



Canadian Journal of Forest Research

Spatial and genetic structure of the lodgepole x jack pine hybrid zone

Journal:	<i>Canadian Journal of Forest Research</i>
Manuscript ID	cjfr-2018-0428.R1
Manuscript Type:	Article
Date Submitted by the Author:	04-Mar-2019
Complete List of Authors:	Burns, Ian; University of Alberta, Department of Biological Sciences James, Patrick; Université du Montréal, Coltman, David; University of Alberta, Department of Biological Sciences Cullingham, Catherine; University of Alberta, Department of Biological Sciences
Keyword:	jack pine, lodgepole pine, distribution modelling, mountain pine beetle, hybridization
Is the invited manuscript for consideration in a Special Issue? :	MPB TRIA-Net

SCHOLARONE™
Manuscripts

1 Spatial and genetic structure of the lodgepole x jack pine hybrid zone

2 Ian Burns¹, Patrick MA James², David W Coltman¹, Catherine I Cullingham¹

3 ¹ Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada, T6G 2E9,

4 idburns@ualberta.ca; dcoltman@ualberta.ca; cathy.cullingham@ualberta.ca

5 ² Département de sciences biologiques, Université du Montréal, Montréal, QC, Canada H3C 3J7,

6 patrick.ma.james@umontreal.ca

7 Corresponding author: Catherine Cullingham, Department of Biological Sciences, CW405,

8 University of Alberta, Edmonton, AB, Canada, T6G 2E9, Phone: 780-492-8368, Fax: 780-492-

9 9234, cathy.cullingham@ualberta.ca

Draft

Abstract

In north-central Alberta, lodgepole (*Pinus contorta* Dougl. ex Loud. var. *latifolia*) and jack pine (*Pinus banksiana* Lamb) form a mosaic hybrid zone, the spatial extent of which remains poorly defined. We sought to refine the genetic and geographic distribution of this hybrid zone in western North America to provide information important in predicting future risk of mountain pine beetle outbreaks. We used 29 SNP markers to discriminate lodgepole, jack pine, and their hybrids. We compared and contrasted spatial patterns of hybridization in northern and southern forest zones based on the colonization history of the two species. We found patterns of introgression were more similar between the zones than expected by chance, but there were significant differences between these regions at specific loci. Using logistic regression, we created a robust predictive model to distinguish among lodgepole, jack pine, and their hybrids using a combination of geographic and environmental predictors. Using AIC-based model selection, we found location, elevation, and moisture are important predictors for species class. Quantification of the genetic differences between these two regions, combined with an accurate model for predicting the spatial distribution of lodgepole, jack pine, and hybrids provide essential information for continued effective management of forest resources.

Keywords: jack pine, lodgepole pine, mountain pine beetle, hybridization, distribution modelling

29 Introduction

30 The Canadian boreal forest extends from the Yukon and northern British Columbia through to
31 the east coast of Canada (Critchfield 1985). It comprises Canada's primary source of timber
32 (Brandt et al. 2013), and includes multiple species of pine, including lodgepole pine (*Pinus*
33 *contorta* Dougl. ex Loud. var. *latifolia*) and jack pine (*Pinus banksiana* Lamb). Lodgepole and
34 jack pine differ in their ecological preferences, varying across elevation, climate (Carlson et al.
35 1999), soil, and shade tolerance (McLeod and MacDonald 1997). Lodgepole and jack pine are
36 sister species (Wheeler et al. 1983, Eckert and Hall 2006) and their divergence time has been
37 estimated to be within the Pleistocene at ~ 500 000 YBP (Dancik and Yeh 1983), with some data
38 suggesting pre-Pleistocene divergence (Eckert and Hall 2006). The current spatial distributions
39 of these two species can be explained by their respective recolonization routes following retreat
40 of the glacial ice (~12,000 yrs BP). Lodgepole pine migrated northward into British Columbia
41 (MacDonald and Cwynar 1985), east into Alberta ~10,000 yrs BP, and eventually northwest into
42 northern British Columbia and the Yukon. Jack pine migrated westward from Quebec and the
43 western Appalachians, eventually reaching Alberta ~8000 yrs BP (Godbout et al. 2005), and
44 slowly migrated northward reaching current latitudes by ~4000 yrs BP. Evidence from the fossil
45 record would suggest that two distinct contact zones formed in succession as these species
46 expanded their ranges (McLeod and MacDonald 1997), first in central Alberta, then in northern
47 Alberta/Northwest Territories.

48
49 The most-complete distribution map of lodgepole and jack pine is from Little (1971). This map
50 was created from a combination of site-surveys of morphological traits and predictions. While
51 Little (1971) acknowledges the existence of a hybrid zone between lodgepole and jack pine, and

shows an overlap in their species ranges in central Alberta and southeast Northwest Territories, the spatial extent of the hybrid zone remains poorly defined. Since 1971, mitochondria (mt), and chloroplast DNA markers have been used to investigate this hybrid zone (Dong and Wagner 1993, Godbout et al. 2012) revealing evidence for hybridization in central Alberta, and lodgepole pine introgression in Saskatchewan. Cullingham et al. (2012) investigated the hybrid zone more thoroughly using microsatellite resources. They developed a distribution model using the genetic data, and found the hybrid zone was more extensive than Little (1971) predicted, but did not find any evidence of lodgepole pine introgression in eastern AB (Saskatchewan was not sampled extensively enough to test for introgression). Their sampling, similar to Dong and Wagner (1993) and Godbout et al. (2012), did not extend into the approximated northern hybrid zone identified in the Little (1971) range maps (Figure 1) and information on the distributions of lodgepole and jack pine in this northern region remains limited. Improved characterization of the species distributions and the lodgepole pine x jack pine hybrid zone is essential to understanding future mountain pine beetle (MPB) epidemic spread risk, and for guiding pro-active management and silvicultural interventions aimed at reducing this risk.

Mountain pine beetle population dynamics are characterized by quasi-cyclic episodic outbreaks followed by endemic periods of low population density (Safranyik and Carroll 2006). Epidemic phases result in landscape level mortality of host trees, and historically have been limited in spatial scale, with localized economic impacts (Taylor and Carroll 2003, Fettig et al. 2014). The spatial extent and impact of the most recent epidemic phase of MPB growth and range expansion have been substantially greater than those previously recorded. From 1999 to 2014, MPB range expanded from southern British Columbia across the Rocky Mountains into the lowlands of

Alberta and has continued north and east towards the Northwest Territories and Saskatchewan respectively (Cullingham et al. 2011, Dhar et al. 2016, Cooke and Carroll 2017). This progression was well publicized as it resulted in the loss of over 17 million hectares of lodgepole pine forest in Canada (Walton 2012, Corbett et al. 2016). With demonstrated successful colonization of jack pine, MPB could now be considered a native-invasive (Cullingham et al. 2011). The presence of a hybrid zone between lodgepole and jack pine in Alberta may have facilitated this range expansion via mechanisms proposed in the “hybrid bridge hypothesis” (Floate and Whitham 1997). This hypothesis suggests that hybrids fill the morphological or genetic gaps between the pure species. This hypothesis has been invoked to describe host-pest interactions between gall-aphids and *Populus sp.* where pests were able to transfer to backcrossed hybrids of their host-species (Floate and Whitham 1993, Floate et al. 2016). These dynamics have also been described in plant-pathogen complexes (reviewed by Stukenbrock 2016), in which hybridization of plant hosts can promote the transfer of fungal pathogens. While *in-situ* examples of host-range expansion of plant pests via a hybrid bridge are rare (Pilson 1999), historically and spatially extensive hybridization of lodgepole and jack pine would suggest that range expansion of the MPB could be a consequence of a hybrid bridge of these two allopatric pine species.

In this study, we sought to fully characterize the current hybrid zone between lodgepole and jack pine using genetic data. To do so, we first examine the genetic characteristics of the hybrid zone to determine if there are differences in genetic introgression between the two hybrid zones, which we will define as north and south. Next, we use spatial information, and climatic predictors to model genetic ancestry using a spatial logistic regression model. Using this fitted

model, we predict the fine-scale distribution of lodgepole pine, jack pine, and their hybrids across much of their Canadian distribution including the northern extent of their range which has not been examined previously. This work is part of the TRIA-Net project (<http://tria-net.srv.ualberta.ca/>), and the over-arching goals of the project are to better understand the MPB system given exposure to a novel environment, and develop tools for industry and government to improve forest management. The development of a distribution model for lodgepole, jack pine and their hybrids will provide essential information regarding the processes and patterns that determine lodgepole and jack pine species occurrence in western Canada, and will allow for improved management of these important forest resources under threat of MPB spread.

Methods

Sample Collection

We collected pine needles from individual pine trees in a total of 61 pine stands in British Columbia (18 stands, $n = 240$), Alberta (20 stands, $n = 426$), Saskatchewan (2 stands, $n = 41$), Manitoba (1 stand, $n = 10$), Ontario (4 stands, $n = 46$), Yukon (7 stands, $n = 43$) and Northwest Territories (9 stands, $n = 151$) for a total of 957 samples (Figure 1). Samples collected from Wood Buffalo National Park were completed under permit: WB-2015-19658. For a summary of sample collection see Table S1 (supplementary information). DNA was extracted from 783 individuals for this study, while 174 were previously genotyped (Cullingham et al. 2013a, b). Geographic locations of all sampled trees were recorded using Garmin GPS units (Garmin International, Olathe, KS, USA). Needles were stored on ice in coolers, then transferred to storage at -20°C or -80°C until DNA extraction was performed.

Of those 61 stands, we extracted DNA for six Yukon stands from lab grown seedlings (23 individuals). Seeds were obtained from the National Tree Seed Centre (Natural Resources Canada) from bulk seed lots (multiple trees). Seeds were sterilized using Tween-20 and 20% bleach before stratification in autoclaved seedbeds built from 10uL micropipette tip boxes and Kim pads. Twenty seeds were germinated in each seedbed for a 12hr light/dark cycle at 25°C and 75% humidity. Seedlings were harvested when the megagametophyte could be removed easily from the seedling. Seedlings were manually ground with a pestle in individual 1.2 mL tubes of 96 well extraction plates.

DNA Extraction

To prepare pine needle tissue for extraction, needles were chopped and ground into a fine powder using a Retsch MM301 mixer mill. We extracted DNA from pine needle and pine seedling tissue using a CTAB (hexadecyl trimethyl ammonium bromide) method optimized for pine by Roe et al. (2010). We modified Roe et al's (2010) method in three ways: 1) we expanded the procedure to permit use of 96-well collection microtube plates for higher throughput; 2) samples were inverted every 30 minutes for the 2 hr 65°C incubation; and 3) all centrifugation steps were performed at 6000 g. We re-suspended pellets in 100 µL of nuclease-free water. DNA was quantified using a NanoDrop 2000 (Thermo Scientific). Only samples with a concentration greater than 20 ng/µL and 260/280 values $1.7 \leq 2.1$ were retained for genotyping.

SNP selection and characterization

Samples were typed at 29 SNP loci (Table 1) previously determined to be completely discriminating (9 SNPs), or highly differentiated (20 SNPs) between lodgepole and jack pine

(Cunningham et al. 2013a, 2013b). Annotation was performed by Cunningham et al. (2014) through comparison with multiple databases: NCBI – non redundant protein filtered for plant taxa, TAIR9 (Lamesch et al. 2012), and the Arborea white spruce gene catalogue (Rigault et al. 2011). SNP typing was performed at Delta Genomics using the Sequenom© system (Gabriel et al. 2009).

We calculated diversity measures with GenAEx ver. 6.51 (Peakall and Smouse 2012), which uses χ^2 tests for allelic diversity measures. Lodgepole, jack pine, and their hybrids were assessed separately, once defined, for the following characteristics: unbiased estimate of expected heterozygosity (H_E), observed heterozygosity (H_O), fixation index (F), and Hardy-Weinberg equilibrium (HWE). We assessed linkage disequilibrium (LD) at each locus for lodgepole, jack pine, and their hybrids separately using composite measure of LD in Genepop ver 4.2 (<http://genepop.curtin.edu.au>), with a Markov chain dememorization of 10000 and 10000 bootstrapping repetitions for significance testing (Raymond and Rousset 1995). We assessed significance with Benjamini-Hochberg False Discovery Rate (BH FDR)-corrected alpha-values (Benjamini and Hochberg 1995).

Hybrid identification

To identify hybrid ancestry, STRUCTURE ver. 2.3.4 (Pritchard et al. 2000) was used. This program implements a Bayesian admixture model to estimate Q -values or admixture proportions for individuals from K number of populations, where K is defined by the user. We ran STRUCTURE for $K = 2$ using the following parameters: burn-in of 50 000, 500 000 MCMC steps for data collection, admixture, and using a correlated allele frequency model (which assumes

independence of all samples, a conservative model). This analysis was performed first for all samples to identify “pure” lodgepole pine ($Q > 0.9$) or jack pine ($Q < 0.1$) with intermediates being considered as hybrids (Cullingham et al. 2011), then again for each group separately using the same set of parameters as above. Separate STRUCTURE analyses were performed to rule out conflicting admixture from inclusion of shore pine (*Pinus contorta* Dougl. ex Loud. var. *contorta*), a subspecies of lodgepole pine coincident with lodgepole pine stands) in western British Columbia, and to rule out possible sub-structuring in jack pine across the large species distribution. Jack pine stands are located over a wide range of environments, and while it has not been examined specifically, sub-structure may be possible due to the large geographic range included.

We used parental allele frequencies from the pure species to identify hybrid classes for sampled individuals, to discern the age and dynamics of the hybrid zone. We used the program NEWHYBRIDS (Anderson and Thompson 2002) to assign individuals to three generations of crosses: first and second generation hybrids (F1 and F2 respectively) and hybrids backcrossed with their parental species including: F1 x lodgepole pine, F1 x jack pine (B1L and B1J respectively), B1L x lodgepole pine, B1J x jack pine (B2L and B2J respectively), B1L x F2 and B1J x F2, and B2L x lodgepole pine and B2J x jack pine. NEWHYBRIDS uses a Bayesian method similar to STRUCTURE to generate genetic heritage proportions. Samples were run with a burn-in of 50 000 and a data collection of 500 000. We then compared the composition of hybrids between the northern and southern hybrid zones.

Introgression analysis

We estimated introgression for each marker using the program INTROGRESS (Gompert and Buerkle 2010). Hybrid index values were calculated for every individual at each of the 29 loci (Figure S1, supplementary). INTROGRESS requires *a priori* parental populations to estimate a hybrid index for each individual (Buerkle 2005). Fifteen individuals each from two sites in British Columbia, and 30 individuals from Ontario (Temiscaming and Algonquin) were selected to cover the parental ranges of lodgepole and jack pine, respectively (Figure 1). Two loci had ~20% missing data, as SNP typing on previous samples was performed at 27 of 29 loci. The two loci with no hybrid individuals (C17954-P346 and C52254-P578) are chloroplast loci and only one was included in the introgression analysis as they are linked.

INTROGRESS also estimates genomic clines for each locus using multinomial regressions, which estimate the effect of a given genotype at each locus on genome-wide admixture (Gompert and Buerkle 2009). Deviations likely indicate selective forces acting on the locus or closely linked regions (Gompert and Buerkle 2009). Genomic clines and hybrid indices were calculated twice; first jointly to investigate introgression across the entire hybrid zone, and then in latitude separated groups in order to investigate the potential differences between the two contact zones. For this, samples were divided by 58° into northern and southern groups (Figure 1). Rainbow Lake (north of 58°) is not continuous with the northern hybrid zone, so sensitivity analysis was used to determine if a change in grouping would result in a change in hybrid index values. There was no significant difference in hybrid indices; therefore Rainbow Lake was included as a southern stand. For each analysis, genomic clines were estimated for 1000 permutations using the parametric approach.

To determine if genomic clines of the northern and southern hybrid zones were significantly different overall, we used the '*compare.clines*' function in INTROGRESS to determine the cline similarity across loci between the two hybrid zones. We then used a binomial test to determine if our observations were different from that expected at random, where we would expect the probability (p) of a locus as having the same cline between the two regions as 0.5. The observed number of loci showing the same cline (k) was compared to the probability of observing the same, or fewer loci with the same cline. This was estimated from the cumulative probability distribution with a success rate of p over n loci.

Distribution modelling

We sought to build a spatial predictive model for the occurrence of lodgepole pine, jack pine, and their hybrids using logistic regression. In this model, Q-values derived from STRUCTURE for all individuals were used as the response variable. Potential predictors included 25 spatial and annual climate variables including those related to elevation, moisture, temperature, and precipitation (Table S3). All climate data were derived from ClimateNA v5.10 (Wang et al. 2016) available at <http://tinyurl.com/ClimateNa>, and from Environment Canada (Table S3). Environmental variables were selected for their demonstrated relationship with pine growth in provenance studies (Yeatman and Teich 1969).

Prior to fitting our model we examined the Pearson correlation between all pairs of predictors. When a correlation > 0.7 was identified, the variable in the pair with the greatest variance inflation factor (VIF) was removed. The final set of potential predictors represented a subset of 10 variables all of which exhibited correlation less than 0.7. We then sought to identify the 'best'

model describing pine species occurrence based on minimization of AIC (Akaike Information Criterion) through step-wise variable selection. Once we had identified a candidate model on the basis of AIC, we examined the VIF associated with each retained predictor and iteratively removed predictors with a $VIF > 5$ (Zurr et al. 2009). High VIFs can indicate multicollinearity between predictors and can result in overfit models and biased coefficient estimates. VIFs were calculated in R using the “car” package version 2.1-6 (Fox and Weisberg 2011).

Once the model was identified based on our AIC and VIF criteria, we assessed model performance using the Receiver Operator Curve (ROC) and the associated Area Under the Curve (AUC) statistic. ROC is a method of k -fold cross-validation that assesses the predictive capacity of a model using new data. ROC analysis was undertaken using a 60/40 training/testing split with 1000 bootstrapped replicates. The resulting “confusion matrix” of true and false positives, as well as true and false negatives, were used to generate the ROC plot of the true positive rate vs. the false-positive rate for varying threshold levels of classification for a true positive. We generated an ROC curve for each bootstrap replicate and calculated an average AUC for all ROCs associated with a single model. ROC analysis was undertaken using the package ROCR v. 1.0-7 (Sing et al. 2005) in R.

We used the selected model to spatially predict species distribution over the extent of our study area. We extracted climate data for model predictors in ArcMap10.5 using the centroids of a 10km grid across the study area, and using our model generated pine predictions for each location. We classified predicted probabilities output from our selected regression model such that predicted values < 0.1 indicated the presence of jack pine and values > 0.9 indicated the

presence of lodgepole pine; intermediate values indicated the presence of hybrids (Cullingham et al. 2011). The resulting predictive layer was masked to include only those regions where pine is found using pine distribution data from Yemshanov et al. (2012) and R. Legare (Energy Mines and Resources, YK) using the Spatial Analyst toolbox in ArcMap10.5.

Results

Genotyping

We genotyped 783 new samples using Sequenom© with > 90% sequencing success. These new data were combined with 174 previously genotyped samples (Cullingham et al. 2013a, 2013b) for a total of 957 pine samples. Five loci each for lodgepole and jack pine were out of HWE with significant heterozygote deficit, except one locus in jack with heterozygote excess. Significant LD was present at 11 locus pairs in lodgepole pine and six locus pairs in jack pine, and almost all locus pairs for hybrid individuals. A summary of diversity measures at each locus for all individuals, and for lodgepole and jack pine separately, can be seen in Table 2.

Hybrid identification

Preliminary analyses using STRUCTURE identified two distinct groups with a range of admixture proportions at $K = 2$. No additional structure was identified when lodgepole pine or jack pine individuals were analyzed separately. Of 957 individuals, 379 were assigned as jack pine ($Q > 0.9$), 436 as lodgepole pine ($Q < 0.1$) and 142 as hybrids ($0.1 \leq Q \leq 0.9$). Assignments from NEWHYBRIDS were similar to our structure results, with a greater number of hybrids assigned: 347 jack pine, 393 lodgepole pine, and 217 hybrids, the mismatches between the two programs were all individuals with split assignment across backcrossed categories. Based on the proportion

of ancestry for each of the hybrid classes (Table 3) there were no early generation hybrids identified, and there were no differences between the northern and southern hybrid zones.

Introgression analysis

Genomic clines were estimated with INTROGRESS using multinomial regressions, of the 28 loci examined, 14 showed significant deviations from neutral expectations across the entire dataset (Table 1; Figure S2, supplementary). Patterns of introgression were similar across all three datasets (complete, north and south). When comparing the northern and southern zones, two loci had significantly different patterns of introgression ($p < 0.05$). This is significantly less than you would expect if the two zones were independent ($p < 0.001$). If the two zones were independent you would expect 19 or more loci to be significantly different. The two loci were C39371-P429 and JpLpc66545p1207, both being proteins of unknown function (Table 1).

Distribution modeling

Our final selected model of pine ancestry included the physical variables of elevation, latitude, and longitude, as well as the climatic variables climate moisture index (CMI) and summer heat moisture (SHM) (Table 4). In evaluating model performance using ROC, we determined that our model had a high degree of predictive accuracy (mean AUC = 0.94; Figure 2). Species prediction with the model coincided well with previous estimates of pure distributions, while the range of hybrids is greater than previously determined by Little (1971, Figure 3).

Discussion

This study aimed to characterize the pine hybrid zone in western Canada and expand upon

research into the distribution of pine species, based on improved sampling across the range, especially in the northern and eastern ranges of lodgepole and jack pine distributions. We found evidence of the northern and southern hybrid zone described in Little (1971), and that they are more similar than expected by chance, which suggests that parallel selection may be acting across the hybrid zone in western Canada. While the zones have similar patterns of introgression overall, they are not identical on a locus by locus basis, and significant differences in clines at two loci suggest that individuals in either zone may be responding differently to selective pressures, which could indicate important adaptive differences between the naïve northern and attacked southern hybrid zones regarding mountain pine beetle exposure. We did not find any evidence of lodgepole pine introgression in eastern Alberta, or Saskatchewan, which has been described previously for mtDNA data (Dong and Wagner 1993, Godbout et al. 2012). Additionally we have used spatial environmental data to accurately predict pine species class across the landscape and create a new, fine-scale distribution map of lodgepole and jack pine. This new predictive map extends previous genomics-based models of pine species ranges (Cullingham et al 2012) into the previously unexamined northern distribution. Taken together, we now have a better understanding of the overall structure of the hybrid zone.

Introgression analysis

The process of introgression between species can result in novel assemblages of genes that can promote adaptation, reduce fitness, or any combination in between these two extremes (Harrison and Larson 2014). When analyzed as one population, all loci were significantly out of HWE (Table 2), showed heterozygote deficit, and had high fixation indices. This is to be expected as the loci were selected to be species discriminating, but the observed deviations from neutral

introgression may indicate selection is acting at these loci (Gompert and Buerkle 2009). Selection at SNP loci can lead to differential survival of individuals and inheritance of genes (Krechenwinkel and Tautz 2013). Genes of selective benefit are more likely to be passed between hybridizing species, which may result in differential introgression between individuals (Harrison and Larson 2014).

To test the hypothesis that lodgepole and jack pine have formed two genetically distinct hybrid zones, we compared patterns of introgression between the northern and southern regions. We found that the northern and southern hybrid zones were more similar than would be expected if they were independent. This may indicate selection acting in a parallel manner on the two zones despite environmental and landscape differences, where average temperatures are lower in the northern region. There were loci that had significantly different patterns of introgression between the two regions. However, these loci are in proteins of unidentified function.

Distribution Modeling

The predictive model we developed contained two main environmental variables, elevation and climate moisture index (CMI). Elevation is highly correlated with many environmental variables; especially those derived from moisture and temperature indices (Körner and Diemer 1987, Guisan and Zimmermann 2000, Körner 2007). Despite these known correlations we are confident in the independent and relevant contributions of all retained variables as their VIF values are all below 5. That is, there is no risk of multi-collinearity biasing our final model selection. Elevation, and similarly longitude, are indirect variables whose gradients greatly influence physiologically relevant variables for plants (water, heat, nutrients). Climate moisture index is

the difference between annual precipitation and potential evapotranspiration (Hogg 1997), and is positively related to the distribution of lodgepole pine and negatively related to jack pine. This is consistent with known growth conditions, as lodgepole pine proliferates in a higher moisture environment than jack pine (Yeatman and Teich 1969, Rweyongeza et al. 2007).

The model we developed revealed that the distribution of hybrids is greater in extent and width of range than that estimated previously by Little (1971) or Cullingham *et al.* (2012). We found hybrid individuals ranging from central Alberta north through to the Northwest Territories, and northeast British Columbia to the Alberta/Saskatchewan border. Notable changes from Little's (1971) assessments include the presence of individuals of hybrid ancestry in the Northwest Territories, and eastern Alberta; however individuals with hybrid ancestry also cover most of northern Alberta, and parts of the British Columbia/Alberta border. Despite improved sampling of both pine species for this study, the distribution of samples in northern Alberta is still limited and the hybrid zone in this area requires further sampling for finer-scale resolution of the hybrid map. Sampling here is challenging because of limited road access to stands in this region. Some hybridization was predicted by the model in central British Columbia. Predictions in this case are likely due to the influence of environmental variables on the model, combined with the absence of samples in those regions. Our work here has made great strides towards characterizing the extent and structure of the hybrid zone and species distributions, however there is still room for improvement with more extensive sampling.

Hybrid zone structure

In general, there are three types of hybrid zones, which can be distinguished based on the fitness

of hybrids and the effect of environment. Clinal or tension hybrid zones (Key 1968) may result from strong gene flow in the presence of selection against hybrids. This results in a narrow hybrid zone maintained by continuous hybridization between parental species and smooth clines, which are spatially correlated (May et al. 1975). Bounded hybrid superiority zones are the opposite, where selection causes hybrids to exhibit superior fitness to parental species in intermediate habitats (Moore 1977). Hybrids also show reduced fitness in parental habitats. The third type of hybrid zone is a mosaic hybrid zone, characterized by multiple generations of hybrids with variable introgression, and patchy distributions of hybrids across the landscape. Mosaic hybrid zones are formed by a combination of superior, inferior, and variable hybrid fitness caused by variable environmental selection across the regions where parental species meet. This leads to differential selection among genotypes (May et al. 1975) and pockets of pure species and hybrids across the landscape (Harrison 1986).

Based on the data we have presented here, we propose that the lodgepole $\times$ jack pine hybrid zone represents a mosaic zone, consistent with a previous assessment of the southern hybrid zone (Cullingham et al. 2012). Three lines of evidence support this idea. First, results from NEWHYBRIDS assignments opposed a tension character for the hybrid zone. In Table 3, first generation hybrids and backcrosses are assigned at low to negligible frequencies, while the majority of hybrids are late generation. This would suggest that the hybrid zone is both self-sustaining, and variable. If the lodgepole $\times$ jack pine hybrid zone were a tension zone hybrid individuals would experience a selective disadvantage (Key 1968), and we would therefore see few late generation hybrids (F2 and F3). Additionally, the absence of F1 hybrids does not necessarily indicate that hybrids are selected against (which would suggest a tension zone), as

incomplete sampling will greatly influence the representation of hybrid generations, especially across such an expansive distribution. Second, we observed differential introgression between the northern and southern zones at two genes. This is consistent with characteristics of a mosaic hybrid zone, where a selection gradient often leads to differential selection among genotypes (May et al. 1975). Finally, discontinuous selection gradients of a mosaic hybrid zone, caused by strong gene flow in the presence of selection, often results in patchy, or discontinuous spatial distributions across the landscape (Gompert et al. 2017). This can be seen in the predicted distribution map (Figure 3) where pure lodgepole and jack pine individuals can be found in multiple stands in Alberta and the Northwest Territories within the described hybrid zone. Practically, the presence of a mosaic hybrid zone translates into a patchy distribution of preferred habitat for MPB. Such patchy distribution can reduce population movement among patches and may help prevent the establishment of novel epidemic populations of MPB in hybrid and/or jack pine stands (Bone et al. 2013).

Implications for management of MPB

Improving our understanding of the distribution and genetics of pine allows for further assessments of the role hybrid zones play in host-shift dynamics. Currently, MPB has not reached the northern limit of lodgepole pine, or hybrid pine. Establishment of MPB in the northern hybrid zone may be limited by temperature, elevation, and the patchy distribution of hosts (Safranyik et al. 2010). It has been shown that naïve lodgepole pines are more susceptible and result in greater reproductive success of MPB than trees in epidemic areas (Cudmore et al. 2010). This is concerning given the northern hybrid zone, and entire boreal forest, comprise naïve trees potentially susceptible to MPB establishment. Also, hybrid trees likely represent an

intermediate phenotypic environment where levels of chemical production by hybrid hosts are intermediate to lodgepole and jack pine (Lusebrink *et al.* 2013, Erbilgin *et al.* 2014). This transitional environment could promote MPB transfer by providing a progressive change in host-environment across the landscape. However the more recent establishment of the northern zone would suggest that introgressive hybridization has occurred less frequently. Also, significant differences exist between the affected southern hybrid zone and the naïve northern hybrid zone in specific genes. A more comprehensive genome analysis would better identify differences between these zones, and whether the environment is playing a role in the differences.

Accurate pine species distribution data are essential to industry and government for resource allocation, reforestation of attacked stands, and for MPB control and spread prediction. As part of the TRIA-Net project, the resource we have created will allow for the prediction of at-risk pine stands and aid in the implementation of appropriate conservation measures to manage the potential spread of MPB through the boreal forest. Beyond the goals of the TRIA-Net project, the model may also be of use for predictive modeling of future distributions of suitable pine habitat under climate change, which will ensure appropriate seed stock is deployed for future forest sustainability

Acknowledgements

We wish to acknowledge our partners, Parks Canada, and volunteers in helping collect and process pine samples from across the distribution especially the following for their efforts: V. Shegelski, S. Pokorny, J. Burke, M. Lucky, M. MacKay, D. Yip, T. Daniels, F. Ross, J. Olesinski, S. Irwin, J. Straka, and T. Ramsfield. Particular thanks to Sebastian Lackey for project

management. An earlier version of this manuscript was improved thanks to valuable feedback from A.J. Eckert, and J. Godbout. This work was part of TRIA-Net, funding through the Natural Science and Engineering Research Council of Canada (grant no. NETGP 434810-12), with contributions from Alberta Agriculture and Forestry, fRI Research, Manitoba Conservation and Water Stewardship, Natural Resources Canada - Canadian Forest Service, Northwest Territories Environment and Natural Resources, Ontario Ministry of Natural Resources and Forestry, Saskatchewan Ministry of Environment, West Fraser, and Weyerhaeuser.

References

- Anderson, E.C., and Thompson, E.A. 2002. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* **160**: 1217–1229.
- Benjamini, Y., and Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal Stat. Soc.* **57**(1): 289-300
- Bone, C., White, J.C., Wulder, M.A., Robertson, C., and Nelson, T.A. 2013. Impact of forest fragmentation on patterns of mountain pine beetle-caused tree mortality. *Forests* **4**(2): 279-295
- Brandt, J.P., Flannigan, M.D., Maynard, D.G., Thompson, I.D., and Volney, W.J.A. 2013. An introduction to Canada's boreal zone: ecosystem processes, health, sustainability, and environmental issues. *Environmental Reviews* **21**(4): 207-226. doi.org/10.1139/er-2013-0040
- Buerkle, C.A. 2005. Maximum-likelihood estimation of a hybrid index based on molecular markers. *Mol. Ecol. Notes* **5**: 684-687
- Carlson, M.R., Murphy, J.C., Berger, V.G., and Ryrie, L.F. 1999. Genetics of elevational

- 467 adaptations of lodgepole pine in the interior. J. Sustain. For. **10**: 35-44
- 468 Cooke, B.J., and Carroll, A.L. 2017. Predicting the risk of mountain pine beetle spread to eastern
469 pine forests: Considering uncertainty in uncertain times. For. Ecol. Manag. **396**: 11–25
- 470 Corbett, L.J, Withey, P., Lantz, V.A., and Ochuodho, T.O. 2016. The economic impact of the
471 mountain pine beetle infestation in British Columbia: provincial estimates from a CGE
472 analysis. Forestry **89**:100-105.
- 473 Critchfield, W.B. 1985. The late quarternary history of lodgepole and jack pine. Can. J. For. Res.
474 **15**: 749-722
- 475 Cudmore, T.J., Björklund, N., Carroll, A.L., and Lindgren, B.S. 2010. Climate change and range
476 expansion of an aggressive bark beetle: evidence of higher beetle reproduction in naïve
477 host tree populations. J. Appl. Ecol. **47**(5): 1036–1043
- 478 Cullingham, C.I., Cooke, J.E.K., Dang, S., Davis, C.S., Cooke, B.J., and Coltman, D.W. 2011.
479 Mountain pine beetle host-range expansion threatens boreal forest. Mol. Ecol. **20**: 2157-
480 2171
- 481 Cullingham C.I., James, P.M.A., Cooke, J.E.K., and Coltman, D.W. 2012. Characterizing the
482 physical and genetic structure of the lodgepole pine x jack pine hybrid zone: mosaic
483 structure and differential introgression. Evol. Appl. **5**: 879-891
- 484 Cullingham, C.I., Cooke, J.E.K., Dang, S., and Coltman, D.W. 2013a. A species-diagnostic SNP
485 panel for discriminating lodgepole pine, jack pine, and their interspecific hybrids. Tree
486 Genet. Genom. **9**: 1119-1127
- 487 Cullingham, C.I., Cooke, J.E.K., and Coltman, D.W. 2013b. Effects of introgression on the
488 genetic populations structure of two ecologically and economically important conifer
489 species lodgepole pine (*Pinus contorta* var. *latifolia*) and jack pine (*Pinus banksiana*).

- 490 Genome **56**: 1-9
- 491 Cullingham, C.I., Cooke, J.E.K., and Coltman, D.W. 2014. Cross-species outlier detection
492 reveals different evolutionary pressures between sister species. *New Phytol.* **204**: 215-229
- 493 Dancik, B.P., and Yeh, F.C. 1983. Allozyme variability and evolution of lodgepole pine (*Pinus*
494 *contorta* var. *latifolia*) and jack pine (*P. banksiana*) in Alberta. *Can. J. Genet. Cytol.* **25**:
495 57-64
- 496 Dhar, A., Parrott, L., and Hawkins, C.D.B. 2016. Aftermath of mountain pine beetle outbreak in
497 British Columbia: Stand dynamics, management response and ecosystem resilience.
498 *Forests* **7**(8): 171
- 499 Dong, J., and Wagner, D.B. 1993. Taxonomic and population differentiation of mitochondrial
500 diversity in *Pinus banksiana* and *Pinus contorta*. *Theor. Appl. Genet.* **86**:573-578.
- 501 Eckert, A., and Hall, B. 2006. Phylogeny, historical biogeography, and patterns of diversification
502 for *Pinus* (Pinaceae): phylogenetic tests of fossil-based hypotheses. *Mol. Phylogenet.*
503 *Evol.* **40**:166-182.
- 504 Erbilgin, N., Ma, C., Whitehouse, C., Shan, B., Najjar, A., and Evenden, M. 2014. Chemical
505 similarity between historical and novel plant hosts promotes range and host expansion of
506 the mountain pine beetle in a naïve host ecosystem. *New Phytol.* **201**: 940-950
- 507 Fettig, C.J., Gibson, K.E., Munson, A.S., and Negrón, J.F. 2014. Cultural practices for
508 prevention and mitigation of mountain pine beetle infestations. *For. Sci.* **60**:450-463.
- 509 Floate, K.D., and Whitham, T.G. 1993. The “hybrid bridge” hypothesis: host shifting via plant
510 hybrid swarms. *Am. Nat.* **141**:651–662.
- 511 Floate, K.D., Godbout, J., Lau, M.K., Isabel, N., and Whitham, T.G. 2016. Plant-herbivore
512 interactions in a trispecific hybrid swarm of *Populus*: assessing support for hypotheses of

- 513 hybrid bridges, evolutionary novelty and genetic similarity. *New Phytol.* **209**:832-844.
- 514 Fox, J., and Weisberg, S. 2011. *An R companion to applied regression*, 2nd Edition. SAGE
- 515 Publications, Thousand Oaks, CA
- 516 Gabriel, S., Ziaugra, L., and Tabbaa, D. 2009. SNP genotyping using the sequenom
- 517 MassARRAY iPLEX platform. *Curr. Protoc. Hum. Genet.* **60**: 2.12.1 – 2.12.18
- 518 Godbout, J., Jaramillo-Correa, J.P., Beaulieu, J., and Bousquet, J. 2005. A mitochondrial DNA
- 519 minisatellite reveals the postglacial history of jack pine (*Pinus banksiana*), a broad-range
- 520 North American conifer. *Mol. Ecol.* **14**(11): 3497-3512
- 521 Godbout, J., Yeh, F.C., and Bousquet, J. 2012. Large-scale asymmetric introgression of
- 522 cytoplasmic DNA reveals Holocene range displacement in a North American boreal pine
- 523 complex. *Ecol. Evol.* **2**(8): 1853-1866.
- 524 Gompert, Z., and Buerkle, C.A. 2009. A powerful regression-based method for admixture
- 525 mapping of isolation across the genome of hybrids. *Mol. Ecol.* **18**: 1207–1224
- 526 Gompert, Z., and Buerkle, C.A. 2010. INTROGRESS: a software package for mapping
- 527 components of isolation in hybrids. *Mol. Ecol. Resour.* **10**: 378-384
- 528 Gompert, Z., Mandeville, E.G., and Buerkle, C.A. 2017. Analysis of population genomic data
- 529 from hybrid zones. *Annu. Rev. Ecol. Evol. Syst.* **48**: 207-229
- 530 Guisan, A., and Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology.
- 531 *Ecol. Model.* **135**: 146-186
- 532 Harrison, R.G. 1986. Pattern and process in a narrow hybrid zone. *Heredity* **56**:337–349
- 533 Harrison, R.G., and Larson, E.L. 2014. Hybridization, introgression and the nature of species
- 534 boundaries. *J. Hered.* **105**(1): 795-809
- 535 Hogg, E.H. 1997. Temporal scaling of moisture and the forest–grassland boundary in western

- 536 Canada. Agric. For. Meteorol. **84**: 115–122
- 537 Key, K.H.L. 1968. The concept of stasipatric speciation. Syst. Zool. **17**(1): 14-22
- 538 Körner, C., and Diemer, M. 1987. In situ photosynthetic responses to light, temperature and
539 carbon dioxide in herbaceous plants from low and high altitude. Funct. Ecol. **1**: 179–194
- 540 Körner, C. 2007. The use of ‘altitude’ in ecological research. Trends Ecol. Evol. **22**(11): 569-574
- 541 Krehenwinkel, H., and Tautz, D. 2013. Northern range expansion of European populations of
542 wasp spider *Argiope breunnichi* is associated with global warming-correlated genetic
543 admixture and population-specific temperature adaptations. Mol. Ecol. **22**: 2232-2248
- 544 Lamesch, P., Baradini, T.Z., Li, D., Swarbeck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher,
545 K., Alexander, D.L., Garcia-Hernandez, M., KArthikeyan, A.S., Lee, C.H., Nelson,
546 W.D., Ploetz, L., Singh, S., Wensel, A., and Huala, E. 2012. The Arabidopsis information
547 resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res. **40**:
548 D1202–D1210
- 549 Little Jr, E.L. 1971. Atlas of the United States trees, conifers and important hardwoods, Vol. 1.
550 U.S. Department of Agriculture Miscellaneous Publication No. 1146.
- 551 Lusebrink, I., Erbilgin, N., and Evenden, M.L. 2013. The lodgepole x jack pine hybrid zone in
552 Alberta, Canada: a stepping stone for the mountain pine beetle on its journey east across
553 the boreal forest? J. Chem. Ecol. **39**: 1209-1220.
- 554 MacDonald, G.M., and Cwynar, L.C. 1985. A fossil pollen based reconstruction of the late
555 quarternary history of lodgepole pine (*Pinus contorta* ssp. *Latifolia*) in the western interior
556 of Canada. Can. J. Res. **15**: 1039-1044
- 557 May, R.M., Endler, J.A., and McMurtrie, R.E. 1975. Gene frequency clines in the presence of
558 selection opposed by gene flow. Am. Nat. **109**(970): 659-676

- 559 McLeod, T.K., and MacDonald, G.M. 1997. Postglacial range expansion and population growth
560 of *Picea mariana*, *Picea glauca* and *Pinus banksiana* in the western interior of Canada. J.
561 Biogeogr. **24**: 865-881
- 562 Moore, W.S. 1977. An evaluation of narrow hybrid zones in vertebrates. Q. Rev. Biol. **52**(3):
563 263-277
- 564 Peakall, R., and Smouse, P.E. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic
565 software for teaching and research – an update. Bioinformatics **28**(19): 2537-2539
- 566 Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using
567 multilocus genotype data. Genetics **155**: 945-959
- 568 Raymond, M., and Rousset, F. 1995. GENEPOP (version 1.2): population genetics software for
569 exact tests and ecumenicism. J. Hered. **86**: 248-249
- 570 Rigault, P., Boyle, B., Lepage, P., Cooke, J.E.K., Bousquet, J., and MacKay, J.J. 2011. A white
571 spruce gene catalogue resource for conifer genome analyses. Plant Physiol. **157**: 14–28
- 572 Roe, A.D., Rice, A.V., Bromilow, S.E., Cooke, J.E.K., and Sperling, F.A.H. 2010. Multilocus
573 species identification and fungal DNA barcoding: insights from blue stain fungal
574 symbionts of the mountain pine beetle. Mol. Ecol. Res. **10**:946–959. Doi:
575 10.1111/j.1755-0998.2010.02844.x.
- 576 Rweyongeza, D.M., Dhir, N.K., Barnhardt, L.K., Hansen, C., and Yang, R.C. 2007. Population
577 differentiation of the lodgepole pine (*Pinus contorta*) and jack pine (*Pinus banksiana*)
578 complex in Alberta: growth, survival, and responses to climate. Can. J. Bot. **85**(6): 545-
579 556
- 580 Safranyik, L., and Carroll, A. 2006. The biology and epidemiology of the mountain pine beetle
581 in lodgepole pine forests. In The Mountain Pine Beetle: A Synthesis of Biology,

- 582 Management and Impacts on Lodgepole Pine. *Edited by*: L. Safranyik, and W. Wilson.
583 Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria,
584 BC pp. 3–66.
- 585 Safranyik, L, Carroll, A.L., Régnière, J., Langor, D.W., Riel, W.G., Shore, T.L., Peter, B.,
586 Cooke, B.J., Nealis, V.G., and Taylor, S.W. 2010. Potential for range expansion of
587 mountain pine beetle into the boreal forest of North America. *Can. Entomol.* **142**:415-
588 442. doi.org/10.4039/n08-CPA01
- 589 Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T. 2005. ROCR: visualizing classifier
590 performance in R. *Bioinformatics* **21**(20): 3940-3941
- 591 Stukenbrock, E.H. 2016. The role of hybridization in the evolution and emergence of new fungal
592 plant pathogens. *Phytopathology* **106**(2):104-112
- 593 Taylor, S.W., and Carroll, A.L. 2003. Disturbance, forest age, and mountain pine beetle outbreak
594 dynamics in BC: A historical perspective. In: *Mountain Pine Beetle Symposium:*
595 *Challenges and Solutions* (editors: T.L. Shore, J.E. Brooks, and J.E. Stone). Natural
596 Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Information Report
597 BC-X-399, Victoria, BC. 298 p.
- 598 Walton, A., 2012. Provincial-Level Projection of the Current Mountain Pine Beetle Outbreak:
599 Update of the infestation projection based on the Provincial Aerial Overview Surveys of
600 Forest Health conducted from 1999 through 2011 and the BCMPB model (year 9). BC
601 Ministry of Forests, Lands and Natural Resource Operations, Victoria, BC.
602 <<http://www.for.gov.bc.ca/hre/bcmpb/Year9.htm>>.
- 603 Wang, T., Hamann, A., Spittlehouse, D.L., and Carroll, C. 2016. Locally downscaled and
604 spatially customizable climate data for historical and future periods for North America.

- 605 PLoS One **11**: e0156720
- 606 Wheeler, N.C., Guries, R.P., and O'Malley, D.M. 1983. Biosystematics of the genus *Pinus*,
607 subsection Contortae. Biochem. Syst. Ecol. **11**:333–340.
- 608 Yeatman, C.W. and Teich, A.H. 1969. Genetics and breeding of jack and lodgepole pines in
609 Canada. For. Chron. **45**: 428-433
- 610 Yemshanov, D., McKenney, D.W., and Pedlar, J.H. 2012. Mapping forest composition from the
611 Canadian National Forest Inventory and land cover classification maps. Environ. Monit.
612 Assess.**184**: 4655–4669. doi:10.1007/s10661-011-2293-2
- 613

Draft

Figure 1 Distribution of lodgepole, jack pine, and hybrid samples. The potential northern and southern hybrid zones predicted by the Little (1971) range maps are included for perspective. The dividing line of 58°N is indicated which was used to test for different patterns of introgression between the two regions of contact between lodgepole and jack pine, individuals used to estimate parental allele frequencies for this analysis are star symbols.

Figure 2 Receiver operator characteristic (ROC) curves for cross-validation tests of model predictive success using 60% training, 40% testing data sets. 100 bootstrapped replicates are shown (out of a total of 1000 replicates each). Average area under the curve (AUC) was calculated for the ROC curves.

Figure 3 Predictive map for lodgepole and jack pine and their hybrids based on the Spatial Model, which includes geographic location. Predicted Q values are indicated by the species class colour at each geographic location with dark green representing lodgepole, light green representing jack, and blue indicating hybrid pine. Yellow circles indicate sample sites included in this study. Grey and dotted black lines indicate the proposed distributions of lodgepole and jack pine respectively from Little (1971).

Table 1 SNP loci, their annotations as determined from transcriptome sequencing, previously published in (Cullingham et al. 2014), and patterns of introgression. Three analyses were completed “All Samples” includes the study wide patterns, “Northern” includes only those sampling regions north of 58° latitude, while “Southern” includes all study areas south of 58° latitude including Rainbow Lake). For the patterns, “High” and “Low” indicate heterozygotes are over or under represented, respectively. An asterisk (*) indicates significant deviation from neutral introgression ($p < 0.001$). Loci in **bold** indicate significant differences between northern and southern genomic clines. We did not test introgression patterns for chloroplast loci because there are only homozygotes.

Locus	Annotation	All Samples	Northern	Southern
C17954-P346	Photosynthetic electron transfer A (chloroplast)	NA	NA	NA
C26372-P562	Calcium-dependent lipid binding (CaLB domain) family protein	Neutral	Neutral	Neutral
C35213-P325	eukaryotic aspartyl protease family protein	Neutral	Neutral	Neutral
C39371-P429	protein of unknown function (DUF3353)	*High	*High	*High
C52254-P578	Photosystem I PsA/PsB protein (chloroplast)	*No hets	*No hets	*No hets
C54523-P103	Translation protein SH3-like family	Neutral	Neutral	Neutral
C55350-P439	chaperone protein dnaJ-related	*Lodge	*Lodge	Neutral
C55378-P723	transcription factor jumonji domain-containing protein	Neutral	Neutral	Neutral
C55401-P415	transcribed locus	Neutral	Neutral	Neutral
C63961-P710	manganese transport protein MntH	*High/Lodge	*High/Lodge	*High/Lodge
C64907-P190	thioredoxin superfamily protein	Neutral	Neutral	Neutral
C66807-P512	beta-amylase/glycosyl hydrolase family 14	*Lodge	Neutral	*Lodge
C84852-P331	CRAL/TRIO domain/Sep14p-like phosphatidylinositol transfer family protein	*Jack	Neutral	*Jack
C85320-P102	DEK domain-containing chromatin associated protein	*Low/Jack	Neutral	*Low/Jack
C85506-P364	transcribed locus	*High/Lodge	Neutral	*High/Lodge
JpLpc04112p131	SSXT family protein	*Jack	Neutral	Neutral
JpLpc36252p1327	histone chaperone/global transcription factor C	Neutral	Neutral	Neutral
JpLpc39993p867	uncharacterized conserved protein (DUF2358)/SnoaL-like domain	Neutral	Neutral	*Lodge
JpLpc41319p340	uncharacterized BCR, YbaB family COG0718	Neutral	Neutral	Neutral

JpLpc44782p470	KNOX1/2 domain/KNOTTED-like	Neutral	Neutral	Neutral
JpLpc45225p571	B-cell receptor-associated protein 31-like	Neutral	Neutral	Neutral
JpLpc47089p1831	Dof-type zinc finger DNA-binding family protein	*High/Lodge	Neutral	*High/Lodge
JpLpc47778p1036	chlorophyll A-B binding family protein	* Low/Jack	* Low/Jack	*Low/Jack
JpLpc50195p453	complex I subunit	* Low/Jack	Neutral	*Low/Jack
JpLpc66545p1207	transcribed locus	Neutral	*Low/Jack	Neutral
JpLpc86157p398	RNA recognition motif/SC35-like splicing factor 28	*Jack	*Jack	*High
Lp-C45579-P117	myb-like HTH transcriptional regulator family protein	Neutral	Neutral	Neutral
Lp_c00150p459	circadian clock associated 1	Neutral	Neutral	Neutral
Lp_c12025p1415	core-2/l-branching beta-1,6-N-acetylglucosaminyltransferase family protein	*Low	*Low/Jack	*Low

Draft

Table 2 Diversity measures for 29 SNP loci across all samples of lodgepole, jack pine, and their hybrids (“All”), as well as estimates for the pure species. Two chloroplast loci with no heterozygotes are marked with an *. Number of individuals sampled at each locus (N), observed heterozygosity (H_O), unbiased estimate of expected heterozygosity (H_E), and the fixation index (F). Measures were calculated in GenAEx 6.51 (Peakall and Smouse 2012). Numbers in bold indicate loci out of Hardy-Weinberg Equilibrium (note that all loci were out of HWE when “all” samples were included).

Locus	All				Jack pine				Lodgepole pine			
	N	H_O	H_E	F	N	H_O	H_E	F	N	H_O	H_E	F
C17954-p346*	954	0.000	0.496	1.000	378	0.000	0.000	NA	436	0.000	0.000	NA
C26372-P562	954	0.127	0.494	0.743	378	0.114	0.112	-0.016	434	0.035	0.034	-0.018
C35213-P325	952	0.112	0.499	0.775	373	0.067	0.070	0.039	436	0.087	0.083	-0.046
C39371-P429	905	0.382	0.463	0.174	356	0.612	0.428	-0.434	417	0.113	0.106	-0.060
C52254-P578*	953	0.000	0.496	1.000	375	0.000	0.000	NA	436	0.000	0.000	NA
C54523-P103	956	0.086	0.499	0.828	378	0.040	0.039	-0.020	435	0.034	0.034	-0.018
C55350-P439	956	0.131	0.489	0.733	377	0.178	0.166	-0.069	436	0.023	0.023	-0.012
C55378-P723	955	0.108	0.498	0.783	378	0.050	0.054	0.069	434	0.060	0.058	-0.031
C55401-P415	957	0.119	0.498	0.761	378	0.063	0.067	0.044	436	0.067	0.064	-0.034
C63961-P710	956	0.159	0.495	0.679	378	0.159	0.160	0.004	436	0.071	0.077	0.077
C64907-P190	957	0.104	0.494	0.788	378	0.114	0.112	-0.016	436	0.011	0.011	-0.006
C66807-P512	957	0.103	0.491	0.789	378	0.122	0.119	-0.023	436	0.014	0.014	-0.007
C84852-P331	947	0.079	0.498	0.841	372	0.051	0.050	-0.026	432	0.019	0.018	-0.009
C85320-P102	956	0.060	0.499	0.880	378	0.026	0.026	-0.013	435	0.005	0.014	0.664
C85506-P364	764	0.136	0.500	0.728	351	0.128	0.149	0.141	308	0.023	0.023	-0.011
JpLpc04112p131	957	0.089	0.499	0.822	378	0.034	0.034	-0.017	436	0.046	0.045	-0.023
JpLpc36252p1327	953	0.118	0.499	0.764	377	0.064	0.062	-0.033	435	0.064	0.062	-0.033
JpLpc39993p867	957	0.159	0.493	0.678	378	0.161	0.179	0.097	436	0.050	0.049	-0.026
JpLpc41319p340	956	0.100	0.499	0.799	377	0.056	0.059	0.058	436	0.039	0.051	0.241

JpLpc44782p470	953	0.185	0.478	0.614	378	0.288	0.304	0.050	432	0.019	0.018	-0.009
JpLpc45225p571	957	0.146	0.493	0.703	378	0.153	0.177	0.131	436	0.041	0.045	0.079
JpLpc47089p1831	956	0.152	0.485	0.687	378	0.190	0.198	0.036	435	0.014	0.014	-0.007
JpLpc47778p1036	745	0.068	0.475	0.856	374	0.005	0.005	-0.003	269	0.074	0.119	0.372
JpLpc50195p453	864	0.109	0.500	0.782	351	0.034	0.039	0.125	403	0.087	0.097	0.101
JpLpc66545p1207	949	0.095	0.498	0.809	374	0.048	0.047	-0.025	434	0.028	0.027	-0.014
JpLpc86157p398	953	0.137	0.500	0.725	378	0.053	0.052	-0.027	432	0.120	0.129	0.069
Lp-C45579-P117	957	0.138	0.488	0.717	378	0.188	0.187	-0.004	436	0.016	0.020	0.214
Lp_c00150p459	948	0.162	0.476	0.658	371	0.261	0.281	0.067	435	0.018	0.018	-0.009
Lp_c12025p1415	955	0.177	0.497	0.644	378	0.040	0.044	0.097	434	0.242	0.237	-0.023

Draft

Table 3 Proportion of ancestry assigned by NEWHYBRIDS to each class: lodgepole pine (L), pure jack pine (J), their hybrids (F1, F2), and backcrossed generations: F1 x lodgepole, F1 x jack (B1L and B1J respectively), B1L x lodgepole, B1J x jack (B2L and B2J respectively), B1L x F2 and B1J x F2 (B1LF2 and B1JF2 respectively) and B2L x lodgepole and B2J x jack (B3L and B3J respectively). Included are the proportions for the entire dataset, together with the northern (North) and southern hybrid zones (South).

Hybrid Class	Proportion of class		
	All	North	South
L	35.65	38.67	34.08
J	40.95	36.13	43.46
F1	0.00	0.00	0.00
F2	0.19	0.14	0.21
B1L	0.02	0.02	0.02
B1J	0.01	0.02	0.01
B2L	6.64	6.44	6.75
B2J	4.92	7.03	3.82
B1LF2	1.45	1.72	1.32
B1JF2	3.99	3.86	4.06
B3L	2.90	3.74	2.46
B3J	3.27	2.24	3.81

Table 4 Summary of final logistic model of pine ancestry chosen using a minimization of AIC *p*-values were calculated using a Likelihood Ratio Test. VIF is the variance inflation factor, which measures correlation between predictors, and p(Lp) refers to the direction of effect on probability of lodgepole pine. The inverse sign refers to the probability of jack pine. All Predictors listed were significant.

Predictor	Coefficient	+/- SE	p-value	Effect on p(Lp)	VIF
Elevation (m)	0.022	0.0009	<0.001	+	4.051
Latitude	-0.414	0.0707	0.025	+	2.591
Longitude	-0.803	0.0163	<0.001	-	2.853
Summer heat		0.0093	0.048	+	3.118
moisture	-0.031				
Climate moisture	0.105	0.0230	0.005	+	1.004

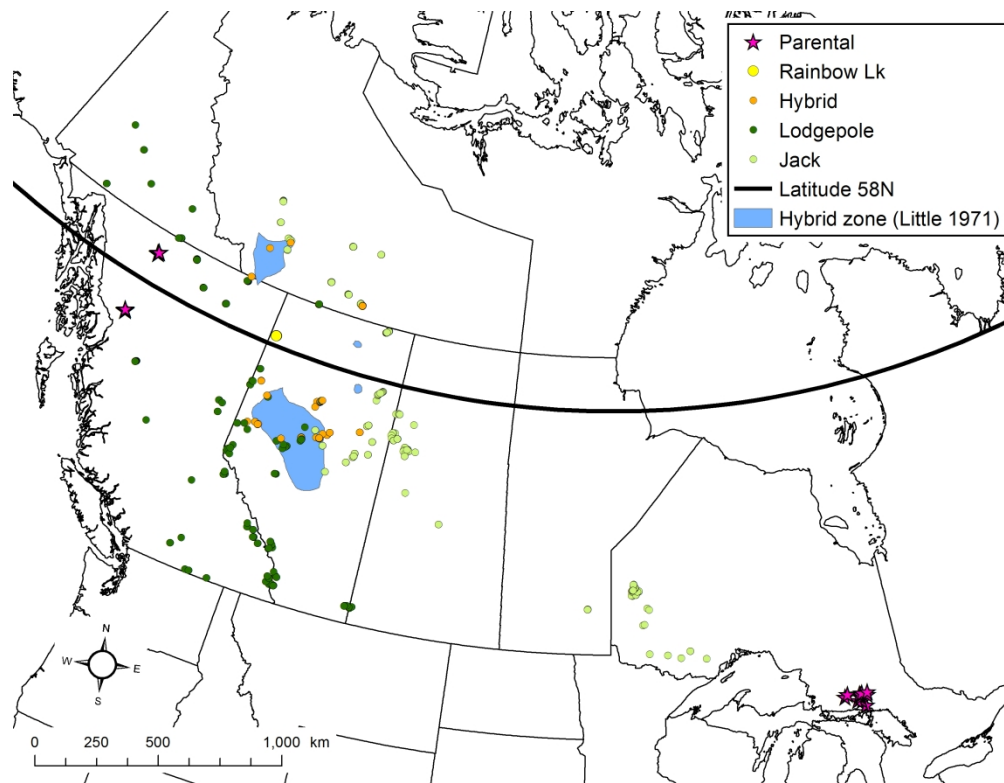


Figure 1 Distribution of lodgepole, jack pine, and hybrid samples. The potential northern and southern hybrid zones predicted by the Little (1971) range maps are included for perspective. The dividing line of 58°N is indicated which was used to test for different patterns of introgression between the two regions of contact between lodgepole and jack pine, individuals used to estimate parental allele frequencies for this analysis are star symbols.

279x215mm (300 x 300 DPI)

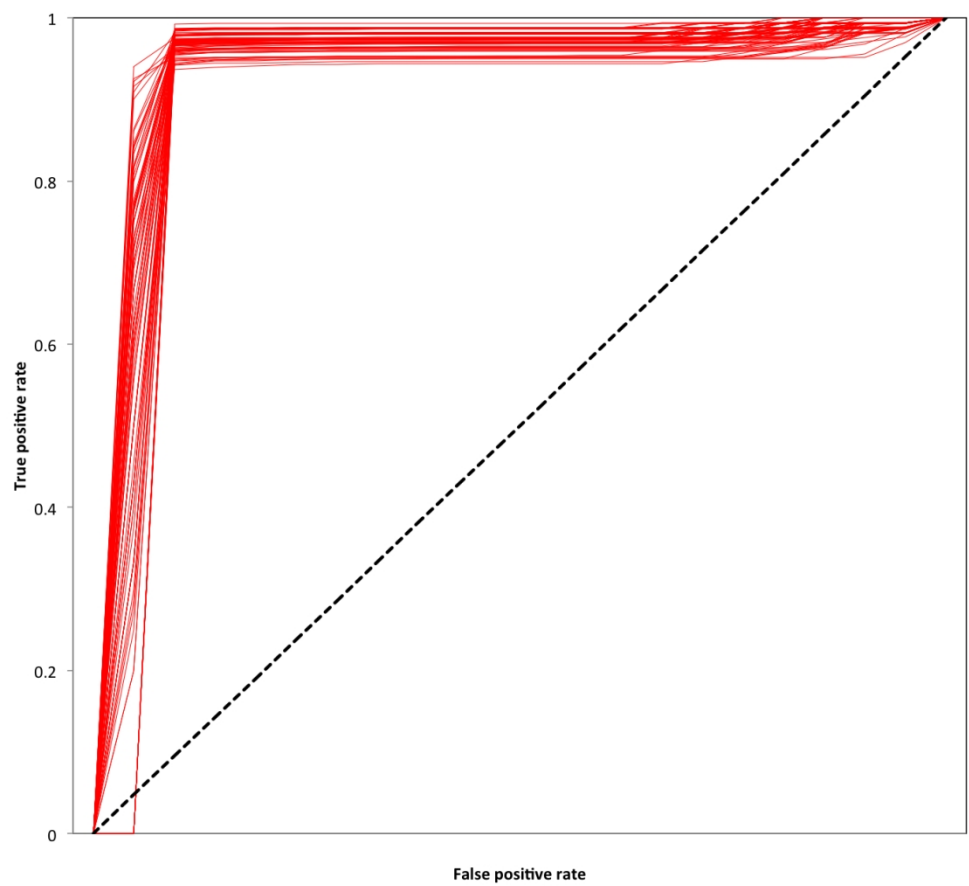


Figure 2 Receiver operator characteristic (ROC) curves for cross-validation tests of model predictive success using 60% training, 40% testing data sets. 100 bootstrapped replicates are shown (out of a total of 1000 replicates each). Average area under the curve (AUC) was calculated for the ROC curves.

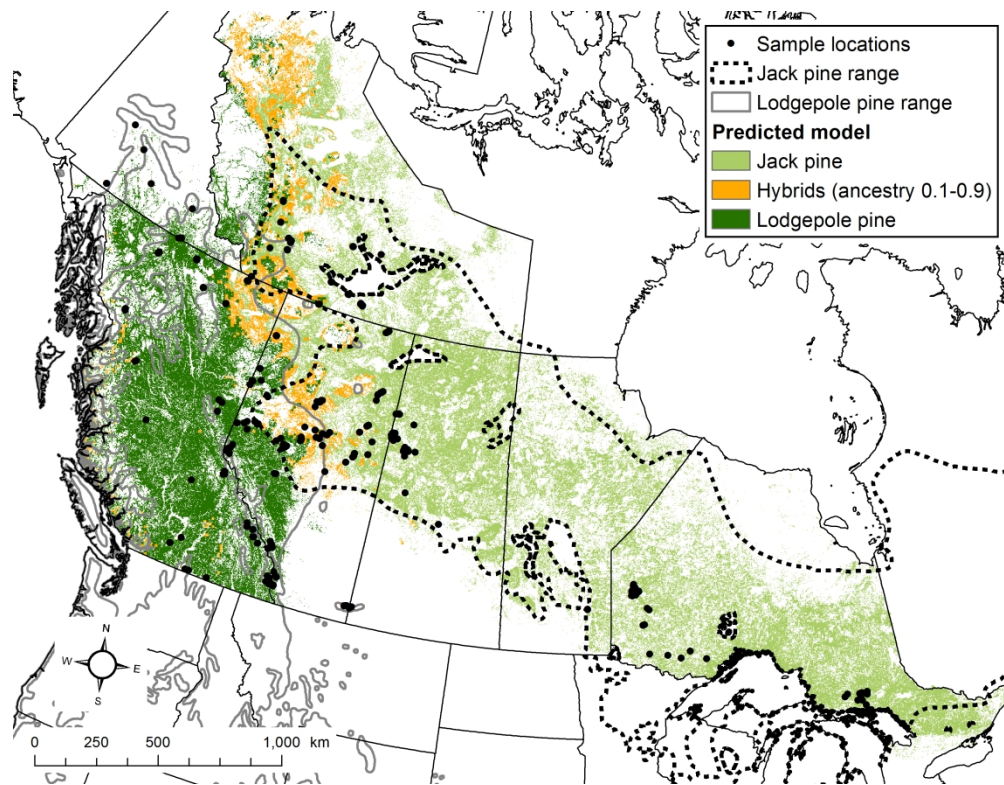


Figure 3 Predictive map for lodgepole and jack pine and their hybrids based on the Spatial Model, which includes geographic location. Predicted Q values are indicated by the species class colour at each geographic location with dark green representing lodgepole, light green representing jack, and blue indicating hybrid pine. Yellow circles indicate sample sites included in this study. Grey and dotted black lines indicate the proposed distributions of lodgepole and jack pine respectively from Little (1971).

279x215mm (300 x 300 DPI)