and jack pine once defined.

### Defining species classes

We used both NEWHYBRIDS 1.1 beta (Anderson and Thompson 2002) and STRUCTURE 2.3.1 (Pritchard et al. 2000; Falush

et al. 2003, 2007) to calculate ancestry using the protocol outlined in Cullingham et al. (2011). For both programs, we used a burn-in of 50 000 and 500 000 Markov chain Monte Carlo (MCMC) sweeps for data collection.

### Hybrid ancestry

We calculated the proportion of lodgepole and jack pine ancestry among hybrid individuals using *INTROGRESS* (Gompert and Buerkle 2010). This method calculates a hybrid index that is based on the proportion of alleles inherited from each parental population, and where alleles are shared between the parental populations, the uncertainty is included in the index (Gompert and Buerkle 2009). The parental populations were defined using jack pine in Ontario and Saskatchewan, and lodgepole pine from British Columbia to ensure no hybrids were included. Next, we randomly selected 100 individuals each of lodgepole pine, jack pine, and hybrids from Alberta, and then calculated the hybrid index based on the parental contributions and generated an ancestry plot. The microsatellite markers we used were initially optimized on lodgepole pine; therefore, to ensure that there was no ascertainment bias in estimating the ancestry of hybrids, we used the simulated hybrid classes generated by Cullingham et al. (2011) using *HYBRID-LAB* ver. 1.0 (Nielsen et al. 2006) and estimated the ancestry plots as above in *INTROGRESS* for comparison.

### Gene flow between the parental species

Gene flow between species was estimated using the software package *MIGRATE* (v3.2; Beerli 2006). *MIGRATE* uses maximum likelihood to estimate migration ( $M$ ) and effective population size ( $\Theta = 4N_E\mu$ ) without assuming equal migration or population size using a coalescent approach (Beerli and Felsenstein 2001). To estimate these two demographic parameters, we used individuals assigned as lodgepole or jack pine at  $Q > 0.95$  in Alberta and selected a random subset of individuals from the lodgepole pine data equal to the number of jack pine in the sample ( $n = 297$ ), to avoid the confounding effects of differences in sample size on the estimate of gene flow. We used maximum-likelihood estimation with 10 short chains (1000 trees used of 1 000 000 sampled) and one long chain (50 000 trees used of 5 000 000 sampled), discarding the first 100 000 trees as the initial burn-in. We ran *MIGRATE* five times with different random number seeds to verify the consistency of our results.

### Distribution modeling

We used logistic regression and molecular information to model the probability of occurrence of jack pine,

lodgepole pine, and hybrids. The output from *STRUCTURE* includes ancestry values ( $Q$ ) for each of the population clusters, where the sum of the ancestry across the population clusters for each individual sums to one. In this case, we used these continuous values for one of the population clusters as the response variable, and a set of spatial variables that included climate, moisture, and elevation (Table 1) were used as predictors. We did not include samples in the Cypress Hills area (southeast Alberta/Saskatchewan border), Saskatchewan, and Ontario because of their geographic discontinuity with the Alberta–British Columbia samples. We selected the ‘best’ model from among the full set of predictors based on the balance between the Akaike Information Criterion (AIC; Burnham and Anderson 2002) and ensuring that the variance inflation factor (VIF) for all predictors was  $<10$  (Zuur et al. 2009). The VIF is a measure of correlation among predictor variables. Model validation was undertaken through bootstrapping 60% subsets of the full data set and comparing the resulting predictions for the remaining 40% to their actual observed values. Model accuracy for each bootstrapped realization was assessed using a receiver operating characteristic (ROC) curve analogue designed for nonbinomial (i.e., continuous) response variables (Obuchowski 2005; Nguyen 2007). ROC curves are a frequently used tool for assessing model performance and sensitivity in predictive habitat models (Fielding and Bell 1997; Jiménez-Valverde 2011). ROC curves are producing by plotting the false-positive rate against the true positive rate at different thresholds based on the cells from a confusion matrix. Overall model performance can be measured by the area under the ROC curve, also known as the AUC. In effect, AUC represents the probability that the model is making the correct predictions (i.e., true occurrence and true absence). Logistic regression was performed using the *glm* function in R (Hastie and Pregibon 1992), and model performance was assessed using the *nonbinROC* package in R (R Core Development Team 2011).

To better understand the individual components of the continuous  $Q$ s-based logistic model (i.e., lodgepole pine, jack pine, and hybrids) and their contributions to overall model performance, three individual logistic regression models were built to test the accuracy of the continuous model described above for each species class independently. Here, the response variable for each model was the presence or absence of each of the three species classes. Model performances were tested using cross-validated subsets (60% training, 40% testing) and ROC analysis. Overall model performance was assessed using the AUC. ROC curves and AUC calculation for binomial data were performed using the *ROCR* package in R (Sing et al. 2009).

**Table 1.** Summary of predictor variables examined when developing a model to spatially predict *Q*-values. The ClimatePP model used to generate spatial climate data is available at: <http://www.ales2.ualberta.ca/RR/people/hamann/data.html>.

Predictor	Description	Source
Drought	Climate Moisture Index (CMI)	CFS; Hogg 1997;
Elevation	Elevation in meters	NASA ASTER DEM – <a href="https://wist.echo.nasa.gov/api/">https://wist.echo.nasa.gov/api/</a>
EXT_Cold	Extreme minimum temperature	ClimatePP v.3.2; Wang et al. 2006; Hamann and Wang 2005; Daly et al. 2002
DD >5	Growing degree-days (>5°C)	ClimatePP v.3.2
MAP	Mean annual precipitation	ClimatePP v.3.2
MAT	Mean annual temperature	ClimatePP v.3.2
MCMT	Mean coldest month temperature	ClimatePP v.3.2
MWMT	Mean warmest month temperature	ClimatePP v.3.2
Continentality	MWMT-MCMT	ClimatePP v.3.2
NFFD	Number of frost-free days	ClimatePP v.3.2
SHM	Summer heat/moisture index	ClimatePP v.3.2
Latitude	Northing in decimal degrees	Centroid of 10-km cell.
Longitude	Easting in decimal degrees	Centroid of 10-km cell.

Following model validation, *Q*s were predicted across Alberta and British Columbia using the continuous logistic regression model. Spatial prediction was carried out at a resolution of 10 km<sup>2</sup> and was restricted to the relevant extent of available environmental data. Thresholding of this predicted *Q*s layer was then undertaken to delineate a molecularly defined jack pine–lodgepole pine hybrid zone,

where hybrids are defined by a *Q*s >0.1 and <0.9 according to Cullingham et al. (2011).

## Results

### Diversity measures

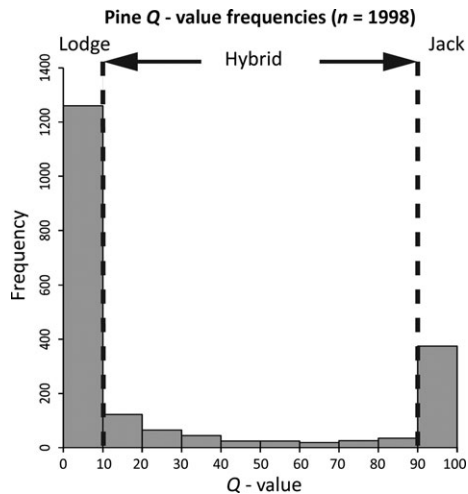
A total of 1998 pine samples were genotyped at 11 micro-satellite loci. The majority of these had complete genotypes, with a small proportion missing one locus ( $N = 123$ ) and two loci ( $N = 11$ ). Data were scored twice independently to ensure minimal scoring errors; there were a number of loci that were out of HWE with a heterozygote deficiency for both jack pine and lodgepole pine (Table 2). One locus had extreme deficiencies and was removed from all analyses. Only one locus pair for jack pine had significant LD and two for lodgepole pine. Allelic diversity measures were calculated for jack pine and lodgepole pine separately and are included in Table 2. Diversity (both number of alleles and heterozygosity) was greatest in lodgepole pine.

### Defining species classes

Species class assignments differed between STRUCTURE and NEWHYBRID for 46 individuals. In all instances, one method assigned either jack or lodgepole pine, while the other method assigned a hybrid. For the method that assigned the hybrid, the majority of the probability belonged to the pure species assigned by the first method. Of the 1998 trees that were genotyped, the final breakdown for species classes gave 386 jack pine, 1264 lodgepole pine, and 348 hybrids. The average assignment for lodgepole and jack pine was >0.98 for both assignment methods. Figure 2 illustrates the U-shaped distribution of *Q*-values (*Q*s) generated in STRUCTURE for all trees sampled.

**Table 2.** Diversity measures for 11 microsatellite loci typed in jack and lodgepole pine. Number of individuals typed at that locus (*N*), number of alleles (*N*<sub>a</sub>), observed heterozygosity (*H*<sub>o</sub>), expected heterozygosity (*H*<sub>e</sub>), and the fixation index (*F*) were calculated in GenAlix 6.4 (Peakall and Smouse 2006). Bold indicates loci out of Hardy–Weinberg equilibrium.

Locus	N	All				Jack pine				Lodgepole pine			
		<i>N</i> <sub>a</sub>	<i>H</i> <sub>o</sub>	<i>U</i> <i>H</i> <sub>e</sub>	<i>F</i>	<i>N</i> <sub>a</sub>	<i>H</i> <sub>o</sub>	<i>U</i> <i>H</i> <sub>e</sub>	<i>F</i>	<i>N</i> <sub>a</sub>	<i>H</i> <sub>o</sub>	<i>U</i> <i>H</i> <sub>e</sub>	<i>F</i>
PtTx2123	1996	6	0.620	0.735	0.156	4	0.528	0.556	0.048	6	0.620	0.629	0.013
PtTx3030	1986	25	0.618	0.847	0.270	16	0.361	0.478	<b>0.244</b>	21	0.683	0.808	<b>0.154</b>
PtTx3127	1982	13	0.519	0.667	0.223	6	0.090	0.117	<b>0.229</b>	13	0.636	0.692	<b>0.081</b>
PtTx3011	1992	55	0.810	0.963	0.159	34	0.803	0.918	<b>0.124</b>	49	0.816	0.954	<b>0.144</b>
PtTx3049	1956	22	0.721	0.906	0.204	15	0.731	0.823	<b>0.111</b>	21	0.716	0.905	<b>0.208</b>
PtTx3025	1984	23	0.749	0.804	0.068	16	0.616	0.684	0.098	23	0.778	0.764	−0.018
Pcon3	1984	34	0.778	0.930	0.163	19	0.779	0.824	0.053	33	0.770	0.930	<b>0.172</b>
PtTx2146	1978	30	0.810	0.865	0.063	14	0.758	0.762	0.003	29	0.824	0.829	0.006
PtTx4139	1993	24	0.739	0.907	0.185	17	0.689	0.694	0.006	23	0.742	0.907	<b>0.181</b>
PtTx4054	1984	23	0.749	0.892	0.160	16	0.335	0.341	0.017	23	0.845	0.892	0.052
Average				0.852	0.165			0.620	0.093			0.831	0.099



**Figure 2** Frequency distribution of  $Q$ -values generated in *STRUCTURE* for trees sampled in Alberta and British Columbia. The cutoff values used to delineate lodgepole pine and jack pine ancestry are indicated by the dashed vertical lines.

### Distribution modeling

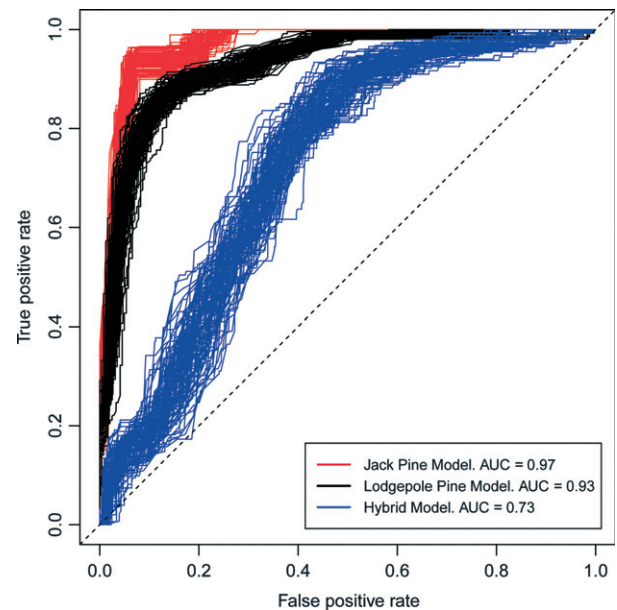
The final best fit model of species class using *STRUCTURE*-derived  $Q_s$  included elevation, drought, precipitation, a summer heat/moisture ratio, extreme cold, and location (latitude and longitude) (Table 3). Cross-validation using 60% training data and 40% testing data and summarized using a nonbinary ROC indicated reasonable overall model performance (mean continuous ROC = 0.758, SD = 0.01,  $N = 100$ ). Performance of individual species classes was assessed using traditional ROC and AUC (Fig. 3). Performance was best for jack pine and lodgepole pine, and less so for the hybrids. Spatial prediction of  $Q_s$  using this model corresponded well with the observed hybrid areas. Figure 4 shows the predicted hybrid zone is more complex, and both the hybrid zone and the distribution of pine in Alberta are more extensive than previously described by Little (1971).

### Hybrid ancestry

There was generally greater lodgepole pine ancestry among hybrids than jack pine, as jack pine only contributed  $\geq 0.5$  ancestry for approximately 10 of the 100 hybrids analyzed (Fig. 5A). In contrast, simulated data predicted equal lodgepole pine and jack pine ancestry (Fig. 5B).

### Gene flow between the parental species

Results were consistent across all five runs. The average migration estimate across runs for lodgepole into jack pine ( $M = 2.47$ , SD = 1.01) was very similar to the average



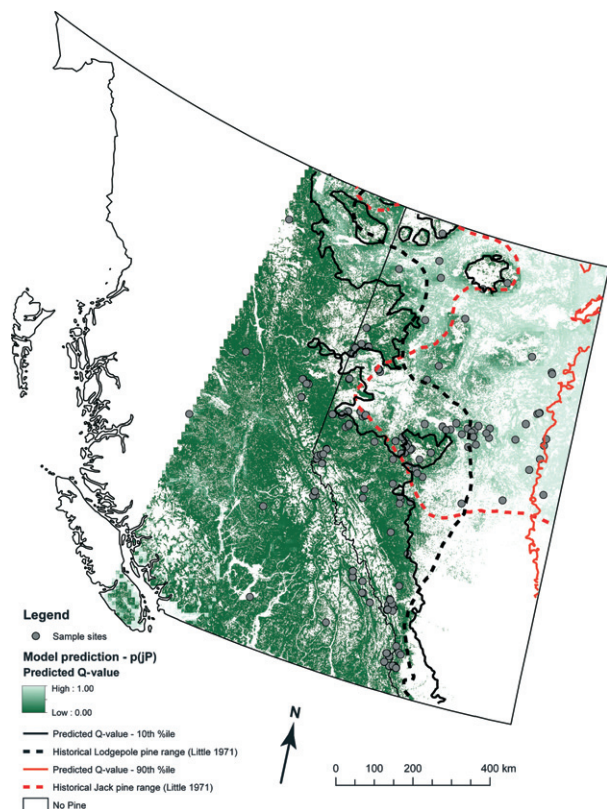
**Figure 3** Receiver operating characteristic (ROC) Curves for bootstrapped cross-validation tests of individual species logistic regression models. In these models, the response variable was a reclassified binary variable derived from the original continuous  $Q$ -values. ROC curves were plotted, and average area under the curve (AUC) was calculated for 100 cross-validation tests using 60% training and 40% testing data. A greater AUC indicates better model performance.

migration estimate from jack into lodgepole pine ( $M = 2.47$ , SD = 0.81).

### Discussion

Genetic analyses of jack and lodgepole pine in Alberta indicate that the hybrid zone is more extensive and complex than that proposed by Little (1971) and that the ranges of jack and lodgepole pine need to be redefined. The probability of genetic ancestry for these species is well predicted in Alberta by geography and habitat. Hybrids were found to occur in patchily distributed intermediate habitats that suggest a mosaic hybrid zone (Harrison 1993; Arnold 1997). Among hybrids, there is a greater proportion of lodgepole pine ancestry that could allow for the transfer of genes associated with effective defense against MPB from the coevolved lodgepole pine populations to naïve jack pine populations.

We found homozygote excess across a number of loci (Table 2), a phenomenon that has been attributed to selfing in other studies of deciduous and coniferous trees (Guries and Ledig 1982; Korshikov et al. 2007; Sutherland et al. 2010). Dancik and Yeh (1983) also reported positive values of  $F$  for lodgepole ( $F = 0.025$ ) and higher values than we found for jack pine ( $F = 0.097$ ), which they



**Figure 4** Species distribution for lodgepole and jack pine modeled by logistic regression ( $QValue \sim Elev + CMI + MAP + SHM + EXT\_Cold + Latitude + Longitude$ ; see Table 3). Color gradient represents the continuous predicted  $Q_s$  from lodgepole pine (dark green) in the west, to jack pine (light green) in the east. Solid lines represent the 10th and 90th percentiles of the predicted  $Q_s$  and indicate the revised 'pure' species boundaries according to our molecular criteria. The area between the black contours represents the revised, genetically determined pine hybrid zone. Also included are the historical eastern and western boundaries of the lodgepole and jack pine distributions, respectively (dashed lines), outlined by Little (1971) for reference. Historical distribution data were obtained from USGS (<http://esp.cr.usgs.gov/data/atlas/little/>, accessed 29 July 2010). The spatial extent of prediction was determined by the extent of the climate data model (PPClimate v3.2).

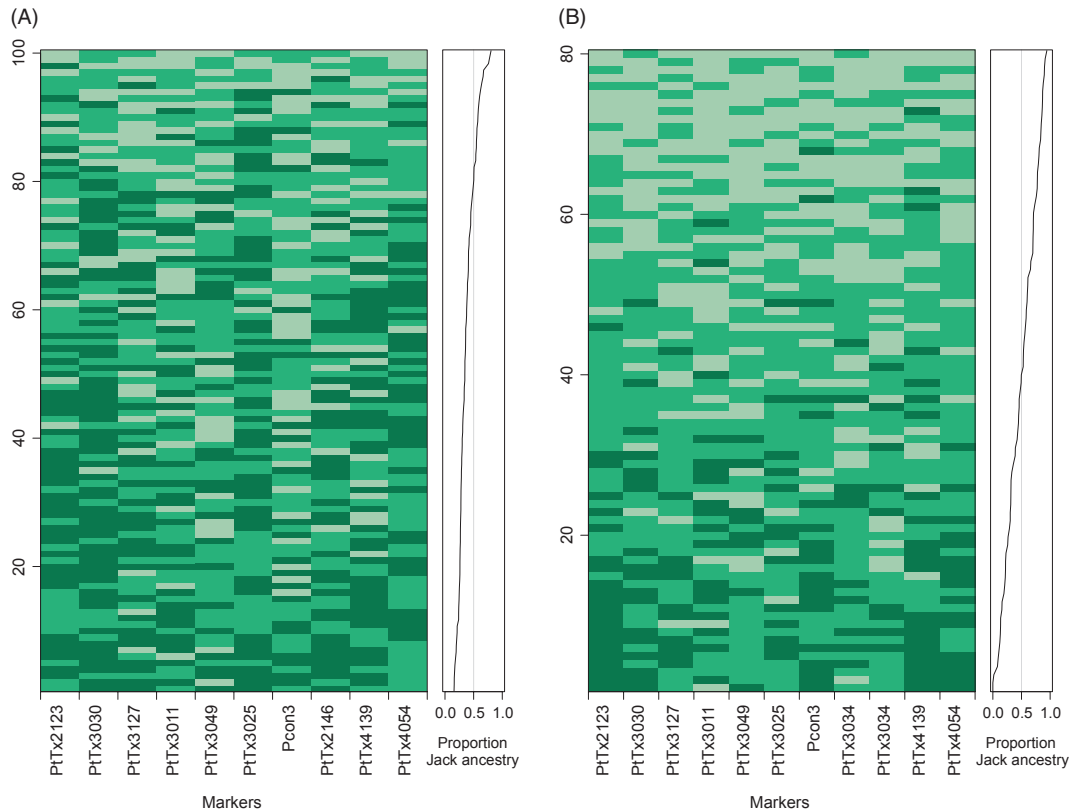
attributed to local inbreeding where seeds may tend to fall close to their maternal parent (Libby et al. 1969). However, in allozyme studies on two lodgepole pine populations, Epperson and Allard (1984, 1989) estimated outcrossing to be almost complete. Yet, their findings may be atypical and they suggested that the high density of trees (~1000/acre) contributed to the high outcrossing estimate. Therefore, a low rate of selfing may contribute to the homozygote excess observed in lodgepole pine. Our data were also collected across a very broad geographic range where some weak population substructure, perhaps owing to isolation by distance, could cause a Wahlund effect (Hedrick 2005); when

**Table 3.** Logistic model summary. Summary of chosen 'best' model based on a minimization of Akaike's Information Criterion (AIC) and variance inflation factor (VIF) factors <10. VIF refers to the variance inflation factor; a measure of correlation among predictor variables. LRT refers to a likelihood ratio test that was used to determine the significance of each predictor. All predictors listed were significant.  $p(jP)$  refers to the probability of the occurrence of jack pine based on the  $Q$ -values. The opposite directions apply to lodgepole pine.

Predictor	VIF	Coefficients	LRT	Effect on $p(jP)$
(Intercept)		51.112		
Elevation (Elev)	6.096	-0.007	724 556	-
Drought index (CMI)	2.303	0.060	329 853	+
Mean Annual Precipitation (MAP)	3.761	-0.001	10 140	-
Summer heat/moisture index (SHM)	3.671	-0.006	1217	-
Extreme min. temp. (EXT_Cold)	3.804	-0.324	111 228	+
Northing - Latitude	5.583	-0.608	340 266	-
Easting - Longitude	2.189	0.235	330 995	+

we analyzed lodgepole pine at the stand level, we did find an average reduction in  $F$  to 0.060 (data not shown). Based on this, the Wahlund effect accounts for approximately 0.04 of the bias we observed and the residual 0.06 represents the potential effects of inbreeding and genotyping error. Genotyping errors are expected to strongly affect parentage and individual identification studies in which correct genotypes are critical (Bonin et al. 2004; Pompanon et al. 2005). On the other hand, the effects of genotyping error on analyses based on allele frequencies can be minor as long as sample sizes are large. Given the large sample size we have analyzed and the fact that detection of hybrids will rely on differences in allele frequency between the parental types, a small percentage of error is not expected to significantly influence our main results. The rate of missing data for each locus (Table 2) in our study is quite low (0.1–2%). Even if all missing data were attributed to null allele homozygotes, the null allele frequency would still be quite low (<4%). Given that our estimated error rate based on analyzing 48 duplicate samples is 0.8% (Cullingham et al. 2011), any cumulative effects of error on our analysis and the main results should be negligible.

The distribution of lodgepole pine, jack pine, and their hybrids is well predicted using geographic and environmental variables (Fig. 3) and supports the hypothesis that this hybrid zone developed following secondary contact of the two tree species that evolved in allopatry (Arnold and Bennett 1993; Harrison 1993; Moore and Price 1993). In these instances, parental genotypes have the potential to outcompete hybrids in their adapted habitat; if this is true for the species under consideration, we would expect to



**Figure 5** Ancestry plots generated in *INTROGRESS* for hybrid individuals for each microsatellite locus: dark green – lodgepole homozygotes; green – lodgepole/jack pine heterozygote; light green – jack pine homozygote and for the entire individual. (A) 100 hybrid individuals genotyped in Alberta; (B) Eighty simulated hybrid individuals generated from reference lodgepole pine and jack pine.

find jack pine at low elevations and highly drained soils, and lodgepole pine to be found at higher elevations and moist soil conditions given their habitat preference (Kenkel et al. 1997; Carlson et al. 1999; Yang et al. 1999). Our statistical model bears this out, and we find co-localization of parental species with hybrids with their preferred habitats. Most importantly, the zone of hybridization is much larger yet patchier than previously described (Little 1971) and represents a significant and spatially complex region.

Species classes were accurately predicted by elevation and drought index, as has been identified in other tree species distribution models (Sykes et al. 1996). Additional significant predictors were related to soil moisture and temperature (Table 3). Extreme cold (ExtCold – summarized as a 30-year normal) was also identified as significant, whereas the average minimum temperature (MCMT) was not. This result supports recent findings that temperature extremes can be more useful in predicting species distribution than averages (Jentsch et al. 2007), especially in trees (Zimmermann et al. 2009). That spatial location (i.e., longitude and latitude) was significant predictors, even in combination with multiple climatic predictors, suggests

that additional sources of variation remain to be elucidated in the spatial distribution of lodgepole and jack pines in western Canada. As currently implemented, the geographic predictors likely capture one or several latent predictors that may relate to soil type or ecoregional classification. The basis of the habitat relationships likely resulted from the adaptation of the tree species to different environments during the Pleistocene glaciations (Godbout et al. 2005, 2008). Using this model, we found that hybrids tend to occur in transition areas between parental habitats. Using continuous habitat data, we predicted the probability of occurrence for the different species classes – including hybrids in Alberta and eastern British Columbia – and redefined the parental distributions and the hybrid zone (Fig. 4).

We were less successful at predicting hybrids in the southern region of Alberta, near the British Columbia border. Here, the model over-predicted the occurrence of lodgepole pine ( $Q_s < 0.1$ ), whereas genetic analysis identified trees in this area as hybrids (Fig. 4). One potential explanation for this over-prediction is that the coarse spatial resolution of the prediction layer (10 km<sup>2</sup>) does not



accurately capture fine-scale habitat variation. More likely, however, is that lodgepole pine individuals were misassigned to the hybrid class in this region given the large distances from jack pine distribution. Analysis of simulated data using these markers did show a percentage (<2%) of pure lodgepole pine that were misassigned as hybrids (Cullingham et al. 2011). Our data also show increased 'noise' for lodgepole pine and hybrid ancestry (Fig. 2). Indeed, the histogram of Qs (Fig. 2) indicates a definite peak for jack pine that drops sharply, whereas lodgepole pine displays a more gradual decline which makes the threshold value that differentiates lodgepole from hybrids less clear. This result is not unexpected given the greater proportion of lodgepole pine ancestry in the hybrid class (Fig. 5A). We hypothesize that model prediction could be further improved using genomic, mitochondrial, and chloroplast species discriminating SNPs which will increase resolution because the parents will not share alleles (Thompson et al. 2010; Väli et al. 2010).

Unequal ancestry within the hybrid population despite equal gene flow between the two species can be explained by several factors. First is the history of the two species. Based on pollen records, lodgepole pine colonized Alberta approximately 6000 years ago, 1000 years before jack pine (Ritchie and Yarranton 1978; MacDonald and Cwynar 1985; McLeod and MacDonald 1997; MacDonald et al. 1998). As well, the initial invading populations of jack pine were small and took a considerable amount of time to reach modern densities (MacDonald et al. 1998) and modern lodgepole pine populations tend to have higher densities than jack pine (Nealis and Peter 2008). These distributional differences could lead to increased opportunities for lodgepole pine to mate with hybrids, both historically and contemporarily, leading to greater ancestry among hybrids. Second, lodgepole pine tends to be more of a habitat generalist (Carlson et al. 1999) in comparison with jack pine. As such, hybrids with increased lodgepole pine ancestry would presumably exhibit greater habitat tolerances and be able to successfully establish in a wider range of habitats. Finally, there could be reproductive barriers. It has been shown for some tree species that hybridize there is differential gene flow because only one of the parents is able to successfully mate with hybrid offspring. For example, Floate and Whitham (1993) looked at a cottonwood (*Populus fremontii* S. Wats. × *P. angustifolia* James) hybrid zone and found F1 offspring only backcrossed with one of the parents. The same has been found for other *Populus* species (*Populus alba* × *P. tremula*, Lexer et al. 2005 and *P. deltoides* Bartr. Ex Marsh. × *P. balsamifera* L., Hamzeh et al. 2007) and to some extent for oak (*Quercus coccifera* L. × *Q. ilex* L.; Ortego and Bonal 2010). Additional analysis of the hybrid zone with mitochondrial and chloroplast species discriminating markers would allow us

to investigate this further, as these markers can help to distinguish maternal and paternal contributions.

### Implications for MPB management

The distribution of the hybrids, as well as the uncertain degree to which hybrids and jack pine compare to lodgepole pine as MPB hosts, will have important consequences for MPB population dynamics and outbreak consequences as the beetle continues to move eastward and northward. When a pest or pathogen encounters a novel host, it must be able to survive the abiotic conditions, find and infect suitable susceptible hosts, reproduce and disperse (Parker and Gilbert 2004). Although we already know that MPB can reproduce in both hybrids and jack pine (Cullingham et al. 2011), we do not know its reproductive success or other factors that impact fitness, such as the ability of larvae to survive harsh winter conditions.

Lodgepole pine has coevolved with MPB, and therefore, it is hypothesized that the defense system of this species has evolved in the presence of selective pressures imposed by MPB. In fact, there is evidence of population structure for resistance to MPB in lodgepole pine populations (Yanchuk et al. 2008). Furthermore, adaptive differences have been documented for lodgepole pine, where trees that have not been exposed previously to epidemic MPB populations support higher beetle reproductive success than do trees that are within epidemic areas (Cudmore et al. 2010). Together, these data demonstrate that MPB reproductive success can be influenced by pine genotype and suggest that in pine–MPB interactions, there is a genetic basis for the capacity of pines to affect MPB reproductive success. These differences could translate into different spread rates in jack pine and hybrid stands than in lodgepole pine, making accurate distribution maps important for predicting the impact of MPB.

Mountain pine beetle spread rates may also differ across the range of jack pine. We found evidence of gene flow between the two species, which would provide an opportunity for introgression of potentially beneficial MPB defense alleles to jack pine in Alberta and Saskatchewan given selective pressure (Zavarin et al. 1969; Rieseberg and Wendel 1993). For example, there is evidence of favorable introgression of jack pine alleles into lodgepole pine, as demonstrated by an increasing resistance to pest species (western gall rust [*Endocronartium harknessii*], stalactiform blister rust [*Cronartium coleosporioides*], needle cast [e.g., *Lophoderium seditiosum*], and sequoia pitch moth [*Synanthedon sequoia*]) in regions closer to the jack pine distribution (Wu et al. 1996). Thus, jack pine in the eastern portion of its range would be far less likely to acquire these beneficial alleles than jack pine that are proximal to lodgepole pine. From this, there could be an increased risk of spread with

increasing distance from lodgepole pine, not taking other factors such as climate suitability into account.

## Conclusion

Using microsatellite genotyping together with geographic and genetic modeling, we described the physical structure of the lodgepole × jack pine hybrid zone in Alberta and the genomic contribution from the parental species. Given the strong association of genetic ancestry with environmental variables and the range of generational hybrids, our results suggest this is a relatively stable hybrid zone that has existed for many generations. As well, the ecological adaptation of the parentals to distinct habitats should naturally maintain their ranges. Characterization of this stable tree species distribution represents useful information for forest management and quantifies one of the many uncertainties surrounding the likely continued eastward spread of the MPB, namely the likely species of host tree to be encountered. The Little (1971) distribution did not cover the full distribution of pine in Alberta (Figs 1 and 4). We determined that the hybrid zone is much more spatially complex and extensive than previously thought and also includes some regions in northern British Columbia not previously documented (Fig. 4). The redefined range maps will help managers make informed decisions and select appropriate genetic stock for reforestation. This study also raises questions for hybridization dynamics and evolutionary processes. Does the differential gene flow of lodgepole pine into the hybrids represents the movement of beneficial genes across species boundaries (Martinsen et al. 2001), or is this simply owing to the differences in tree distribution and density? A more comprehensive analysis using functional markers and species discriminating SNPs could provide more insight into the factors contributing to differential gene flow.

## Data archiving statement

Data for this study are available at Dryad. DOI: 10.5061/dryad.456q26k3.

## Acknowledgements

The authors would like to thank Sophie Dang, Matthew Bryman, Brad Jones, Darryl Edwards, Ed Hunt, and Stephane Bourassa (University of Alberta); Daniel Lux, Sunil Ranasinghe, and Tom Hutchinson (Alberta Sustainable Resources Development); Barry Cooke and Jim Weber (Canadian Forest Service, Natural Resources Canada); Michael Carlson (Government of British Columbia); Rory McIntosh and Rob Moore (Saskatchewan Ministry of the Environment); and Denys Yemshanov and Daniel

McKenney (Canadian Forest Service, Great Lakes Forestry Centre). We acknowledge funding for this research from the Government of Alberta (AAET/AFRI-859-G07), as well as grants from Genome Canada, the Government of Alberta through Genome Alberta, and Genome British Columbia in support of the Tria I and Tria II projects (<http://www.thetriaproject.ca>) of which J.E.K.C and D.W.C. are principle investigators. P.M.A.J. was additionally supported by a Killam Postdoctoral fellowship at the University of Alberta.

## Literature cited

- Anderson, E., and E. A. Thompson. 2002. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* **160**:1217–1229.
- Arnold, M. L. 1997. *Natural Hybridization and Evolution*. Oxford University Press, Oxford, UK.
- Arnold, M. L., and B. D. Bennett. 1993. Natural hybridization in Louisiana irises: genetic variation and ecological determinants. In H. G. Harrison, eds. *Hybrid zones and the Evolutionary Process*, pp. 115–139. Oxford University Press, Oxford, UK.
- Barton, N. H., and K. S. Gale. 1993. Genetic analysis of hybrid zones. In R. G. Harrison, ed. *Hybrid Zones and the Evolutionary Process*, pp. 13–45. Oxford University Press, Oxford, UK.
- Barton, N. H., and G. M. Hewitt. 1985. Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16**:113–148.
- Barton, N. H., and G. M. Hewitt. 1989. Adaptation, speciation and hybrid zones. *Nature* **341**:497–503.
- Berli, P. 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* **22**:341–345.
- Berli, P., and J. Felsenstein. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America* **98**:4563–4568.
- Bentz, B.J., J. Régnière, C.J. Fettig, E.M. Hansen, J.L. Hayes, J.A. Hicke, R.G. Kelsey et al. 2010. Climate change and bark beetles of the western United States and Canada: direct and indirect effects. *BioScience* **60**:602–613.
- Bonin, A., E. Bronken Eidesen, F. Pompanon, C. Borchmann, and P. Taberlet. 2004. How to track and assess genotyping errors in population genetic studies. *Molecular Ecology* **13**:3261–3273.
- Bridle, J. R., S. J. E. Baird, and R. K. Butlin. 2001. Spatial structure and habitat variation in a grasshopper hybrid zone. *Evolution* **55**:1832–1843.
- Burgarella, C., Z. Lorenzo, R. Jabbour-Zahad, R. Lumaret, E. Guichoux, R.J. Petit, A. Soto, and L. Gil. 2009. Detection of hybrids in nature: application to oaks (*Quercus suber* and *Q. ilex*). *Heredity* **102**:442–452.
- Burnham, K. P., and D. R. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd edn. Springer-Verlag, New York, NY.
- Cabria, M.T., J.R. Michaux, B.J. Gómez-Moliner, D. Skumatov, T. Maran, P. Fournier, J. Lopez, et al. 2011. Bayesian analysis of hybridization and introgression between the endangered European mink (*Mustela lutreola*) and the polecat (*Mustela putorius*). *Molecular Ecology* **20**:1176–1190.
- Carling, M. D., and B. Zuckerberg. 2011. Spatio-temporal changes in the genetic structure of the Passerina bunting hybrid zone. *Molecular Ecology* **20**:1166–1175.

- Carlson, M. R., J. C. Murphy, V. G. Berger, and L. F. Ryrie. 1999. Genetics of elevational adaptations of lodgepole pine in the interior. *Journal of Sustainable Forestry* **10**:35–44.
- Carson, E., M. Tobler, W. L. Minckley, R. J. Ainsworth, and T. E. Dowling. 2012. Relationships between spatio-temporal environmental and genetic variation reveal an important influence of exogenous selection in a pupfish hybrid zone. *Molecular Ecology* **21**:1209–1222.
- Cudmore, T. J., N. Björklund, A. L. Carroll, and B. S. Lindgren. 2010. Climate change and range expansion of an aggressive bark beetle: evidence of higher beetle reproduction in naïve host tree populations. *Journal of Applied Ecology* **47**:1036–1043.
- Cullingham, C. I., J. E. K. Cooke, S. Dang, C. S. Davis, B. J. Cooke, and D. W. Coltman. 2011. Mountain pine beetle host-range expansion threatens the boreal forest. *Molecular Ecology* **20**:2157–2171.
- Daly, C., W. P. Gibson, G. H. Taylor, G. L. Johnson, and P. Pasteris. 2002. A knowledge-based approach to the statistical mapping of climate. *Climate Research* **22**:99–113.
- Dancik, B. P., and F. C. Yeh. 1983. Allozyme variability and evolution of lodgepole pine (*Pinus contorta* var. *latifolia*) and jack pine (*P. banksiana*) in Alberta. *Canadian Journal of Genetics and Cytology* **25**:57–64.
- Dodd, R. S., and Z. Afzal-Rafii. 2004. Selection and dispersal in a multi-species oak hybrid zone. *Evolution* **58**:261–269.
- Epperson, B. K., and R. W. Allard. 1984. Allozyme analysis of the mating system in lodgepole pine populations. *Journal of Heredity* **75**:212–214.
- Epperson, B. K., and R. W. Allard. 1989. Spatial autocorrelation analysis of the distribution of genotypes within populations of lodgepole pine. *Genetics* **121**:369–377.
- Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**:1567–1587.
- Falush, D., M. Stephens, and J. K. Pritchard. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* **7**:574–578.
- Fielding, A. H., and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* **24**:38–49.
- Floate, K. D., and T. G. Whitham. 1993. The “hybrid bridge” hypothesis: host shifting via plant hybrid swarms. *The American Naturalist* **141**:651–662.
- Gay, L., P.-A. Crochet, D. A. Bell, and T. Lenormand. 2008. Comparing clines on molecular and phenotypic traits in hybrid zones: a window on tension zone models. *Evolution* **62**:2789–2806.
- Gee, J. M. 2004. Gene flow across a climatic barrier between hybridizing avian species, California and Gambel’s quail (*Callipepla californica* and *C. gambelli*). *Evolution* **58**:1108–1121.
- Godbout, J., J. P. Jaramillo-Correa, J. Beaulieu, and J. Bousquet. 2005. A mitochondrial DNA minisatellite reveals the postglacial history of jack pine (*Pinus banksiana*), a broad-range North American conifer. *Molecular Ecology* **14**:3497–3512.
- Godbout, J., A. Fazekas, C. H. Newton, F. C. Yeh, and J. Bousquet. 2008. Glacial vicariance in the Pacific Northwest: evidence from a lodgepole pine mitochondrial DNA minisatellite for multiple genetically distance and widely separated refugia. *Molecular Ecology* **17**:2463–2475.
- Gompert, Z., and C. A. Buerkle. 2009. A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology* **18**:1207–1224.
- Gompert, Z., and C. A. Buerkle. 2010. INTROGRESS: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources* **10**:378–384.
- Groome, M. C., S. R. Axler, and D. J. Gifford. 1991. Hydrolysis of lipid and protein reserves in loblolly pine seeds in relation to protein electrophoretic patterns following imbibition. *Physiologia Plantarum* **83**:99–106.
- Guries, R. P., and F. T. Ledig. 1982. Genetic diversity and population structure in pitch pine (*Pinus rigida* Mill). *Evolution* **36**:387–402.
- Hamann, A., and T. Wang. 2005. Models of climate normals for genecology and climate change studies in BC. *Agricultural and Forest Meteorology* **128**:211–221.
- Hamzeh, M., C. Sawchyn, P. Périnet, and S. Dayanandan. 2007. Asymmetrical natural hybridization between *Populus deltoids* and *P. balsamifera* (Salicaceae). *Canadian Journal of Botany* **85**:1227–1232.
- Harrison, R. G. 1993. Hybrids and hybrid zones: historical perspective. In R. G. Harrison ed. *Hybrid Zones and the Evolutionary Process*, pp. 3–12. Oxford University Press, Oxford, UK.
- Hastie, T. J., and D. Pregibon. 1992. Generalized linear models. In J. M. Chambers, and T. J. Hastie, eds. *Statistical Models in S*, pp. 195–247. Wadsworth & Brooks/Cole, University of Michigan, MI.
- Hedrick, P.W. 2005. *Genetics of Populations*, 3rd edn. Jones and Bartlett Publishers, Sudbury, MA.
- Hicke, J. A., J. A. Logan, J. Powell, and D. S. Ojima. 2006. Changing temperatures influence suitability for modeled mountain pine beetle (*Dendroctonus ponderosae*) outbreaks in the western United States. *Journal of Geophysical Research* **111**:G02019.
- Hoban, S. M., T. S. McCleary, S. E. Schlarbaum, and J. Romero-Severson. 2009. Geographically extensive hybridization between the forest trees American butternut and Japanese walnut. *Biology Letters* **5**:324–327.
- Hogg, E. H. 1997. Temporal scaling of moisture and the forest–grassland boundary in western Canada. *Agricultural and Forest Meteorology* **84**:115–122.
- Jentsch, A., J. Kreyling, and C. Beierkuhnlein. 2007. A new generation of climate-change experiments: events, not trends. *Frontiers in Ecology and the Environment* **5**:365–374.
- Jiggins, D. D., and J. Mallet. 2000. Bi-modal hybrid zones and speciation. *Trends in Ecology and Evolution* **15**:250–255.
- Jiménez-Valverde, A. 2011. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography* doi: 10.1111/j.1466-8238.2011.00683.x.
- Kelley, S. T., and B. D. Farrell. 1998. Is specialization a dead end? The phylogeny of host use in *Dendroctonus* Bark Beetles (Scolytidae). *Evolution* **52**:1731–1743.
- Kenkel, N. C., M. L. Hendrie, and I. E. Bella. 1997. A long-term study of *Pinus banksiana* population dynamics. *Journal of Vegetative Science* **8**:241–254.
- Key, K. H. L. 1968. The concept of stasipatric speciation. *Systematic Zoology* **17**:14–22.
- Korshikov, I. I., N. N. Pirko, E. A. Mudrik, and Y. V. Pirka. 2007. Maintenance of genetic structure in progenies of marginal mountainous and steppe populations of three species of *Pinaceae* Lindl. Family in Ukraine. *Silvae Genetica* **1**:1–10.
- Lexer, C., M. F. Fay, J. A. Joseph, M.-S. Nica, and B. Heinze. 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.
- Libby, W. J., R. F. Stettler, and F. W. Seitz. 1969. Forest genetics and forest-tree breeding. *Annual Review Genetics* **3**:469–494.

- Little Jr, E. L. 1971. Atlas of the United States Trees, Conifers and Important Hardwoods, Vol. 1. U.S. Department of Agriculture Miscellaneous Publication No. 1146.
- Lusibrink, I., M. L. Evenden, F. G. Blanchet, J. E. K. Cooke, and N. Erbilingin. 2011. Effect of water stress and plant defense stimulation on monoterpene emission from historical and new pine hosts of the mountain pine beetle. *Journal of Chemical Ecology* **37**:1013–1026.
- MacDonald, G. M., and L. C. Cwynar. 1985. A fossil pollen based reconstruction of the late Quaternary history of lodgepole pine (*Pinus contorta* ssp. *latifolia*) in the western interior of Canada. *Canadian Journal of Forest Research* **15**:1039–1044.
- MacDonald, G. M., L. C. Cwynar, and C. Whitlock. 1998. The late Quaternary dynamics of pines in northern North America. In D. M. Richardson, ed. *Ecology and Biogeography of Pinus*, pp. 122–136. Cambridge University Press, Cambridge.
- Martinsen, G. D., T. G. Whitham, R. J. Turek, and P. Keim. 2001. Hybrid populations selectively filter gene introgression between species. *Evolution* **55**:1325–1335.
- May, R. M., J. A. Endler, and R. E. McMurtrie. 1975. Gene frequency clines in the presence of selection opposed by gene flow. *The American Naturalist* **109**:659–676.
- McLeod, T. K., and G. M. MacDonald. 1997. Postglacial range expansion and population growth of *Picea marina*, *Picea glauca*, and *Pinus banksiana* in the western interior of Canada. *Journal of Biogeography* **24**:865–881.
- Moore, W. S., and J. T. Price. 1993. Nature of selection in the northern flicker hybrid zone and its implications for speciation theory. In R. G. Harrison, ed. *Hybrid zones and the Evolutionary Process*, pp. 196–225. Oxford University Press, Oxford, UK.
- Nealis, V., and B. Peter. 2008. Risk assessment of the threat of mountain pine beetle to Canada's boreal and eastern pine forests. Report BC-X-417, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC.
- Nguyen, P. 2007. nonbinROC: ROC-type Analysis for Non-Binary Gold Standards. R package Version 1.0.1. <http://cran.r-project.org/web/packages/nonbinROC/nonbinROC.pdf> (accessed on April 27 2012).
- Nielsen, E. E. G., L. Arvebach, and P. Kotlicki. 2006. HYBRIDLAB (version 1.0): a program for generating simulated hybrids from population samples. *Molecular Ecology Notes* **6**:971–973.
- Obuchowski, N. A. 2005. Estimating and comparing diagnostic tests' accuracy when the gold standard is not binary. *Academic Radiology* **12**:1198–1204.
- Oosterhout, C. V., W. F. Hutchinson, D. P. M. Wills, and P. Shipley. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**:535–538.
- Ortego, J., and R. Bonal. 2010. Natural hybridization between kermes (*Quercus coccifera* L.) and holm oaks (*Q. ilex* L.) revealed by microsatellite markers. *Plant Biology* **12**:234–238.
- Parker, I. M., and G. S. Gilbert. 2004. The evolutionary ecology of novel plant-pathogen interactions. *Annuals Reviews of Ecology, Evolution and Systematics* **35**:675–700.
- Peakall, R., and P. E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**:288–295.
- Pompanon, F., A. Bonin, E. Bellemain, and P. Taberlet. 2005. Genotyping errors: causes, consequences and solutions. *Nature Reviews, Genetics* **6**:847–859.
- Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**:945–959.
- Quintela, M., C.-G. Thulin, and J. Höglund. 2010. Detecting hybridization between willow grouse (*Lagopus lagopus*) and rock ptarmigan (*L. muta*) in central Sweden through Bayesian admixture analyses and mtDNA screening. *Conservation Genetics* **11**:557–569.
- Raffa, K.F., B.H. Aukema, J. Berntz, A.L. Carroll, J.A. Hicke, M.G. Turner, and W.H. Romme. 2008. Cross-scale drivers of natural disturbances prone to anthropogenic amplification: the dynamics of bark beetle eruptions. *BioScience* **58**:501–517.
- Raymond, M., and F. Rousset. 1995. Genepop (Version-1.2) - Population-genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**:248–249.
- Rice, W. R. 1989. Analyzing tables of statistical tests. *Evolution* **43**:223–225.
- Rieseberg, L. H., and J. F. Wendel. 1993. Introgression and its consequences in plants. In R. G. Harrison, ed. *Hybrid Zones and the Evolutionary Process*, pp. 70–109. Oxford University Press, Oxford, UK.
- Ritchie, J. C., and G. A. Yarranton. 1978. The late-Quaternary history of the boreal forest of central Canada, based on standard pollen stratigraphy and principal components analysis. *Journal of Ecology* **66**:199–212.
- Robertson, C., T. A. Nelson, D. E. Jelinski, M. A. Wulder, and B. Boots. 2009. Spatial-temporal analysis of species range expansion: the case of the mountain pine beetle, *Dendroctonus ponderosae*. *Journal of Biogeography* **36**:1446–1458.
- Rweyongeza, D. M., N. K. Dhir, L. K. Barnhardt, C. Hansen, and R.-C. Yang. 2007. Population differentiation of lodgepole pine (*Pinus contorta*) and jack pine (*Pinus banksiana*) complex in Alberta: growth, survival, and responses to climate. *Canadian Journal of Botany* **85**:545–556.
- Safranyik, L., and A. Carroll. 2006. The biology and epidemiology of the mountain pine beetle in lodgepole pine forests. In L. Safranyik, and W. Wilson, eds. *The Mountain Pine Beetle: A Synthesis of Biology, Management and Impacts on Lodgepole Pine*, pp. 3–66. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC.
- Safranyik, L., A.L. Carroll, J. Règnière, D.W. Langor, W.G. Riel, T.L. Shore, B. Peter, et al. 2010. Potential for range expansion of mountain pine beetle into the boreal forest of North America. *Canadian Entomologist* **142**:415–4442.
- Seehausen, O. 2004. Hybridization and adaptive radiation. *Trends in Ecology and Evolution* **19**:198–207.
- Sing, T., O. Sander, N. Beerwinkler, and T. Lengauer. 2009. ROCr: Visualizing the performance of scoring classifiers. R package version 1.0-4. <http://CRAN.R-project.org/package=ROCr> (accessed on May 2001).
- Sutherland, B.G., A. Belaj, S. Nier, J.E. Cottrell, S.P. Vaughan, J. Hubert, and K. Russell. 2010. Molecular biodiversity and population structure in common ash (*Fraxinus excelsior* L.) in Britain: implications for conservation. *Molecular Ecology* **19**:2196–2211.
- Swenson, N. G. 2006. GIS-based niche models reveal unifying climatic mechanisms that maintain the location of avian hybrid zones in a North American suture zone. *Journal of Evolutionary Biology*, **19**:717–725.
- Sykes, M. T., I. C. Prentice, and W. Cramer. 1996. A bioclimatic model for the potential distributions of north European tree species under present and future climates. *Journal of Biogeography* **23**:203–233.
- Thompson, S. L., M. Lamothe, P. G. Meirmans, P. Périnet, and N. Isabel. 2010. Repeated unidirectional introgression towards *Populus balsamifera* in contact zones of exotic and native poplars. *Molecular Ecology* **19**:132–145.

- Thulin, C.-G., J. Stone, H. Tegelström, and C. W. Walker. 2006. Species assignment and hybrid identification among Scandinavian hares *Lepus europaeus* and *L. timidus*. *Wildlife Biology* **12**:29–38.
- Väli, Ü., P. Saag, V. Dombrovski, B.-U. Meyburg, G. Maciorowski, T. Mizera, R. Treinys, and S. Fagerberg. 2010. Microsatellites and single nucleotide polymorphisms in avian hybrid identification: a comparative case study. *Journal of Avian Biology* **41**:34–49.
- Vines, T.H., S.C. Köhler, M. Thiel, I. Ghira, C.J. MacCallum, N.H. Barton, and B. Nurenberger. 2003. The maintenance of reproductive isolation in a mosaic hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*. *Evolution* **57**:1876–1888.
- Wang, T., A. Hamann, D. Spittlehouse, and S. N. Aitken. 2006. Development of scale-free climate data for western Canada for use in resource management. *International Journal of Climatology* **26**:383–397.
- Wheeler, N. C., and R. P. Guries. 1987. A quantitative measure of introgression between lodgepole and jack pines. *Canadian Journal of Botany* **65**:1876–1885.
- Wheeler, N. C., R. P. Guries, and D. M. O'Malley. 1983. Biosystematics of the genus *Pinus*, subsection *Contortae*. *Biochemical Systematics and Ecology* **11**:333–340.
- Wu, H. X., C. C. Ying, and J. A. Muir. 1996. Effect of geographic variation and jack pine introgression on disease and insect resistance in lodgepole pine. *Canadian Journal of Forestry Research* **26**:711–726.
- Yanchuk, A. D., J. C. Murphy, and K. F. Wallin. 2008. Evaluation of genetic variation of attack and resistance in lodgepole pine in the early stages of a mountain pine beetle outbreak. *Tree Genetics and Genomes* **4**:171–180.
- Yang, R.-C., Z. Ye, and Y. Hiratsuka. 1999. Susceptibility of *Pinus contorta* – *Pinus banksiana* complex to *Endocronartium harknessii*: host-pathogen interactions. *Canadian Journal of Botany* **77**:1035–1043.
- Yang, R.-C., F. C. Yeh, and T. Z. Ye. 2007. Multilocus structure in the *Pinus contorta* – *Pinus banksiana* complex. *Canadian Journal of Botany* **85**:774–784.
- Ye, T. Z., R.-C. Yang, and F. C. Yeh. 2002. Population structure of a lodgepole pine (*Pinus contorta*) and jack pine (*P. banksiana*) complex as revealed by random amplified polymorphic DNA. *Genome* **45**:530–540.
- Zavarin, E., W. B. Critchfield, and K. Snajberk. 1969. Turpentine composition of *Pinus contorta* × *Pinus banksiana* hybrids and hybrid derivatives. *Canadian Journal of Botany* **47**:1443–1453.
- Zimmermann, N.E., N.G. Yoccoz, T.C. Edwards, E.S. Meier, W. Thuiller, A. Guisan, D.R. Schmatz, and P.B. Pearman. 2009. Climatic extremes improve predictions of spatial patterns of tree species. *Proceedings of the National Academy of Sciences, USA* **106**:19723–19728.
- Zuur, A. F., E. N. Ieno, N. J. Walker, A. A. Saveliev, and G. Smith 2009. *Mixed Effects Models and Extensions in Ecology with R*. Springer, New York.