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

By

Ian David Burns

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

In

Systematics and Evolution

Department of Biological Sciences

University of Alberta

© Ian Burns, 2018

Abstract

In north-central Alberta, lodgepole (*Pinus contorta* Dougl. ex Loud. var. *latifolia*) and jack pine (*Pinus banksiana* Lamb) form a stable mosaic hybrid zone, which remains poorly defined. I characterized the genetic composition of the hybrid zone using samples collected from British Columbia to Ontario, and the previously un-sampled Yukon Territory and Northwest Territories, typed at 29 discriminating SNPs. I found differences in genomic clines between the northern and southern historical pine contact zones at specific loci, which could indicate important adaptive differences between the naïve northern and attacked southern hybrid zones during future mountain pine beetle range expansions. Understanding the exogenous processes influencing pine distributions in the hybrid zone is relevant to preventing pine mortality from future mountain pine beetle expansions. To characterize the spatial structure of the hybrid zone, I used logistic regression to create statistically accurate predictive models for pine species composition from a combination of geographic and environmental variables. I found that location, elevation and moisture indices are important predictors for species class. The hybrid zone takes the form of a mosaic across the entire distribution, which extends north and east of current estimates. I suggest that current species ranges be updated.

Preface

This thesis is an original work by Ian Burns. No part of this thesis has been previously published.

Acknowledgements

Thank-you to my supervisor Dr. Dave Coltman and my co-supervisor Dr. Cathy Cullingham at the University of Alberta for their endless support, encouragement, patience and guidance these past years. I am endlessly grateful. Thank-you to Dr. Patrick James at Université Laval for the education, and his contributions to my data analysis.

I would like to thank NSERC TRIA-Net for funding my research, especially Dr. Janice Cooke for her oversight. Thank-you to the provincial and territorial governments of BC, AB, SK, MB, ON, NWT and YT and our government and industrial partners for providing their time, resources and manpower to aid in sample collection and research efforts.

Thank-you to Victor Shegelski for his field expertise and to my undergraduate students Marty and Meghan for their hours spent in the lab. Finally, thank-you to my lab-mates for their support and camaraderie, and my family and friends for keeping me sane.

Table of Contents

Abstract ii

Preface iii

Acknowledgements iv

List of Tables vi

List of Figures vi

Chapter 1: Introduction 1

Chapter 2: Genetic structure of the lodgepole x jack pine hybrid zone 5

- Abstract 5
- Introduction 6
- Methods 11
- Results 19
- Discussion 22

Chapter 3: Spatial and environmental influence on pine species distributions in western Canada 38

- Abstract 38
- Introduction 39
- Methods 43
- Results 47
- Discussion 49

Chapter 4: Conclusion 60

Literature Cited 65

Appendices 69

List of Tables

Table 2.1: SNP annotations and trends in introgression 30

Table 2.2: SNP diversity measures 32

Table 2.3: Newhybrids hybrid class assignment 34

Table 3.1: Climate predictor variables 54

Table 3.2: Summary of logistic models 55

List of Figures

Figure 2.1: Map of introgression values 35

Figure 2.2: Genomic clines 36

Figure 2.3: Map of interpolated species distributions 37

Figure 3.1: Receiver operator characteristic AUCs 56

Figure 3.2: Spatial model prediction map 57

Figure 3.3: Non-spatial model prediction map 58

Figure 3.4: Spatial v. non-spatial differences map 59

Chapter 1: Introduction

The Canadian boreal forest extends from the southwest of British Columbia (BC) northeast through Alberta (AB) and the Northwest Territories (NWT), then southeast into Ontario (ON) and the southeast coast of Canada. It comprises Canada's primary source of timber, and is primarily composed of multiple species of pine, including lodgepole pine (*Pinus contorta* Dougl. ex Loud. var. *latifolia*) and jack pine (*Pinus banksiana* Lamb).

Lodgepole pine is an ecological generalist, and can be found across a range of soil types, elevations and climates, and exhibits a range in growth potential across these ranges, with high variation among stands (Carlson *et al*. 1999). Jack pine can also be found across a variety of soil types, but prefers dry, sandy soil types. However because of this soil preference and shade intolerance (McLeod & MacDonald 1997), jack pine trees rarely reach the diameters of lodgepole pine trees. Jack pine also grows at a reduced elevation range than lodgepole pine, which extends from sea level to elevations of over 3000m, while jack pine reaches only 800m. Despite these differences, lodgepole and jack pine are ecologically similar. Both species are shade intolerant and highly fire-adapted with serotinous cones, geographically concurrent, and both are threatened by the mountain pine beetle (MPB; *Dendroctonus ponderosae* Hopkins).

A major biotic threat to Canadian pine populations is from the MPB. Mountain pine beetle and lodgepole pine have a historically cyclic relationship of epidemic, and endemic growth of MPB populations (Safranyik & Carroll 2006). These cycles depend on the relative availability of susceptible host trees and MPB population size, which depends on climate suitability for MPB

brood success. Generally epidemic phases, characterized by landscape-wide mortality of host trees, are limited in geographic scale and economic impact. However, the scale and impact of the most recent epidemic phase of MPB growth and range expansion were substantially greater than those previously recorded. From 1999 to 2014, MPB range expanded from southern BC across the Rocky Mountains into the lowlands of AB and has continued north and east towards NWT and SK respectively (Cullingham *et al.* 2012; Dhar *et al*. 2016; Cooke & Carroll 2017). This progression was well publicized as it resulted in the loss of over 17 million hectares of lodgepole pine forest.

The relationship between lodgepole, MPB, and its associated blue-stain fungi has been well documented in recent years post-epidemic, and has improved our understanding of the factors promoting MPB expansion and host mortality (Safranyik 2010; James *et al*. 2011; Erbilgin *et al.* 2013). As the host species for MPB, the distribution of pine across Canada will greatly influence dispersal and establishment of MPB. The density of pine populations is inconsistent across the landscape, which may limit establishment of epidemic populations, especially in the absence of high-volume MPB long-distance dispersal (Safranyik & Carroll 2006). Epidemic populations capable of overcoming large-diameter healthy trees requires mass attacks (>50 attacks) by MPB. Patchy pine populations may limit this dispersal (Bone *et al*. 2013). Successful reproduction by only a few adult females may be enough to establish a resulting epidemic population (Safranyik & Carroll 2006), however this is improbable, and low numbers of large, suitable hosts on the landscape are likely to prevent epidemic populations.

The chemical differences between lodgepole, jack and their hybrids have also been shown to impact MPB establishment. It has been proposed that hybrid pine may act as a bridge between lodgepole and jack pine for MPB host-transfer because of intermediate traits related to MPB success. Hybrids have intermediate phenotypes between lodgepole and jack pine in traits related to host-defense including phloem chemistry and monoterpene production (Lusebrink *et al*. 2013). Additionally, novel hosts of both lodgepole and jack pine show greater monoterpene production including α-pinene and 3-carene. These compounds are used by MPB to produce the aggregation pheromone *trans*-verbenol, which attracts MPB mates. Females emerging from naïve (previously un-attacked) jack pine also had greater indicators of success, including body and brood size, than females from lodgepole pine (Erbilgin *et al*. 2014). These interactions between host chemistry and MPB success are also greatly influenced by environment. Drought stress increases monoterpene production in both pure species and their hybrids (Lusebrink *et al*. 2013), and significantly changes the influence of the blue-stain fungi *Grossmania clavigera* in both lesion length and water conductivity (Arango-Velez *et al*. 2016). These studies show that pine species will greatly influence MPB spread across the landscape. Thus it is important to have a thorough understanding of both host-pine distribution and the influence of environment on pine distribution to be able to predict future dispersal of MPB.

Unfortunately the distribution of pine species across Canada has been poorly defined in the past. The most-complete distribution map to date is from Little (1971). This map was created from a combination of site-surveys of morphological traits and predictions. While Little (1971) acknowledges the existence of a hybrid zone between lodgepole and jack pine, and shows an overlap in their species ranges in central AB and southeast NWT, the range of the hybrid zone is

not specifically delineated. Since 1971, improved methods of species delineation and hybrid detection have been developed (Cullingham *et al.* 2011, Cullingham *et al*. 2013), however an updated range map of pine has not been produced.

Improved characterization of the lodgepole pine x jack pine hybrid zone is essential to understanding and preventing future MPB epidemic spread risk. It was recently discovered that MPB has successfully attacked jack pine (Cullingham *et al.* 2011). This constitutes a major host transfer that may have been mediated by the hybrid zone between the two species (Floate & Whitham 1997). Hybridization has been found across a diverse variety of taxa, and hybridization in plants is especially common. Hybrid identification of pine has been performed using morphological and genetic methods (Critchfield 1985; Wheeler & Guries 1987; Cullingham *et al*. 2013) however the most accurate way to identify species is with genetic data.

The objective of this study is to better characterize the current hybrid zone between lodgepole and jack pine. This will be completed in three steps: First, a genetic analysis of the hybrid zone will be performed to determine if there are genetic differences across the landscape. Second, spatial and environmental analysis will be used to predict the species distributions and create a complete, fine-scale distribution map of lodgepole and jack pine in Canada. Third, genetic and spatial data will be analyzed together to determine the overall structure of the hybrid zone.

Chapter 2: Genetic structure of the lodgepole pine x jack pine hybrid zone in Canada

Abstract

Lodgepole pine (*Pinus contorta* Dougl. ex. Loud. var. *latifolia*) and jack pine (*Pinus banksiana* Lamb) form a stable, complex hybrid zone in Alberta, however species distributions are poorly defined. I characterized the hybrid zone using 939 samples collected from British Columbia to Ontario, and the previously un-sampled Yukon Territory and Northwest Territories, typed at 29 discriminating single nucleotide polymorphisms. Using introgression analysis I calculated hybrid indices and genetic hybrid proportions to determine if the previously uncharacterized northern hybrid zone is consistent with the southern distribution of pine. I found differences in genomic clines between the northern and southern historical contact zones at loci of physiological and phenotypic effect, which could indicate important adaptive differences between the naïve northern and attacked southern hybrid zones during future mountain pine beetle range expansions.

Introduction

Mountain pine beetle (MPB, *Dendroctonous ponderosae* Hopkins) is a native forest pest in British Columbia (BC). It infests multiple pine species, including lodgepole pine (*Pinus contorta* Dougl. ex Loud. var. *latifolia,* Taylor & Carroll, 2003), and more recently jack pine (*Pinus banksiana* Lamb, Cullingham *et al*. 2011). With the movement of MPB into the boreal forest and into the jack pine range, MPB could now be considered an invasive pest. The presence of a hybrid zone between lodgepole and jack pine in Alberta (AB) may have mediated this movement.

Hybridization between host species has been suggested to influence the success of host transfer in pests. The hybrid bridge hypothesis (Floate & Whitham 1993) suggests that hybrids fill the morphological or genetic gaps between pure species. This has been suggested in the host-pest interaction of gall-aphids on *Populus sp.* where pests were able to transfer to backcrossed hybrids of their host-species (Floate & Whitham 1993). This has also been seen in plant-pathogen complexes (reviewed by Stukenbrock 2016), where 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 the hybrid zone could represent a hybrid bridge of two allopatric species.

These two *Pinus* species were historically one continuous species prior to the Pleistocene (Dancik & Yeh 1983), but evolved separately during the Wisconsin glaciation. The current

distribution of lodgepole and jack pine can be explained by their colonization routes following retreat of the glacial ice (~12,000 yrs BP). Lodgepole pine migrated northward into BC after the Wisconsin glaciation (MacDonald & Cwynar 1985), east into AB ~10,000 yrs BP, and eventually northwest into northern BC and the Yukon (YT). Jack pine migrated westward from Quebec and the western Appalachians, eventually reaching AB ~8000 yrs BP (Godbout *et al*. 2005), and slowly migrated northward reaching current latitudes by ~4000 yrs BP. Evidence from the fossil record would suggest that two distinct contact zones formed in succession as these species expanded their ranges (McLeod & MacDonald 1997), first in central AB, then in northern AB/Northwest Territories (NWT).

The most complete distribution map of lodgepole and jack pine was produced in 1971 (Little 1971), which indicates the co-occurrence of lodgepole and jack pine in central AB and NWT, two regions of potential hybridization. The presence of lodgepole-jack pine hybrids has since been confirmed using genetic and genomic analyses (Godbout *et al*. 2005; Cullingham *et al.* 2012). In the central region (referred to as the southern contact zone henceforth), jack and lodgepole pine had secondary contact around 8000 yrs BP (Godbout *et al*. 2005), possibly as recently as 6000 yrs BP (MacDonald & Cwynar 1985). The second hybrid zone (the northern contact zone) likely formed as lodgepole expanded to northeastern BC by ~5500 yrs BP (MacDonald & Cwynar 1985). This would suggest that the northern contact zone in NWT/BC is considerably newer, and may therefore show fewer introgressed genes and late-generation hybrids than the southern contact zone.

Clines in a hybrid zone may result from discontinuous selection gradients, caused by strong gene flow in the presence of selection. Given the importance of environment in determining pine species distributions (Cullingham *et al*. 2012) there are likely environmental differences between the northern and southern regions of the hybrid zone, which produce the selection gradients of a hybrid zone. McLeod & MacDonald *et al*. (1997) described the progression of jack pine north through AB after the Wisconsin glaciation. They suggested that environmental differences between central AB and northern AB resulted in a slowed migration and population growth of jack pine compared to jack pine expansion across southern Canada.

Species discrimination becomes increasingly difficult with time because of continued introgression. Pure lodgepole and jack pine can be morphologically distinguished from each other primarily through differences in cone and seed morphometry (Critchfield 1985, Wheeler & Guries 1987), however the morphology of hybrids are highly variant throughout the hybrid zone, and offspring resembling one parent species are often miss-assigned. Studies comparing protein polymorphisms (Wheeler & Guries 1987), chemical profiles including turpentine and monoterpenes (Zavarin *et al*. 1969; Pollack & Dancik 1985, Critchfield 1985), and other traits including wood quality and fiber types (Wood *et al.* 2009) were unable to significantly discriminate hybrids from parental species.

Until recently, these similarities caused spatial structure of the species distribution to be poorly defined along the hybrid zone (Yang *et al.* 2007). Single nucleotide polymorphisms (SNPs), single allelic differences in DNA, have been successfully used to discriminate hybrids from parental species in complex plant systems including poplar and corn (Meirmans *et al*. 2007; Kim

et al. 2016) using multiple SNP-typing methods. As well, Cullingham *et al*. (2013) designed a species-diagnostic SNP panel using cDNA sequencing to improve hybrid discrimination in pine. This SNP panel was compared to a previously designed microsatellite panel (Cullingham *et al.* 2011) and was shown to significantly increase discrimination of hybrid classes from pure individuals, even after multiple generations of backcrossing. This panel was used to assess levels of introgression across a portion of the hybrid zone, however small numbers of hybrid and jack pine samples, together with a narrow sample distribution, limited these analyses (Cullingham *et al*. 2013).

The aim of the current study is to better characterize the hybrid zone and improve our understanding of the species distributions of lodgepole and jack pine. This will be done by analyzing the rate and extent of introgression across the hybrid zone with two main hypotheses. I hypothesize that 1) lodgepole and jack pine have formed two separate hybrid zones, one in the south, and one in the north and 2) that these two hybrid zones will differ in their patterns of introgression, likely because of variation in both age and environment between the northern and southern contact zones.

As discussed above, the hybrid zone was formed over an extended period of time, which involved multiple contact zones over an environmentally diverse and extensive spatial distribution. Previous studies concluding that the southern hybrid zone takes the form of a mosaic, were limited by a small sample size, and cannot be applied to the remainder of the hybrid zone. However with improved sampling and analysis we aim to further our understanding of hybrid zone structure, and create a new, fine-scale distribution map of lodgepole and jack

pine. This work will update our knowledge of pine from Little (1971), especially north and eastward of current assessments. Improving our understanding of the distribution and genetics of pine will allow further assessments of the role hybrid zones play in the mechanisms underlying host-shift dynamics. This is directly applicable to the pine-MPB system as MPB expands its range across the hybrid zone and into naïve jack pine stands, as variation in host genomic clines across the hybrid zone may impact the host-shift dynamics of the MPB. MPB has currently impacted only the southern hybrid zone, and not the northern hybrid zone. This may be due to difference in genetic structure of the hybrid zone, or environmental variation limiting survival or expansion of the MPB into northern stands.

Methods

Sample Collection

We collected needles from 62 pine stands in British Columbia (19 stands, *n* = 254), Alberta (24 stands, *n* = 416), Saskatchewan (4 stands, *n* = 78), Manitoba (1 stand, *n* = 10), Ontario (4 stands, *n* = 46), and the Yukon (5 stands, *n* = 39) and Northwest Territories (5 stands, *n* = 96) for a total of 939 samples (Table A1). Seven hundred and sixty-five samples were newly extracted 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 62 stands, I extracted DNA for four stands (in YT) from lab grown seedlings (20 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 kimpads. 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 first chopped into approximately 1 mm sections from frozen using a razor blade, and placed into 1.2 mL tubes of 96 well collection microtubes (Qiagen #19560). One sterile tungsten carbide grinding ball (Qiagen #69997) was added to each tube for grinding in a mixer mill. Samples were transferred to a -80°C freezer for 30 min before the first grinding step, and for 10 minutes between each grind. Plates were ground for 1 minute at 25 Hz for one grind, and 20 Hz for two additional grinds in a Retsch MM301 mixer mill.

I extracted DNA from pine needle and pine seedling tissue using a CTAB (hexadecyl trimethyl ammonium bromide) gDNA extraction method optimized for pine by Roe *et al.* (2010). Additional changes include expansion of procedure to 96 well collection microtube plates for higher throughput. As well, the samples were inverted every 30 minutes for the 2 hr 65°C incubation, and all centrifugation steps were performed at 6000 g. I re-suspended pellets in 100 µL of nuclease-free water. Then I quantified the DNA spectrophotometrically in a Thermo Scientific NanoDrop 2000, and assessed for extraction quality with 260/280 absorbance values, and estimated DNA concentration. Samples with concentration greater than 20 ng/µL and 260/280 values $1.7 \leq 2.1$ were retained for sequencing.

SNP selection and characterization

Samples were typed at 29 SNP sites (Table 2.1) previously determined to be completely discriminating (9 SNPs), or they had high frequency differences (20 SNPs) between lodgepole and jack pine (Cullingham *et al*. 2013a; 2013b). Seventeen of these were previously used to

identify individuals of hybrid origin (Cullingham *et al.* 2013a) and the remaining 12 were used for population genetic analyses (Cullingham *et al.* 2013b). Annotation was performed by Cullingham *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). Sequenom© uses multiplex reactions and MassArray technology to differentiate SNPs by molecular weight. Two sets of primers were designed for each SNP site. First, forward and reverse primers were designed to amplify the regions of DNA containing the SNP. These regions were amplified using a multiplex polymerase chain reaction (PCR). Second, internal primers were designed to selectively amplify only the polymorphic site in an iPLEX reaction.

I calculated diversity measures with GenAlEx ver. 6.5 (Peakall & Smouse 2012), which uses Chi-Squared tests of significance for allelic diversity measures. Lodgepole and 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). I assessed linkage disequilibrium (LD) at each locus for lodgepole, jack and their hybrids separately in Genepop ver 4.2 (available online at http://genepop.curtin.edu.au), with a Markov chain dememorization of 10000 and 10000 bootstrapping repetitions for significance testing (Raymond & Rousset 1995). I assessed significance with Benjamini-Hochberg False Discovery Rate (BH FDR)-corrected alpha-values (Benjamini & Hochberg 1995).

Hybrid identification

To identify hybrid ancestry, Structure (Pritchard *et al.* 2000) was used. This program uses 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. A cutoff of *Q* values ≥0.1 and ≤0.9 denotes hybrid individuals as determined by Cullingham *et al*. (2011) with simulated data. I 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 was performed first for all samples to identify pure lodgepole pine (*Q* > 0.9) or jack pine (Q <0.1), then again for each group separately. 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 coincident with lodgepole pine stands) in western BC, and to rule out possible sub-structuring in jack pine across the large species distribution respectively. Jack pine stands are located over a wide range of environments, and while it has not been examined specifically, substructuring may be possible due to variation in environment.

Hardy-Weinberg (HW) genotype frequencies were calculated for a series of hybrid generations and backcrosses using Hybridlab ver. 1.0 (Nielsen *et al*. 2006). Three generations of crosses were calculated: first and second generation hybrids (F1 and F2 respectively) and hybrids backcrossed with their parental species including: F1 x lodge, F1 x jack (B1L and B1J respectively), B1L x lodge, B1J x jack (B2L and B2J respectively), B1L x F2 and B1J x F2 (B1LF2 and B1JF2 respectively) and B2L x lodge and B2J x jack (B3L and B3J respectively). One hundred individuals of each generation resulting from the above crosses were simulated to

generate representative HW frequencies.

I used the calculated HW frequencies to designate representative classes for individuals in the sampled hybrid zone, and the program NewHybrids to assign individuals to the calculated generation classes. Newhybrids (Anderson & Thompson 2002) uses a Bayesian method similar to Structure to generate genetic heritage proportions, which may be used to discern the age and dynamics of the hybrid zone. Samples were run for five iterations with a burn-in of 50 000 and a data collection of 500 000, the average of which was taken as the final genetic heritage proportion for each class.

Introgression analysis

I estimated the level of introgression for each marker using the program Introgress (Gompert & Buerkle 2010). This program creates a hybrid index for each individual based on the proportion of alleles inherited from parental populations at both fully discriminating (fixed differences) and highly differentiated SNPs. Similar to the admixture proportion of Structure, hybrid index is denoted by a value between 0.0 and 1.0, which indicates assignment to each parental species at each locus. When fixed differences between parental populations are not present, uncertainty in inheritance is accounted for in the hybrid index when using the parametric approach (Gompert & Buerkle 2010). Hybrid index values were calculated for every individual at each of the 29 loci (Figure 2.1). Assignment to each parental category was informed from Structure to determine parental allele frequencies for each locus, as Introgress requires *a priori* parental populations to estimate a hybrid index for each individual (Buerkle 2005). Individuals are then assigned as parental or hybrid type at each locus. 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. The parametric approach corrects for the inclusion of multiple types of markers by calculating parental allele frequencies (Gompert & Buerkle 2009).

Introgress also estimates genomic clines for each locus using multinomial regressions, which involve more than two discrete outcomes. These regressions estimate the effect of a given genotype at each locus on genome-wide admixture (Gompert & Buerkle 2009). A significant deviation from neutral introgression at a locus may occur in one of five ways: overdominance, underdominance, lodgepole or jack skew (dominance of the alleles of one parental species over the other), or epistasis, an interaction between genes. Deviations likely indicate selective forces acting on the locus or closely linked regions (Gompert & Buerkle 2009). Genomic clines can be viewed graphically as plots of hybrid index (admixture) versus probability of genotype.

Genomic clines and hybrid indices were calculated twice, first jointly to investigate introgression across the entire hybrid zone, and then in latitude separated groups. The purpose of the latitude separation was to investigate the potential differences between the two contact zones. For this, samples were divided by 56 degrees into northern and southern groups. Fort McMurray (north of 56 degrees) 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 Fort McMurray 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, binomial distribution tests were performed in Excel®. The observed patterns of introgression were compared to the expected degree of similarity, i.e. the average probability of sharing a pattern of introgression at a locus in both clines, to determine if the null expectation of no difference between zones can be rejected. This was performed both on all genomic clines, and on only the significant genomic clines. To calculate the expected frequency of each type of cline, I calculated the observed frequency of each type of cline across both hybrid zones. The probability (p) of a locus sharing the same cline in both zones if the zones are independent was estimated as the sum of the squared cline frequencies. 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:

$$\Pr(X \leq k) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}$$

Prediction mapping

Hybrid location prediction mapping was performed in ArcMAP (version 10.5.1, Copyright 1995-2016 Esri) using the kriging algorithm. Kriging is an interpolation method, which uses inverse distance weighted interpolation and autocorrelation between spatial distance and an included z value (Oliver 1990), in this case hybrid index values from Introgress. I used a spherical model and a sample radius of 100 individuals to limit distance where spatial autocorrelation between data points is predictable and minimized by the semivariogram as a means to prevent overestimation of the model. The semivariogram was optimized using an ordinary predictive

model, with a nugget of 0.021, a major range of 15.200, and a partial sill of 0.252. Lodgepole pine from Cypress Hills were removed from prediction mapping analyses as they do not occur continuously with other lodgepole pine forests.

Results

Genotyping

Seven hundred and sixty-five new samples were sequenced using Sequenom© with >90% sequencing success. 43 samples were previously removed from analyses due to low success, with <50% of loci sequenced due to poor DNA quality or spotting errors. The remaining 765 samples were combined with an additional 174 samples previously sequenced with Sequenom© and Illumina© (Cullingham *et al.* 2013a, 2013b) for a total of 939 samples. The individuals previously sequenced were sequenced at 27 of 29 SNP loci. All but four loci had <5% missing data. JpLpc47778p1036 and C85506-P364 had not been included in previous studies, and therefore were not typed for ~20% of samples, while C39371-P429 and JpLpc50195p453 had 8% and 12% missing data respectively due to sequencing error. 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 2.2.

Hybrid identification

Preliminary analyses using Structure revealed two distinct groups with a range of admixture proportions at K=2. No additional admixture was revealed when lodgepole or jack individuals were analyzed separately, indicating a lack of sampling of shore pine on the western front of the sampling region, and a lack of substructuring in jack pine. Pure (Q = 1.0 and Q=0.0) lodgepole

and jack (30 each, as suggested by Structure documentation) individuals were selected as representative of parental allele frequencies for introgression analyses. Fifteen individuals each from Bell 2 and Simmons Lake in BC, and 10 individuals each from ON (Temiscaming and Algonquin), MB (Traverse Bay), and SK (north Saskatchewan) were selected to cover the parental ranges of lodgepole and jack respectively.

Of 939 individuals, 328 were assigned as jack pine ($Q>0.9$), 446 as lodgepole pine ($Q<0.1$) and 165 as hybrids ($0.1\leq Q\leq 0.9$), Pure lodgepole and jack individuals also made up >70% of genetic hybrid proportion assignment from NewHybrids (Table 2.3). First generation hybrids and their respective backcrosses had the lowest proportion assignment. Of the remaining classes, individuals were most likely to be assigned as second-generation backcrosses.

Introgression analysis

Genomic clines were estimated with Introgress using multinomial regressions, three of which for the locus JpLpc04112p131 can be seen in Figure 2.2. This locus is representative of both significant deviation from neutral introgression, and differential introgression between latitudinal groups. P values <0.028 indicate significant deviation using the BH FDR method with $\alpha=0.05$ (Benjamini & Hochberg 1995). In this case, the separation of samples into latitudinal groups resulted in significant deviation from neutral introgression at the locus with differing patterns in each group. The remaining genomic cline plots can be found in Figure A1 in the appendices.

Of 29 loci, 24 showed significant deviations from neutral expectations (Table 2.1). Overall, there were no significant differences between the northern and southern hybrid zones based on

binomial distribution tests. This was true for all loci, and for only those loci that experienced significant introgression ($p=1.02x10^{-9}$ and $p=0.444$ respectively). However when examined for all loci there were significantly more similarities between the two hybrid zones than expected by chance. Two loci were completely discriminating for all samples, however this was expected given they are chloroplast loci and therefore show significant underdominance (lack of hybrids). In addition to JpLpc04112p131 (above), two other loci, JpLpc41319p340 and JpLpc66545p1207 also showed significant and different patterns of introgression among northern, southern, and combined data sets (Table 2.1, in bold). In all 21 other cases of significant deviation, differences between latitudinally separated groups were either absent, or not significant.

Prediction Mapping

I used hybrid index values to create a prediction map of hybrid and species location with ArcGIS (Figure 2.3). The distribution of hybrids (dark blue) extends from NWT south into AB and BC.

Discussion

The aim of this study was to characterize the hybrid zone to improve our understanding of the species distributions of lodgepole pine, jack pine and their hybrids. I have done this by substantially increasing sampling of pine across the hybrid zone, then analyzing the rate and extent of introgression across the hybrid zone to determine if there are two distinguishable hybrid zones in the north and south of the hybrid distribution. The extent of the lodgepole x jack pine hybrid zone has not been well characterized in the past, as previous studies were limited in sampling (Cullingham *et al.* 2012) and could not be applied to the remainder of the hybrid zone, and therefore genomic analyses of the hybrid zone were incomplete. With expanded sampling and analysis we also created a new, fine-scale map (Figure 2.3) of lodgepole and jack pine to update our knowledge of pine distributions from Little (1971).

I found that of 939 samples, 446 were assigned as lodgepole pine, 328 were assigned as jack and 165 were of hybrid class based on Q value ranges from Cullingham *et al*. (2011). When analyzed as one population, all loci were significantly out of HWE (Table 2.2), and showed heterozygote deficit and high fixation indices (F). When separated by latitude, five loci in lodge and five loci in jack were out of HWE with significant heterozygote deficit (Table 2.2). Heterozygote deficit (homozygote excess) has been attributed in the past to selfing resulting in a Wahlund effect, which is common in coniferous trees (Stoeckel *et al.* 2006; Manka *et al*. 2011), however in this case deficit is likely due to the selection of species discriminating markers.

Sampling and mapping revealed that the distribution of hybrids is greater in extent and width of range than that estimated by Little (1971). I found individuals of hybrid character ranging from central AB north through NWT, and northeast BC to the AB/SK border. Hybrids were previously designated using only phenotypic data, however these have been shown to be less reliable for hybrid typing than genetics (Zavarin *et al*. 1969; Pollack & Dancik 1985; Critchfield 1985; Wheeler & Guries 1987; Wood *et al.* 2009, Cullingham *et al.* 2011). Notable changes from Little's (1971) assessments include the presence of individuals of hybrid ancestry in the NWT, and eastern AB; however individuals with hybrid ancestry also cover most of northern AB, and parts of the BC/AB border. Despite improved sampling of both pine species for this study, the distribution of samples in northern AB is still limited and the hybrid zone in this area requires further sampling to discern a finer scale structure before this can be properly assessed. This has been difficult because of limited road access to stands in this region. Additionally, predictive mapping would be further improved by taking into account environmental variables, as this has been shown to greatly influence the distribution of pine species across the landscape (Cullingham *et al*. 2012).

To test the hypothesis that lodgepole and jack pine have formed two separate hybrid zones, I divided the hybrid zone into two latitude-separated groups based on the historical secondary contact zones of lodgepole and jack pine (MacDonald & Cwynar 1985). I found that these two hybrid zones differed in their patterns of introgression, but may not be completely distinct. I cannot definitively state that there are two hybrid zones based on this study because of the continuous distribution of pine between the northern and southern hybrid zones, and the lack of significant differences between the zones overall based on binomial distribution tests.

Interestingly, there was significantly more similarity found between the northern and southern hybrid zones than expected by random. This may indicate selection acting in a parallel manner on the two zones despite environmental and landscape differences. However there is evidence of significant differences in the patterns of selection at specific loci between these two regions from introgression analysis, which supports that responses to this parallel selection may not be the same between individuals in each region.

Introgression analyses first required establishment of parental allele frequencies, which was performed using Structure admixture values. Parental populations identified using Q value cutoffs as determined by Cullingham *et al.* (2011) were used to determine parental allele frequencies for hybrid index calculations. Individuals were selected to best represent the parental allele frequencies across their respective native ranges. The absence of admixture in lodgepole and jack pine stands to the west and east (respectively) of the sampling distribution supports the conclusion that shore pine individuals were not sampled, or that sampled individuals are not hybrids at the typed SNPs.

Introgression may form novel complexes of genes promoting adaptation, reduce fitness, or any combination in between these two extremes (Harrison & Larson 2014). Deviations from neutral introgression as seen in 24 of the 29 SNP loci examined (Table 2.1) may be indications of selective effect acting at these loci (Gompert & Buerkle 2009). When selection acts on variation at SNP loci, this can lead to differential survival of individuals, and inheritance of genes (Krehenwinkel & Tautz 2013). Genes of selective benefit are more likely to be passed between

hybridizing species, which may result in differential introgression between individuals (Harrison & Larson 2014).

Differential introgression was seen at three of the 29 SNP loci analyzed. Two of the three differentially introgressed loci are found in genes involved in the regulation of key physiological processes and may affect growth & survival. JpLpc04112p131, an SSXT family protein involved in transcription regulation and cell proliferation during flower and leaf growth in *Arabidopsis thaliana* (Vercrussyen *et al.* 2014); and JpLpc41319p340, a member of the YbaB protein involved in DNA binding and regulation of thylakoid membrane biosynthesis (Bedard *et al.* 2017). Allelic differences at these loci may be caused by environmental differences related to climate or elevation, driven by selection. I suggest that the phenotypic impact of variation at these loci be examined in more detail to discern if changes in environmental variables have an impact on fecundity of individuals of different hybrid origin.

Differential environmental pressures on two parental species can create selection gradients, which often lead to differential selection among genotypes (May *et al*. 1975), and in some cases a variable hybrid landscape. As discussed above, differential introgression between the northern and southern zones was observed at three genes, two involved in physiological regulation, all potentially under selection. Discontinuous selection gradients, caused by strong gene flow in the presence of selection can result in patchy, or discontinuous spatial distributions across the landscape (Gompert *et al*. 2017). This can be seen in the created distribution map (Figure 2.3) where pure lodgepole pine (pink) and jack pine (light blue) individuals can be found in multiple

stands in AB and NWT within the described hybrid zone, as well as within their parental distributions. This suggests that environment has an influence on hybrid genetic composition.

Further, results from NewHybrids assignments oppose a clinal or tension character for the hybrid zone. As seen in Table 2.3, first generation hybrids and backcrosses were assigned at low to negligible frequencies, while pure lodgepole or jack pines (not listed) made up the remaining ~70% of assignments, however taking into account only those individuals assigned as a hybrid class, the majority of individuals are backcrossed or late generation hybrids. This would suggest that the hybrid zone is both self-sustaining, and variable. If the lodgepole x jack pine hybrid zone was a tension clinal zone hybrid individuals would experience selective disadvantage (Key 1968), and we would therefore see few late generation hybrids, which is not the case. Additionally, the absence of F1 hybrids does not necessarily indicate that hybrids are selected against, or that hybridization between pure individuals is not continuing to occur, as incomplete sampling will greatly influence the representation of hybrid generations, especially across such an expansive distribution. Indeed, the presence of steep transitions between pure species would suggest continued hybridization is likely. Genetic hybrid proportions are reliable for distinguishing between classes when using SNP loci up to the third generation of backcrossing (Boecklen & Howard 1997; Cullingham *et al.* 2011). This penetrance of late-generation hybrids is also apparent when considering the significant LD found in the hybrid pine, as substantial LD is expected with hybrid divergence and age (Gompert *et al*. 2017), as well as greater fitness of hybrid individuals (Arnold & Hodges 1995)

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 the mechanisms underlying host-shift dynamics. Currently, MPB has not reached the northern limit of lodgepole, or hybrid pine, and establishment of MPB in the northern hybrid zone may continue to be limited by multiple factors, including environment (temperature, elevation) and the patchy distribution of hosts (Bone *et al*. 2013).

The hybrid bridge hypothesis (Floate &Witham 1993), suggests that hybrid zones act as genetic bridges between host species. This hypothesis could influence the lodgepole x jack pine system in two ways. First, if pest-host co-adaptation exists between lodgepole and MPB in southern BC, this co-adaptation could be conferred through introgressive hybridization into hybrid and jack pine in the southern hybrid zone. This would then promote the successful establishment of MPB in pure jack pine. However it would also be true that genes of adaptive benefit to resisting MPB would also be passed through introgressive hybridization. At the same time, the hybrids of the southern zone have greater lodgepole pine and jack pine skew in alleles (Table 2.1), which would promote MPB host-transfer into pure jack pine. Secondly, the hybrid zone represents an intermediate phenotypic environment to either pure species. This has been seen in multiple studies of pine-MPB interactions (Lusebrink *et al.* 2013; Erbilgin *et al.* 2014) where levels of chemical production by hybrid hosts are intermediate to lodgepole and jack pine. This transitional environment could promote MPB transfer by providing a progressive change in host-environment across the landscape. This hypothesis would suggest that the southern hybrid zone

is more conducive to bridging the host-transfer of MPB into jack pine than the northern hybrid zone.

Lodgepole and jack pine are both susceptible hosts for the reproductive success of MPB and it has been shown in lodgepole pine that naïve trees are more susceptible and result in greater reproductive success 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. However the more recent establishment of the northern zone would suggest that introgressive hybridization has occurred less frequently, and genes of adaptive or maladaptive effect are less likely to be frequent across the zone. Additionally, significant differences exist between the affected southern hybrid zone and the naïve northern hybrid zone in specific genes related to physiological and phenotypic expression of pine individuals. Selectively advantageous phenotypic differences between lodgepole, jack and their hybrids may affect survival and brood success of the MPB when attempting to infest new hosts, or of the hosts when resisting MPB establishment. As observed by Janes *et al*. (2014) signatures of selection at SNPs in MPB may indicate adaptations to novel environments, and jack pine is a novel environment to MPB.

Conclusions

This study has aimed to characterize the hybrid zone, and has done so with improved sampling across the range, especially in the NWT, YT, eastern AB, SK and MB. Using these data, I have created a new species distribution map, which will allow for better prediction of the species distributions across the landscape. New, more accurate maps will inform both management, and

future research into the pine-MPB system. Finally, I found significant differences in genetic introgression at three loci between the northern and southern hybrid zones suggesting that two regions of the hybrid zone are potentially responding differently to selective pressures within the continuous pine distribution in western Canada. These two regions have introgressed genes that differ in genes of physiological and phenotypic effect, which could indicate important adaptive differences between the naïve northern and attacked southern hybrid zones.

Table 2.1 SNP loci and their annotations determined from transcriptome sequencing, previously published in (Cullingham *et al.* 2014). Samples grouped into all samples, samples north of 56 degrees latitude (Northern), and samples south of 56 degrees latitude (Southern). Fort McMurray is included in the southern contact zone. An asterisk (*) indicates significant deviation from neutral introgression. Loci in **bold** show both significant deviation from neutral introgression and differences in the pattern of introgression between sample groups (i.e. Northern, Southern, etc.).

Locus	Annotation	All Samples	Northern	Southern
C17954-P346	Photosynthetic electron transfer A (chloroplast)	*Underdominance	*Underdominance	*Underdominance
C26372-P562	Calcium-dependent lipid binding (CaLB domain) family protein	*Jack	*Jack	Jack
C35213-P325	eukaryotic aspartyl protease family protein	Neutral	Neutral	Neutral
C39371-P429	protein of unknown function (DUF3353)	*Overdominance	*Overdominance	*Overdominance
C52254-P578	Photosystem I PsaA/PsaB protein (chloroplast)	*Underdominance	*Underdominance	*Underdominance
C54523-P103	Translation protein SH3-like family	*Underdominance	Underdominance	*Underdominance
C55350-P439	chaperone protein dnaJ-related	*Jack	*Jack	Neutral
C55378-P723	transcription factor jumonji domain-containing protein	Neutral	Neutral	Neutral
C55401-P415	transcribed locus	Neutral	Neutral	*Jack
C63961-P710	manganese transport protein MntH	Neutral	Neutral	Neutral
C64907-P190	thioredoxin superfamily protein	Neutral	Neutral	Neutral
C66807-P512	beta-amylase/glycosyl hydrolase family 14	*Jack	Overdominance	*Jack
C84852-P331	CRAL/TRIO domain/Sep14p-like phosphatidylinositol transfer family protein	*Epistasis	Epistasis	*Epistasis
C85320-P102	DEK domain-containing chromatin associated protein	*Underdominance	Underdominance	*Underdominance
C85506-P364	transcribed locus	*Jack	Overdominance	*Jack
JpLpc04112p131	**SSXT family protein**	***Lodge**	***Underdominance**	***Epistasis**
JpLpc36252p1327	histone chaperone/global transcription factor C	*Epistasis	Lodge	Neutral
JpLpc39993p867	uncharacterized conserved protein (DUF2358)/SnoaL-like domain	*Jack	Overdominance	*Jack
JpLpc41319p340	**uncharacterized BCR, YbaB family COG0718**	***Lodge**	***Underdominance**	**Lodge**
JpLpc44782p470	KNOX1/2 domain/KNOTTED-like	*Jack	Jack	Jack
JpLpc45225p571	B-cell receptor-associated protein 31-like	*Jack	Neutral	Neutral
JpLpc47089p1831	Dof-type zinc finger DNA-binding family protein	*Overdominance	*Overdominance	*Overdominance
JpLpc47778p1036	chlorophyll A-B binding family protein	*Underdominance	*Underdominance	*Underdominance

JpLpc50195p453	complex I subunit	*Underdominance	Underdominance	*Underdominance
JpLpc66545p1207	**transcribed locus**	***Epistasis**	***Overdominance**	***Lodge**
JpLpc86157p398	RNA recognition motif/SC35-like splicing factor 28	*Lodge	*Lodge	Lodge
Lp-C45579-P117	myb-like HTH transcriptional regulator family protein	Jack	Jack	Jack
Lp_c00150p459	circadian clock associated 1	*Jack	*Jack	Jack
Lp_c12025p1415	core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein	Lodge	Lodge	*Lodge

Table 2.2 Diversity measures for 29 SNP loci in lodgepole (lodge) and jack pine (jack), and all samples. 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 GenAlEx 6.5 (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).

	All				Lodge				Jack			
Locus	N	Ho	H_E	F	N	Ho	He	F	N	Ho	H_E	F
C17954-P346*	939	0.000	0.489	0.998	446	0.000	0.000	0.000	328	0.000	0.000	0.000
C26372-P562	910	0.125	0.486	0.742	435	0.032	0.032	-0.016	317	0.095	0.096	0.013
C35213-P325	920	0.114	0.494	0.769	441	0.082	0.078	-0.043	316	0.047	0.052	0.093
C39371-P429	863	0.364	0.450	0.191	419	0.112	0.106	-0.059	297	**0.599**	**0.423**	**-0.419**
C52254-P578*	933	0.000	0.488	0.998	446	0.000	0.000	0.000	323	0.000	0.000	0.000
C54523-P103	935	0.090	0.495	0.818	445	**0.040**	**0.053**	**0.229**	327	0.028	0.027	-0.014
C55350-P439	934	0.119	0.481	0.753	443	0.023	0.022	-0.011	327	0.144	0.139	-0.037
C55378-P723	928	0.109	0.493	0.779	438	0.059	0.058	-0.031	326	0.049	0.054	0.086
C55401-P415	939	0.130	0.493	0.736	446	0.067	0.065	-0.035	328	0.064	0.062	-0.033
C63961-P710	929	0.157	0.489	0.678	443	0.063	0.070	0.092	324	0.139	0.140	0.007
C64907-P190	937	0.095	0.487	0.805	445	0.013	0.013	-0.007	328	0.076	0.079	0.034
C66807-P512	939	0.113	0.480	0.765	446	0.009	0.009	-0.005	328	0.122	0.120	-0.018
C84852-P331	907	0.080	0.490	0.836	437	0.018	0.018	-0.009	312	0.045	0.044	-0.023
C85320-P102	927	0.065	0.494	0.869	439	**0.009**	**0.018**	**0.495**	325	0.028	0.027	-0.014

C85506-P364	714	0.127	0.497	0.743	308	0.019	0.019	-0.010	290	**0.107**	**0.126**	**0.148**
JpLpc04112p131	938	0.091	0.496	0.817	445	0.040	0.040	-0.021	328	0.018	0.018	-0.009
JpLpc36252p1327	928	0.122	0.495	0.754	443	0.061	0.059	-0.031	325	0.065	0.063	-0.033
JpLpc39993p867	938	0.160	0.485	0.670	445	0.049	0.048	-0.025	328	0.165	0.176	0.065
JpLpc41319p340	936	0.104	0.495	0.790	444	**0.041**	**0.053**	**0.229**	328	**0.040**	**0.045**	**0.113**
JpLpc44782p470	933	0.183	0.466	0.607	441	0.016	0.016	-0.008	327	0.284	0.302	0.057
JpLpc45225p571	924	0.150	0.485	0.690	441	0.039	0.042	0.086	320	0.153	0.168	0.085
JpLpc47089p1831	937	0.154	0.474	0.676	445	0.013	0.013	-0.007	328	0.192	0.184	-0.047
JpLpc47778p1036	710	0.070	0.483	0.854	272	**0.074**	**0.136**	**0.460**	324	0.003	0.003	-0.002
JpLpc50195p453	818	0.112	0.496	0.773	403	0.089	0.094	0.053	294	**0.031**	**0.037**	**0.166**
JpLpc66545p1207	926	0.104	0.493	0.789	441	0.029	0.029	-0.015	323	0.053	0.051	-0.027
JpLpc86157p398	927	0.143	0.499	0.712	438	0.121	0.130	0.066	325	0.049	0.048	-0.025
Lp-C45579-P117	938	0.135	0.480	0.718	446	**0.018**	**0.022**	**0.191**	327	0.174	0.184	0.052
Lp_c00150p459	916	0.153	0.462	0.669	442	0.018	0.018	-0.009	313	0.246	0.275	0.105
Lp_c12025p1415	934	0.186	0.500	0.627	442	0.242	0.237	-0.025	328	**0.037**	**0.042**	**0.124**

Table 2.3 Newhybrids class assignment of pure lodge (L), pure jack (J), hybrid (F1, F2), and backcrossed generations: F1 x lodge, F1 x jack (B1L and B1J respectively), B1L x lodge, B1J x jack (B2L and B2J respectively), B1L x F2 and B1J x F2 (B1LF2 and B1JF2 respectively) and B2L x lodge and B2J x jack (B3L and B3J respectively).

Hybrid Class	Proportion of Classes
L	39%
J	37%
F1	0.0%
F2	0.067%
B1L	0.0%
B1J	0.043%
B2L	5.2%
B2J	7.4%
B1LF2	3.9%
B1JF2	2.6%
B3L	4.0%
B3J	0.30%

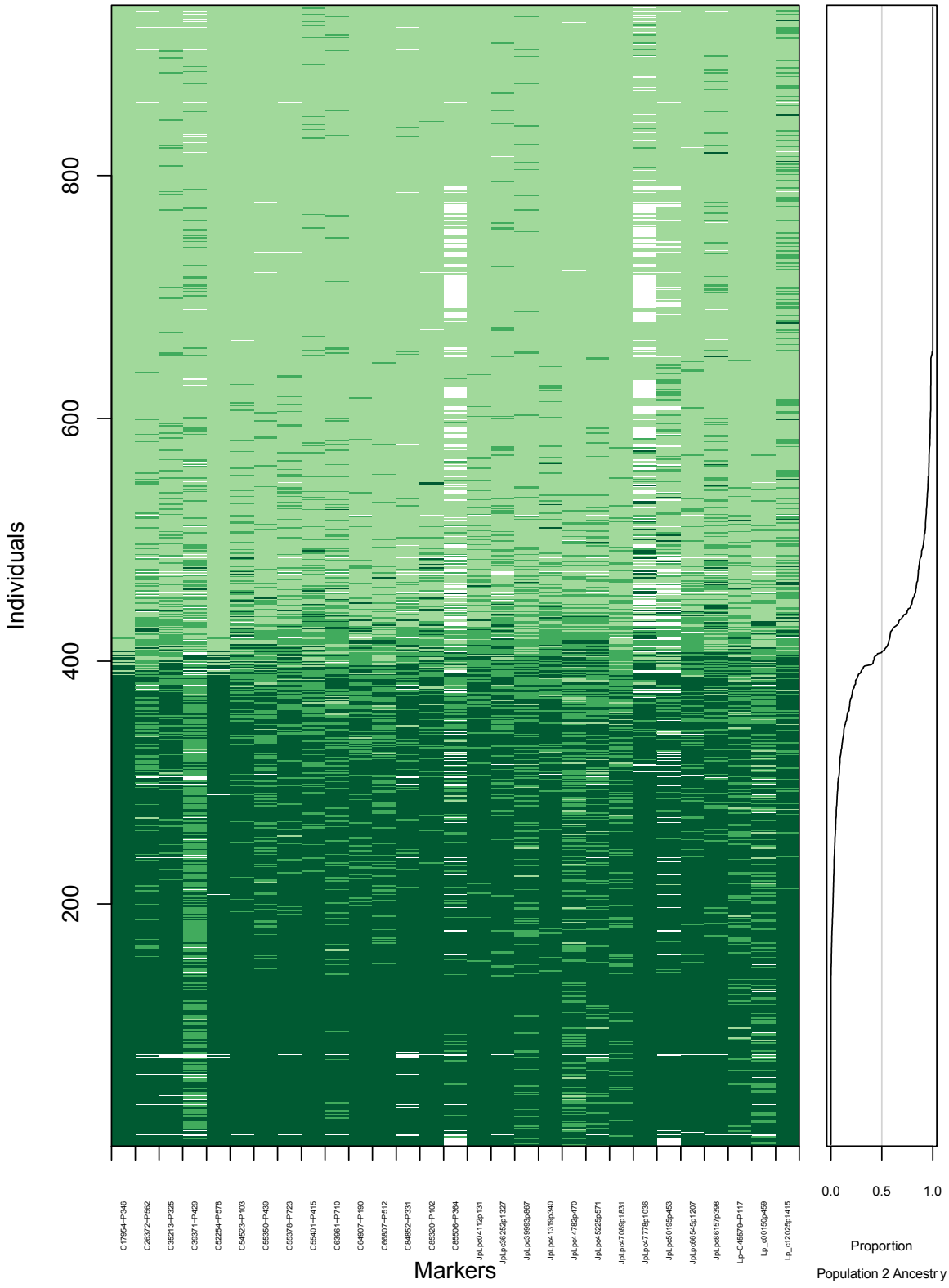


Figure 2.1 Introgression map of individuals (rows) for each SNP locus (columns). Colour indicates assignment to lodgepole pine (light green), jack pine (dark green), hybrid pine (intermediate green) or missing data (white).

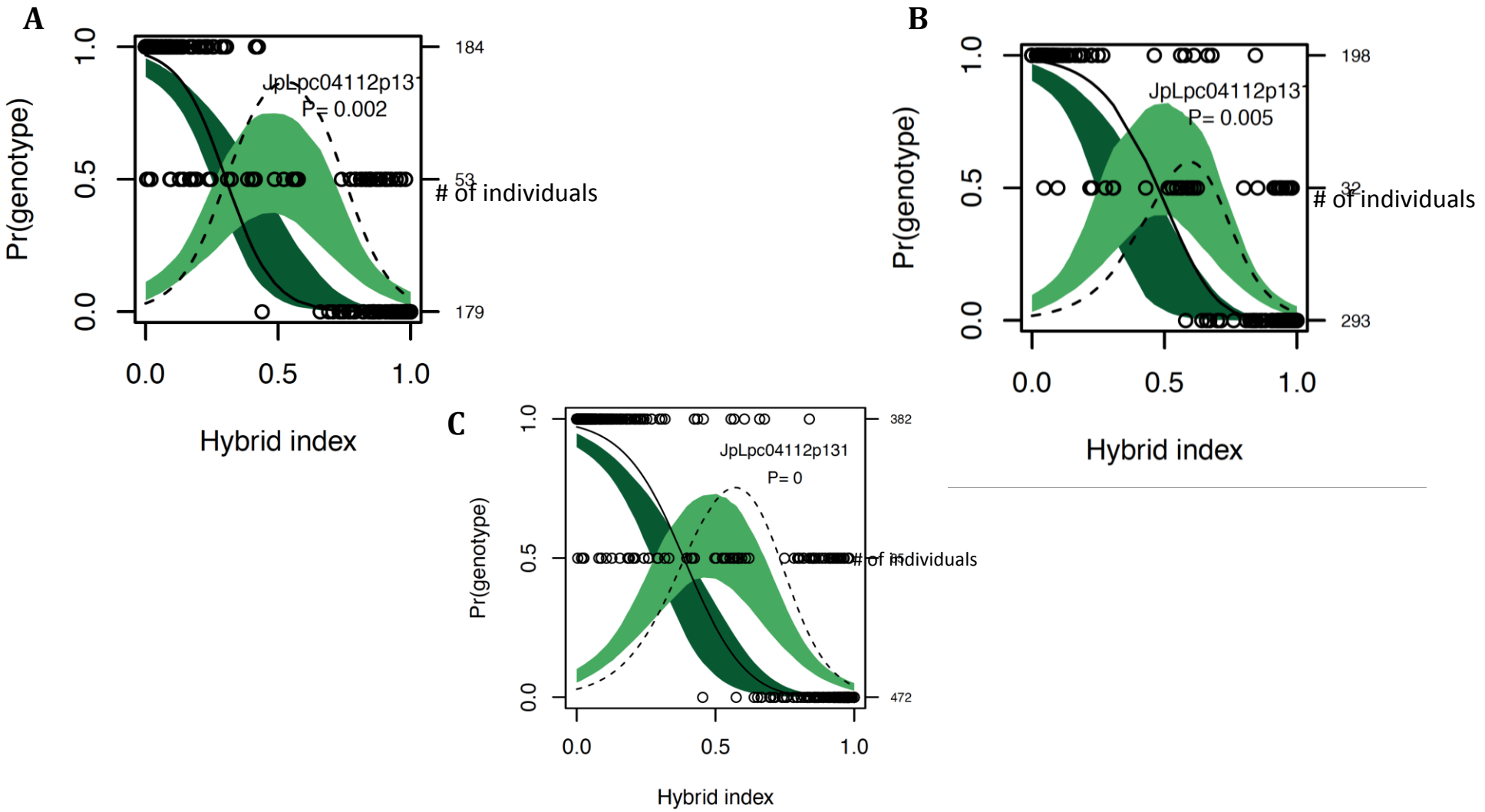


Figure 2.2 Genomic clines resulting from multinomial regressions for the JpLpc04112p131 locus. Part A: Above 56°, B: below 56°, C: All samples. Thick bands represent the 95% confidence interval of the homozygous and heterozygous (dark and light respectively) genotypes, while the solid and dashed line denote the genomic clines of the same. Open circles are the individual samples included in the regression. Numbers on the right axis are the numbers of individuals included for each genotype.

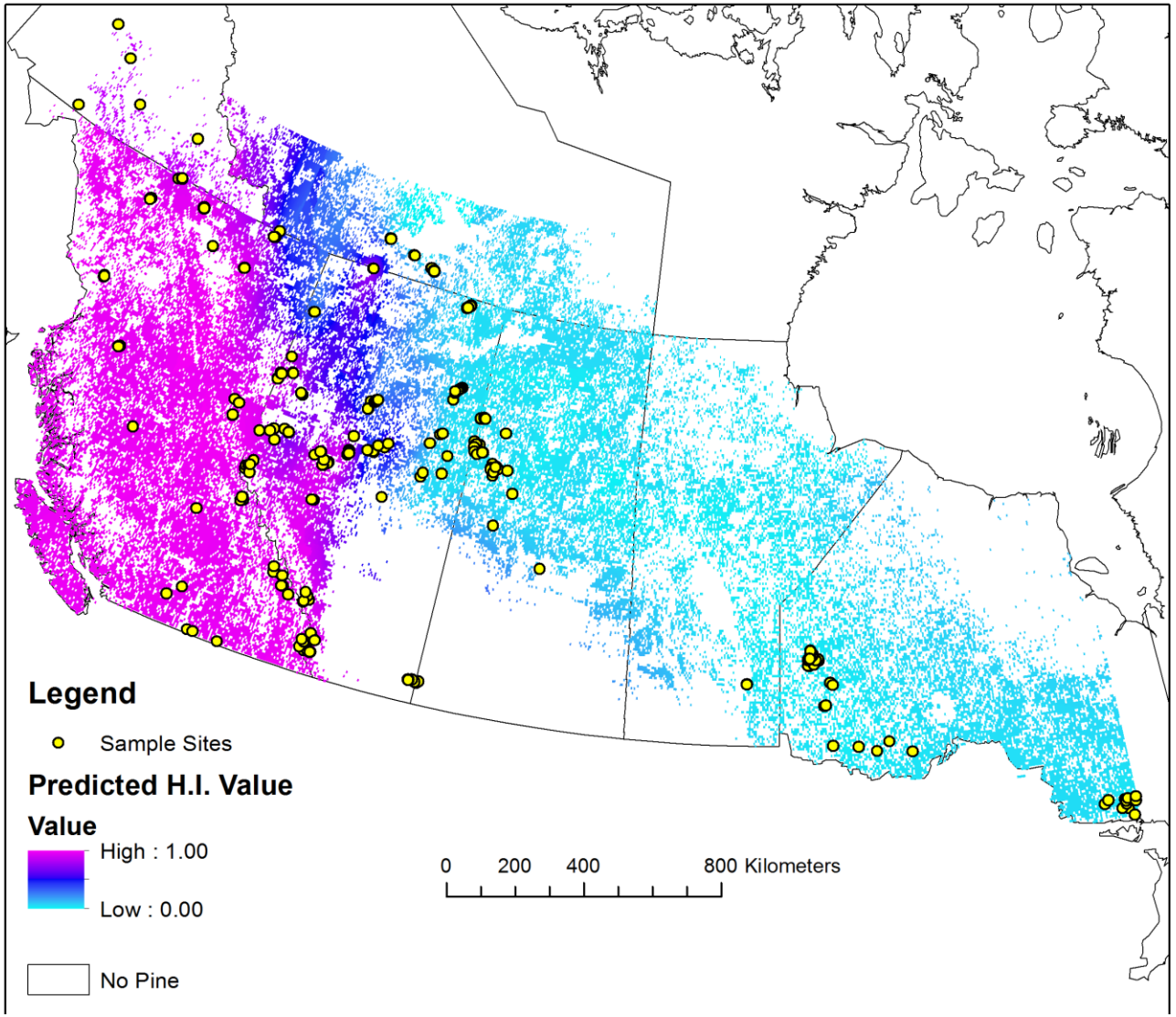


Figure 2.3 Prediction map of foliage type based on hybrid index (H.I.) values estimated from Introgress. Predicted H.I. value is the interpolated hybrid index value as determined by the distribution of hybrids and pure individuals across the landscape. High (1.0) H.I. value represents lodgepole genotypes, and low (0.0) represents jack genotypes. Prediction range covers extent of sampling distribution (yellow circles) included in the current study. Foliage presence/absence information has been modified from Yemashanov *et al*. (2012) and R. Legare (Energy Mines and Resources, YK).

Chapter 3: Spatial and environmental influence on pine species distributions in western Canada

Abstract

In north-central Alberta, lodgepole (*Pinus contorta* Dougl. ex Loud. var. *latifolia*) and jack pine (*Pinus banksiana* Lamb) form a stable mosaic hybrid zone, which remains poorly defined. Understanding the exogenous processes influencing pine distributions in the hybrid zone is relevant to reducing pine mortality from future mountain pine beetle expansions. I used 928 pine samples previously typed for genetic ancestry (Q) from British Columbia, Alberta, Saskatchewan, Manitoba, Ontario, and the Yukon and Northwest Territories to predict the Canadian range of lodgepole, jack and their hybrids. Using logistic regression I created statistically accurate predictive models for pine species composition from a combination of geographic and environmental variables. Models were validated using variance inflation factor and receiver operator characteristics. With our final model, I predicted pine species across western and central Canada and created a map of current pine species distributions. I found that location, elevation and moisture indices are important predictors for species class, and that the hybrid zone extends north and east of current estimates, suggesting current species ranges need to be updated.

Introduction

The distribution of pine species across western Canada is highly variable, and does not follow clinal patterns of transition between lodgepole pine (*Pinus contorta* Dougl. ex Loud. var. *latifolia)* and jack pine (*Pinus banksiana* Lamb*;* Cullingham *et al.* 2012). In Canada, lodgepole pine can be found in British Columbia (BC), Alberta (AB), and Yukon Territory (YT), and jack pine in the boreal forest of AB, Northwest Territories (NWT), Saskatchewan (SK), Manitoba (MB), Ontario (ON), and the eastern provinces. These species form a stable, variable hybrid zone (Moss 1949; Mirov 1956; Critchfield 1985) where the two pure species are sympatric in AB and BC. While increased sampling has improved our understanding of the distribution and genomics of pine across the hybrid zone (Chapter 2 unpublished), many regions remain un-sampled. Most regions are inaccessible from the ground year-round, and aerial sampling is too costly an endeavor for complete sampling. A complete, accurate distribution map based solely on sampling data is not practical as a result, and other methods must be pursued.

Predictive modeling has been used to successfully describe the distribution of plant species using multiple methods (summarized in Guisan & Zimmermann 2000). The preferred group of methods for species modeling is linear regression, which models a dependent variable against one or multiple explanatory variables. This is often performed using a glm (generalized linear model). Linear regression has also been used to successfully predict species distributions based on environmental and biotic variables. For example, Hicks (1980) found that tree canopy dominance and moisture indices were associated with understory species presence using multiple

linear regressions. Also, sea floor habitats of *Ophiodon elongates,* a coastal fish, were modelled using glm with success (Bassett *et al*. 2018). Importantly, linear regression modeling has also been successful for predicting hybridization, such as in the hybridization of sparrows *Ammodramus caudacutus* x *A. nelsoni* by regressing microsatellite data against local habitat variables (Walsh *et al.* 2016).

I used linear regression in this study to predict the hybridization of two plant species, lodgepole and jack pine. Hybridization in plants is common, and often mediated by climatic and geographic variables due to the close association between plant growth and climate. For example, Schweitzer *et al.* (2002), examined differences in fitness traits of hybridizing *Populus sp.* which differ in habitat elevation, and Raudnitschka *et al*. (2007) used a combination of RAPDs and morphology to examine the introgression of elevation-divided *Senecio sp.*. Additionally, the speciation of *Pinus densata,* a pine variant in China, was related to altitudinal gradients and the climate variables associated with altitude using a hybrid index (Wang 1994). In all three examples, elevation was an important predictor of hybridization between species variants. Increases in elevation/altitude (metres above sea level) are directly related to decreases in temperature and atmospheric pressure, and increases in radiation (Körner 2007), which all impact photosynthetic rates, especially in alpine trees where photosynthetic optima are variable (Körner 1987). Habitat elevation is known to differ between lodgepole and jack pine (Sykes *et al.* 1996), and was used in a previous study of pine distribution (Cullingham *et al*. 2012) as a predictor variable in linear regression.

Here I extend the research done previously by Cullingham *et al*. (2012) to include the relationship between climate and hybridization of jack pine and lodgepole pine. This has been examined previously with microsatellite loci (Cullingham *et al*. 2012), where they used a panel of 11 microsatellite loci in a logistic regression to model the distribution of pine species in BC and AB using13 predictor variables. The final model predicted the distribution of lodgepole and jack pine with a high degree of accuracy. Here I will use a panel of SNP loci, which were shown to illustrate hybrid ancestry with greater confidence (Cullingham *et al*. 2013) than with morphological, or microsatellite variation. This study will expand on previous research into predictive mapping of jack, lodgepole and hybrid pine with an enlarged data set including samples from MB, YT and NWT, regions absent from previous studies, additional sampling in AB, BC, SK and MB, and an expanded dataset of 25 predictor variables.

Specifically for jack and lodgepole pine, it has previously been shown that geography including elevation, longitude and latitude; and climate including drought and extreme minimum temperature are key indicators of species class (Cullingham *et al*. 2012) when performing predictive modeling. Multiple studies have also suggested that habitats suitable for hybridization are intermediate to those of lodgepole and jack pine (Yeatman & Teich 1969; Rweyongeza *et al.* 2007). Given this strong influence of climate on species distributions, I hypothesize that a predictive model excluding location will be able to accurately describe the distribution of pine. Additionally, I hypothesize that geography and climate will accurately co-predict for species designation across the entire population distribution, as they have in the southern provinces (Cullingham *et al*. 2012).

To assess these hypotheses, I will construct two predictive models for pine species distribution: one including geography and climate, to expand on previous modeling efforts; and a second to determine if a model excluding location will equally predict species distribution. These models will be used to create a complete distribution map to improve knowledge of the distribution of these pine species and their hybrids across Canada.

Methods

Sample Collection

928 pine samples are included in this study, 376 jack, and 409 lodgepole and 143 hybrid individuals from 67 pine stands in British Columbia (19 stands, $n = 246$), Alberta (24 stands, $n = 393$), Saskatchewan (4 stands, $n = 44$), Manitoba (1 stand, $n = 10$), Ontario (4 stands, $n = 46$), and the Yukon (5 stands, $n = 38$) and Northwest Territories (9 stands, $n = 151$) for a total of 928 samples (Table A1). A subset of samples from Chapter 2 were included, and newly collected and extracted samples in NWT were added to expand the geographic range of samples. Seven hundred and fifty-four samples were extracted 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).

Samples were typed at 29 SNP sites previously determined to be completely discriminating (9 SNPs), or they had high frequency differences (20 SNPs) between lodgepole and jack pine. Seventeen of these were previously used to identify individuals of hybrid origin (Cullingham *et al.* 2013) and the remaining 12 were used for population genetic analyses (Cullingham *et al.* 2013b). All 29 SNPs were previously used for introgression analysis of the hybrid zone (Chapter 2 unpublished) for which sample identity was determined using Structure (Pritchard *et al*. 2000) admixture proportions (Q) where $Q \leq 0.1$ indicates pure lodgepole, $Q \geq 0.9$ pure jack, and intermediate values indicate hybrid individuals (Cullingham *et al*. 2012). For more information

about sampling, DNA extraction, and sequencing please refer to sample collection information included in Chapter 2.

Distribution modeling

I initially used 25 spatial and annual climate variables including elevation, moisture and precipitation (Table 3.1) as predictors with species assignment as the response variable in a logistic regression to create a predictive model for the occurrence of jack pine, lodgepole pine, and hybrids. Logistic regression is a non-linear regression that estimates a dependent/response variable from one or more continuous explanatory variables. I used the Q values from Structure (above) as the response variable. Q values are continuous ancestry values between 0 and 1, with 0 indicating pure lodgepole pine, and 1 indicating pure jack pine based on parental allele frequencies with K=2, where K is the number of populations. The sum of ancestry values for each individual equals one. Climate data were generated with ClimateNA v5.10 (based on methodology described in Wang *et al.* 2016) available at http://tinyurl.com/ClimateNa, and from Environment Canada as indicated in Table 3.1. Environmental variables were selected for their relationship to pine growth in provenance studies (Yeatman & Teich 1969).

I started with a model including all annual climate variables, run using a binomial distribution (designation as either lodge or jack, 0 or 1) with a logit transformation. Predictors were fit as linear main effects then I selected the best model based on AIC (Akaike Information Criterion) and VIF (Variance Inflation Factor). These were performed in R using the "car" package version 2.1-6 (Fox & Weisberg 2011). AIC evaluates variable influence on the model (Akaike 1973).

Minimizing AIC during model building prevents over-explanation of the data by extraneous variables. VIF minimizes correlation between variables in the model. I used a stringent threshold of VIF<5 in the model, below the usual threshold of VIF <10 (Zurr *et al.* 2009). A VIF >10 can indicate problematic multi-co-linearity, which interferes with optimal model performance. Variables were excluded first based on VIF through sequential retention of one variable in each correlated group. Original VIF values can be found in Table A2 of the supplementary information. Once reduced, variables were then excluded based on their effect on AIC. For the final models I retained all variables with VIF<5 and significant effect on AIC.

Receiver operator characteristics (ROCs) were used to validate the selected models. ROC is a method of k-fold cross validation utilizing confusion matrices to qualify the predictive capacity of a model. The data set is split into training data on which I built the model, and testing data on which I tested the effectiveness of the created model. I used a 60/40, training/testing split (percent of data) with 1000 bootstrapping replicates. Testing data is predicted based on the training data, and confusion matrices are used for a series of "true positive" thresholds. I created ROC curves by plotting the true positive rate against the false-positive rate for each replicate and threshold. ROCs are summarized using AUC (area under the curve) to validate performance with higher AUC indicating model predictive accuracy (Figure 3.1). Cross-validation and performance were assessed with the package ROCR ver 1.0-7 (Sing *et al*. 2005) in R.

Following model validation I applied the models to create species class distribution maps in ArcMAP (version 10.5.1, Copyright 1995-2016 Esri). Environmental data was extracted at the

centroid for each 10 km grid of GPS points reaching from BC and YT to ON the extent of sampling in this study. I applied the models to this dataset to produce a predicted Q-value for each GPS point. The predictive layers were then edited to include only those locations where pine trees are located, using pine presence data modified from Yemashanov *et al*. (2012), and R. Legare (Energy Mines and Resources, YK). The differences between the predictive layers were calculated using the Map Algebra extension in ArcMAP©.

Results

I selected two final best-fit models, one including location as an attribute and one without (AIC=200.49 and AIC=572.6 respectively compared to the null model AIC=1284.8). The first model (Spatial Model) of species class as Q includes location (latitude and longitude), drought (CMI) and summer heat moisture index (SHM). The second model (Non-Spatial Model) includes elevation, drought, annual heat moisture index (AHM), relative humidity (RH), and mean annual precipitation (MAP).

Spatial*: Q ~ Elevation + Latitude + Longitude + Drought + SHM*

Non-Spatial*: Q ~ Elevation + Drought + AHM + RH + MAP*

The VIF and LRT for all variables in each model can be seen in Table 3.2. VIF for all variables included in the models are <5. Both models result in high average AUC from 1000 bootstrapping replicates (Figure 3.1), with AUC=0.9408 for the Spatial Model, and AUC=0.9412 for the Non-Spatial Model (standard deviation ±0.0227 and ±0.0232) respectively, which indicate high-accuracy predictive performance of the models.

For the two selected models, a predictive map of species class from Canada-wide GPS data was produced. The resulting maps for the Spatial Model (Figure 3.2) and the Non-Spatial Model (Figure 3.3) predict the identity of pine tree species as one of three categories: lodgepole pine ($Q \leq 0.1$), hybrid ($0.1 < Q < 0.9$), or jack pine ($Q \geq 0.9$) (Cullingham *et al*. 2011). Species prediction with the Spatial Model coincided well with previous estimates of pure distributions, while the range of hybrids is greater than previously determined. Figure 3.2 clearly shows hybrids outside

the species ranges determined by Little (1971), extending north, east, and west of previous estimates. The Non-Spatial Model predicts hybrids far outside the current estimated range of the hybrid zone (Figure 3.3), from central BC to ON. The numerical differences between the predictions of the Spatial and Non-Spatial models can be seen in Figure 3.4 as the difference in predicted Q-value at each cell.

Discussion

I created two logistic regression models of species class (Q) to accurately map the hybrid zone between lodgepole and jack pine in western Canada. Both models predict the location of lodgepole pine, jack pine, and their hybrids with high statistical accuracy based on ROC and AUC. The Spatial Model, which includes location, better predicts the current location of the hybrid zone based on current and previous sampling efforts (Figure 3.2).

The Spatial Model (Figure 3.2) predicts hybrids primarily in the known hybrid zone extending from northeast BC across central and northern AB towards SK, and from BC north into NWT along the eastern slopes of the Rocky Mountains. Some hybridization is predicted in central BC, which is unlikely given the current distributions of jack pine, and lack of evidence from genetic surveys of individuals in the area. Predictions in this case are likely due to the influence of environment on the model, combined with the absence of spatial influence on the model in unsampled regions. The influence of environment on predictions is also apparent in northern Alberta where pockets of predicted pure lodgepole have a steep transition to jack pine with little or no hybridization. This suggests that predicted habitat suitability has steeper transitions in this area than in other regions of the hybrid zone. Pockets of lodgepole pine are present in areas of high elevation (Figure 3.2), habitat more suitable to lodgepole pine persistence than jack pine or hybrids (Yeatman & Teich 1969; Carlson *et al*. 1999).

It is clear from the Spatial Model that current distributions of pine may be accurately predicted using a combination of environmental variables and geographic variables, however predictions in the absence of geographic information may not be useful for describing current distributions. The Non-Spatial Model appears to over-predict hybrid distributions compared to current sampling efforts (Figure 3.3) in regions occupied by pure jack and pure lodgepole pine. Similar to the Spatial Model, the over-estimation of hybridization in the Non-Spatial Model is likely due to the lack of sampling in over-predicted areas, especially in the distribution range of jack pine (blue regions in Figure 3.4). This sparse sampling results in the over-estimation of accuracy based on AUC. As the Non-Spatial Model does not include geographic location as a predictor, it may however be used as an indicator of regions with environment suitable to the success of each species class, especially regions suitable to hybridization and hybrid growth success. As can be seen in Figure 3.3, regions predicted as suitable for hybrid class extend from the current hybrid zone in AB, BC and NWT, east to ON, across habitat currently occupied by jack pine. These regions may represent habitats suitable for colonization by hybrid pine.

Parallel analysis of ecological and spatial processes is known to induce spatial autocorrelation. Spatial dependence in plants is primarily driven by exogenous processes (Fortin & Dale 2005), however these exogenous processes are themselves spatially auto-correlated. Environmental data (Wang *et al.* 2016) are produced with a downscaled grid using linear regression to extrapolate intermediate data points. This interpolation may induce spatial autocorrelation in the response variable of a linear model, in this case the ancestry value Q, in nearby individuals. While this could lead to underestimation of variation in Q at small scales, the large-scale geographic

distributions and regressions included in this study are unlikely to be affected. Additionally, the GPS grid from which I extracted environmental data for the predictive layer is larger (10 x 10km and 4 x 4km respectively) than that of the original environmental grid used by Wang *et al*. (2016) to interpolate their data, which would reduce induced spatial autocorrelation in Q.

Similarity between the predictive success of the two models as indicated by AUC values is likely caused by the influence of elevation in the predictive models. Elevation is highly correlated to environmental variables; especially those derived from moisture and temperature indices (Körner 1987; Guisan & Zimmermann 2000; Körner 2007). Elevation, and similarly latitude and longitude, are indirect variables whose gradients greatly influence direct/resource variables for plants (water, heat, nutrients). As such they can be expected to be of great predictive importance for habitat. Thus, when elevation was removed from the model, AIC increased dramatically, indicating that elevation explains most of the variation in the response variable.

After elevation, climate moisture index (CMI, drought) was the most important predictor of species class in both models. CMI, the difference between annual precipitation and potential evapotranspiration (Hogg 1997), is positively related to the distribution of lodgepole and negatively related to jack pine. This is consistent with known growth conditions, as lodgepole proliferates in a higher moisture environment than jack (Yeatman & Teich 1969; Rweyongeza *et al*. 2007).

As the host species for MPB, lodgepole, and now hybrid and jack pine are at risk of future MPB epidemics with climate change and MPB range expansion (Safranyik *et al*. 2010). Analysis of system dynamics will allow for the prediction of at-risk pine stands and aid in the implementation of appropriate conservation measures to prevent the continued spread of MPB through the boreal forest. These models may be used for predictive modeling of future distributions of suitable pine habitat, as has been performed with MPB. With climate change, habitat suitable for lodgepole, jack or hybrids may shift. This model could be used to predict future habitats for seed-stock management. Multiple predictive climate models are available (Wang *et al.* 2016) which provide suitable environmental data for use in this model. Additionally, the future distribution of pine is especially important for determining MPB spread risk.

Spread risk will be influenced by both climate and habitat availability for MPB. I found that precipitation (mean annual precipitation) and temperature (annual/summer heat moisture indices) are positively and negatively associated with lodgepole pine distribution respectively. Precipitation (summer precipitation) and temperature (winter temperature anomaly) have also been found to influence future MPB spread risk (Liang *et al*. 2014). Summer precipitation was negatively associated with MPB mortality, while temperature was positively associated. These corresponding climate trends can be modeled along with stand location to predict MPB spread. MPB is more successful in naive stands (Cudmore *et al*. 2010), which includes hybrid stands and jack pine stands east and north of the current attack area.

This study has expanded upon previous research into the distribution of pine species. The last Canada-wide distribution map of pine species (Little 1971) did not account for hybrid distribution except for overlaps between presumed lodgepole and jack pine distributions, and was based primarily on morphology. Multiple studies have shown that hybrids cannot be accurately discriminated from their parental species using phenotypic data (Zavarin *et al*. 1969; Critchfield 1985; Pollack & Dancik 1985; Wheeler & Guries 1987; Wood *et al.* 2009). More recently predictive mapping was performed using microsatellite loci (Cullingham *et al.* 2012) but these predictions were limited in range to BC and AB.

The Spatial and Non-Spatial models may be used to accurately predict the species class of pine using environmental data, which can be derived from georeferenced data. Pine is an agriculturally and economically important species in Canada. Accurate stand mixture data is essential to resource allocation for industry, for reforestation of attacked stands and annual logging assessments, and government for control and spread prevention.

Table 3.1 Twenty-five climate and location variables included in distribution model creation as predictors of Q-values. The ClimateNA program used to collect environmental data is available at http://tinyurl.com/ClimateNa (Wang *et al*. 2016). Predictors in bold are included in the final models.

Predictor Variable	Description	Source
Elevation	**Elevation in meters**	**NASA ASTER DEM – https://wist.echo.nasa.gov/api/**
Latitude	**Northing in decimal-degrees**	**Centroid of 10-km cell**
Longitude	**Easting in decimal-degrees**	**Centroid of 10-km cell**
Drought	**Mean climate moisture index (CMI)**	**Natural Resources Canada; Hogg 1997**
MAT	Mean annual temperature	ClimateNA v5.21; Wang *et al.* 2016
MWMT	Mean winter maximum temperature	ClimateNA v5.21
Mean CMT	Mean convective momentum transport	ClimateNA v5.21
TD	Continentality	ClimateNA v5.21
MAP	**Mean annual precipitation**	**ClimateNA v5.21**
MSP	Mean summer precipitation	ClimateNA v5.21
AHM	**Annual heat moisture index**	**ClimateNA v5.21**
SHM	**Summer heat-moisture index**	**ClimateNA v5.21**
DD0	Degree days < 0°C	ClimateNA v5.21
DD5	Growing degree days > 5°C	ClimateNA v5.21
DD18	Degree days >18°C	ClimateNA v5.21
NFFD	Number Frost Free Days	ClimateNA v5.21
FFP	Length of frost free period	ClimateNA v5.21
bFFP	Beginning of frost free period	ClimateNA v5.21
eFFP	End of frost free period	ClimateNA v5.21
PAS	Precipitation as snow	ClimateNA v5.21
EMT	Extreme minimum temperature over 30 years	ClimateNA v5.21
EXT	Extreme maximum temperature over 30 years	ClimateNA v5.21
Eref	Reference evaporation	ClimateNA v5.21
CMD	Climatic moisture deficit	ClimateNA v5.21
RH	**Average relative humidity**	**ClimateNA v5.21**

Table 3.2 Summary of logistic models, for the Spatial Model 1 (A) and the Non-Spatial Model (B) chosen with minimization of AIC and VIF<5. LRT refers to the likelihood ratio test performed in R to determine each predictor's significance in the model. 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.

A

Predictor	VIF	Coefficient	+/- Std. Err.	p-value	LRT	Effect on p(Lp)
Elevation (m)	4.051	0.022	0.0009	<0.001	568.32	+
Latitude	2.591	-0.414	0.0707	0.025	86.93	+
Longitude	2.853	-0.803	0.0163	<0.001	58.01	-
SHM	3.118	-0.031	0.0093	0.048	0.19	+
Drought (CMI)	1.004	0.105	0.0230	0.005	110.26	+

B

Predictor	VIF	Coefficient	+/- Std. Err.	p-value	LRT	Effect on p(Lp)
Elevation (m)	2.629	0.010	0.0008	<0.001	568.32	+
Drought (CMI)	1.002	0.059	0.0086	<0.001	91.10	+
AHM	1.543	-0.163	0.0351	<0.001	5.04	-
RH	2.772	0.092	0.0419	<0.001	1.00	+
MAP	2.737	-0.004	0.0009	0.002	24.62	-

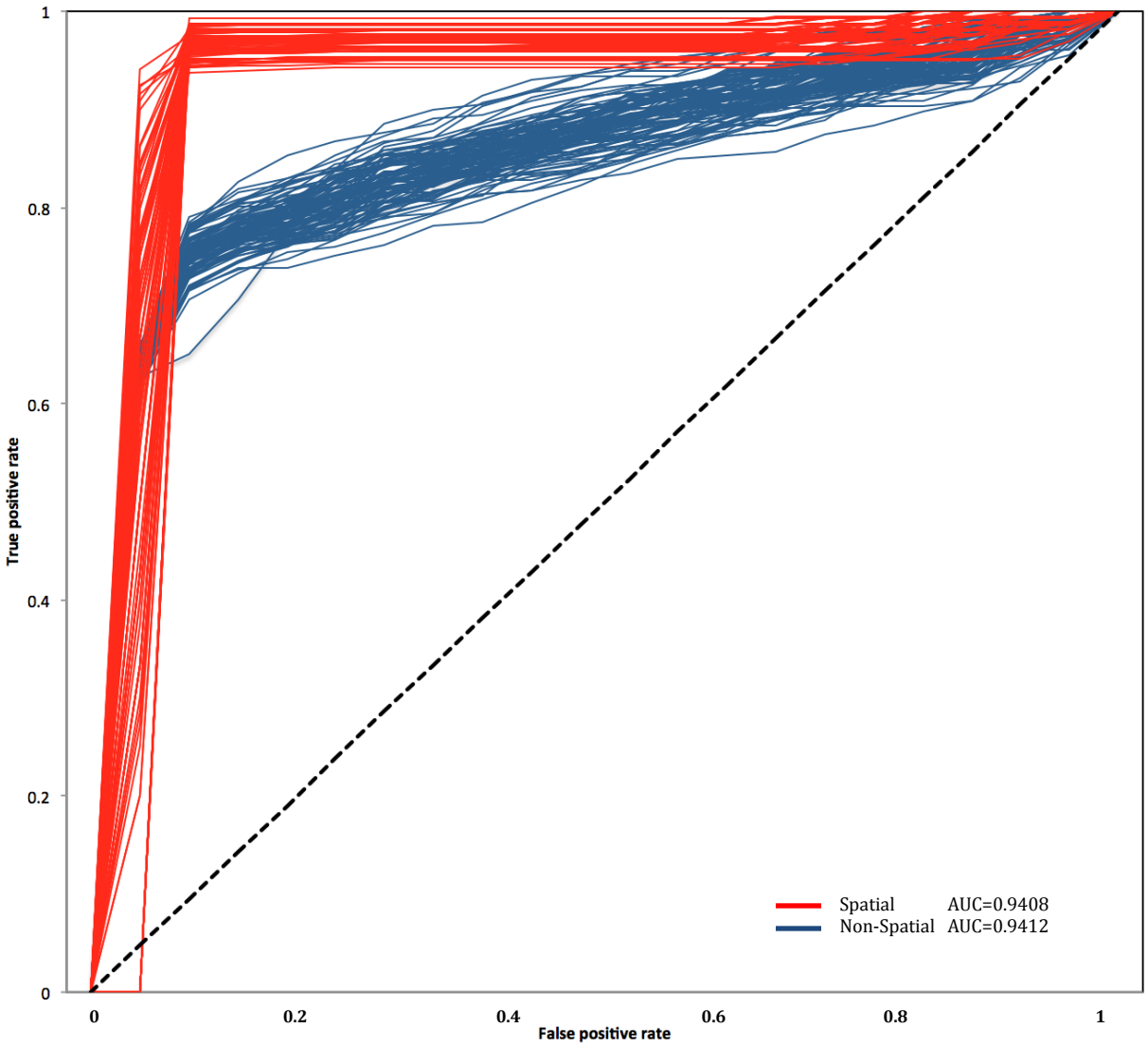


Figure 3.1 Receiver operator characteristic (ROC) curves for cross-validation tests of model predictive success using 60% training, 40% testing data sets. 100 bootstrapped replicates for each model are shown (out of a total of 1000 replicates each). Average area under the curve (AUC) was calculated for each set of ROC curves.

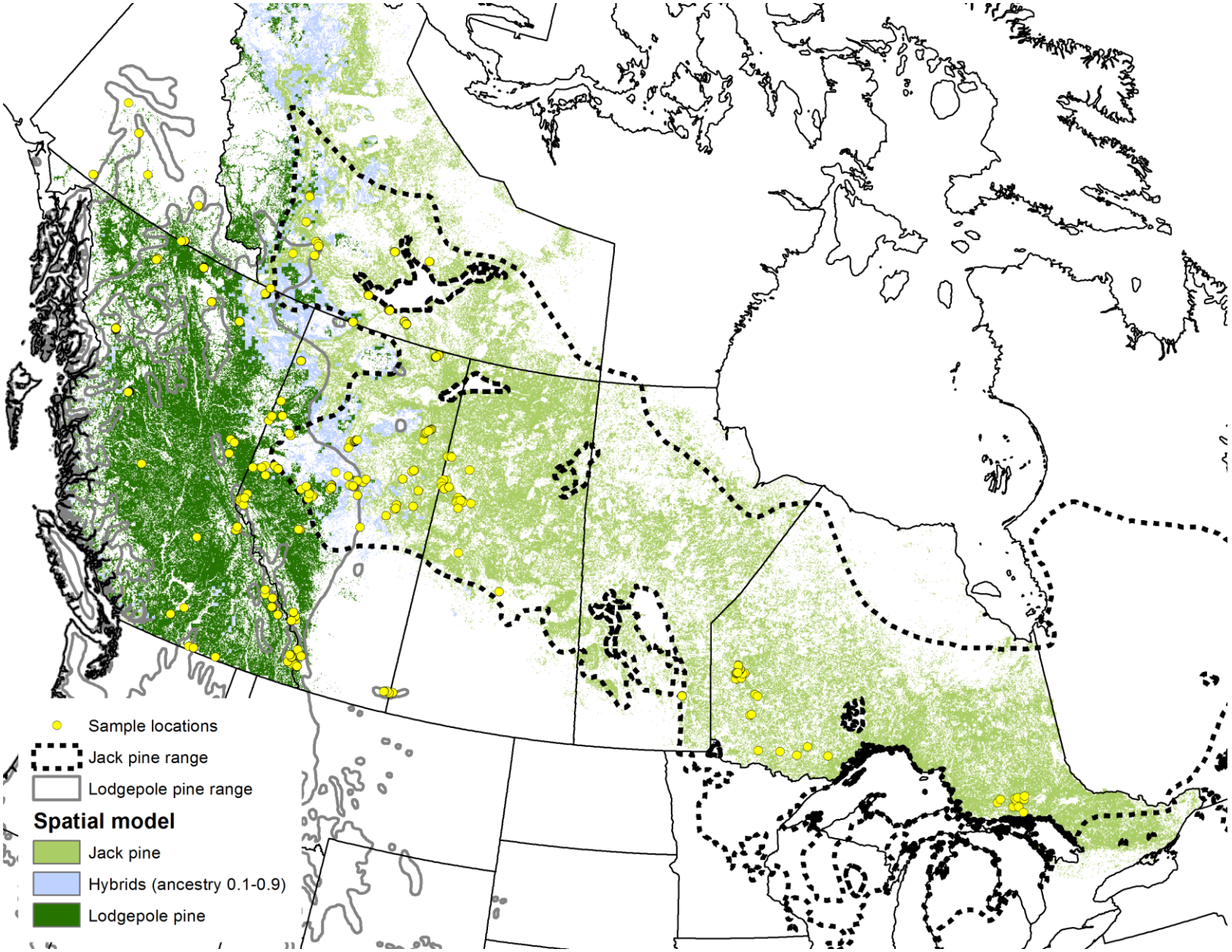


Figure 3.2 Predictive map for lodgepole and jack pine and their hybrids based on the Spatial Model, which includes geographic location. Predicted Q values are indicated by the species class colour at each geographic location with dark green representing lodgepole, light green representing jack, and blue 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).

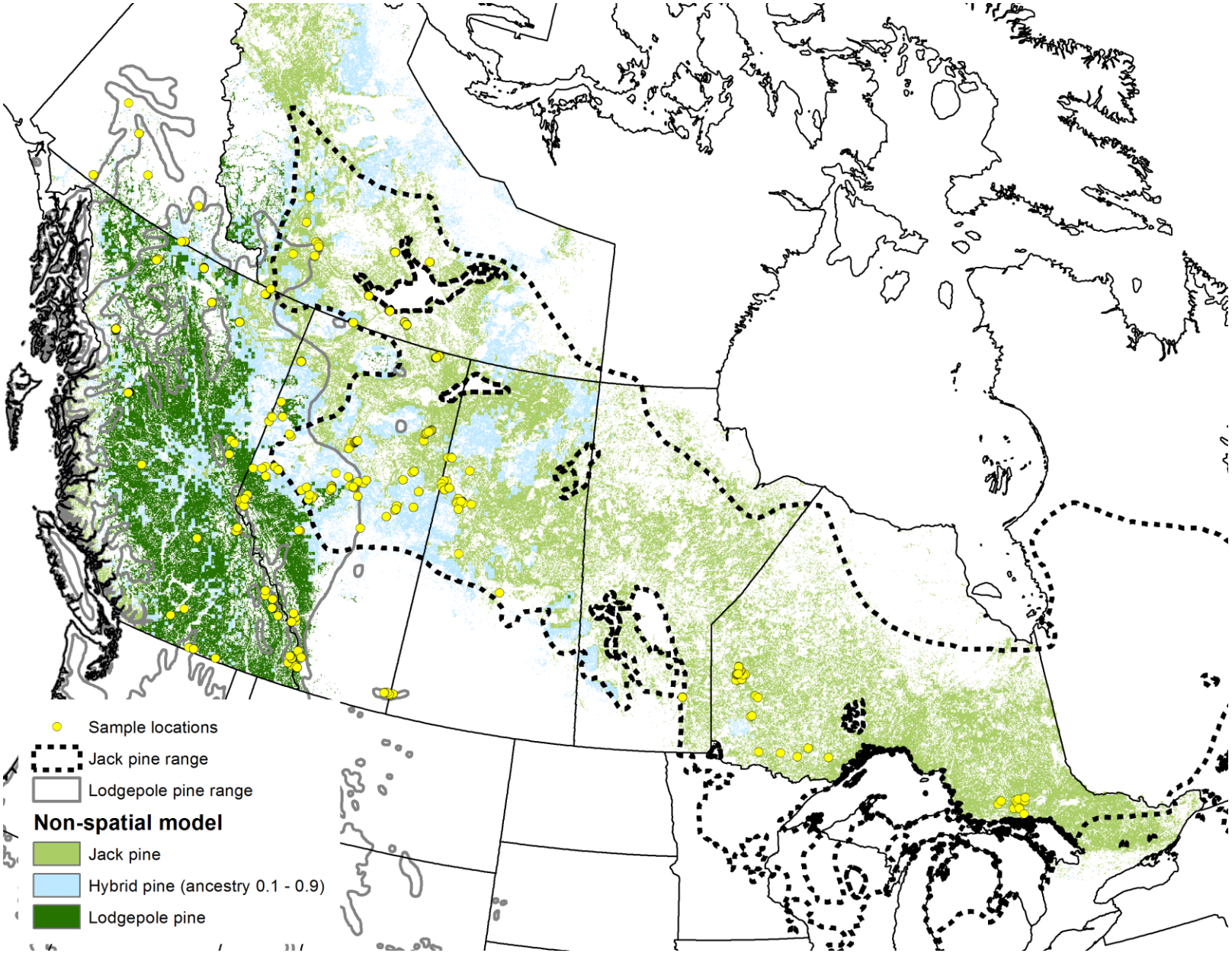


Figure 3.3 Predictive map for lodgepole and jack pine and their hybrids based on the Non-Spatial Model, which includes solely environmental variables. Predicted Q values are indicated by the species class at each geographic location with dark green representing lodgepole, light green representing jack, and blue indicating hybrids 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).

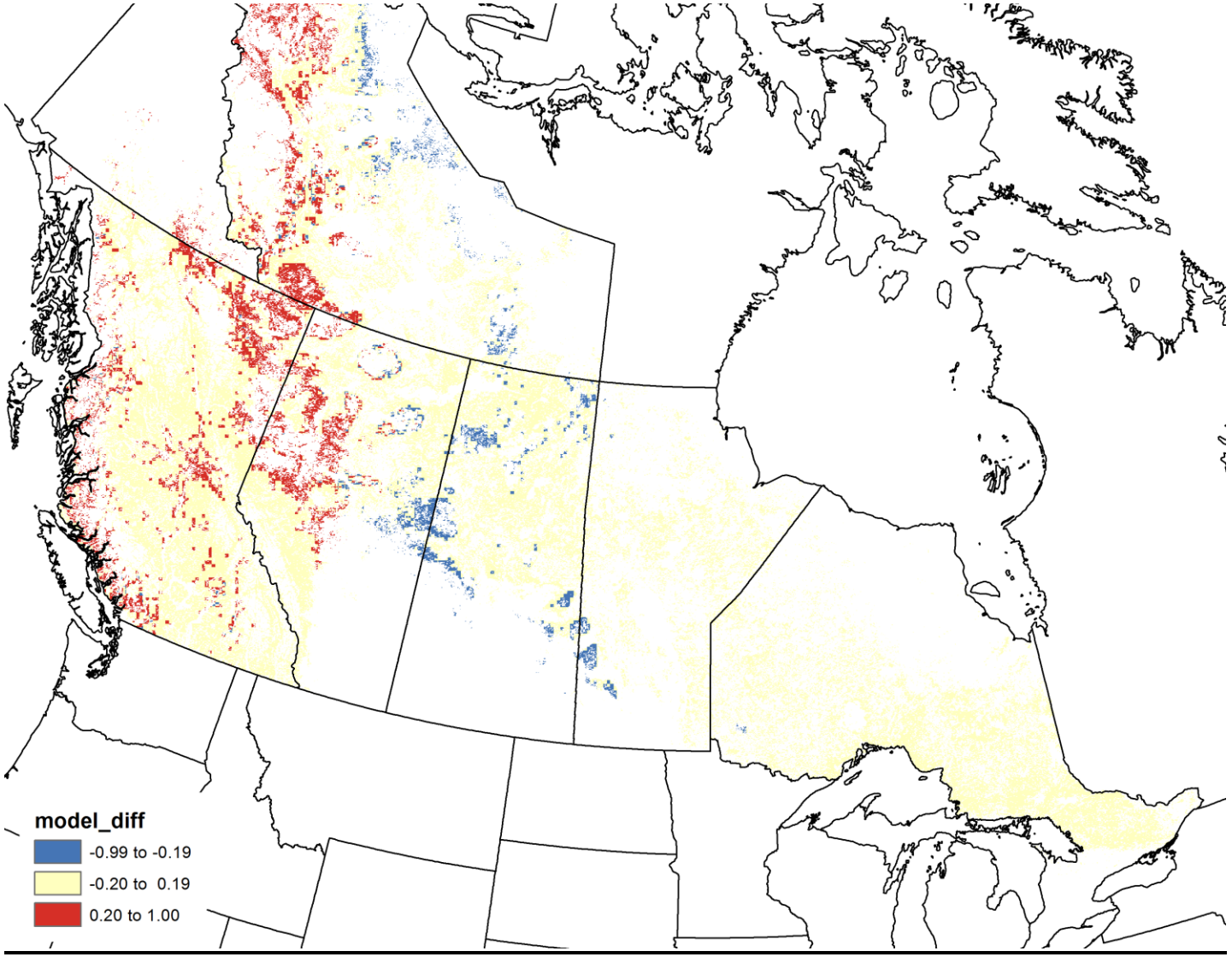


Figure 3.4 The differences between the predictions made by the Spatial and Non-Spatial models. Values indicate the arithmetic difference in Q-value prediction (Spatial minus Non-Spatial) of each model for each cell value (centroid of 2km cell). Blue and red colouration indicate regions where the predicted class differs between models.

Chapter 4: Conclusion

The overall aim of this study was to characterize the lodgepole pine x jack pine hybrid zone across the entirety of its Canadian range. In the previous chapters I have discussed both the genomic variation and spatial variation of lodgepole and jack pine and their hybrids in western Canada. First, I determined that the southern regions, and previously uncharacterized northern regions of the hybrid zone comprise a single geographically continuous hybrid zone, whose loci potentially respond differently to selective pressures. Then, I used spatial and environmental data to accurately predict pine species class across the landscape. Through improved sampling and genetic analyses, I have furthered our understanding of the hybrid zone structure, and created a new, fine-scale distribution map of lodgepole and jack pine. With this collective information the overall structure of the hybrid zone can be discerned. Previous studies concluding that the southern hybrid zone takes the form of a mosaic were limited in sampling (Cullingham *et al.* 2012), and cannot necessarily be applied to the remainder of the hybrid zone.

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 1985).

I have previously shown that genetic differences between the northern and southern regions of the hybrid zone exist in introgressed genes, and that environmental variables can be used to accurately predict hybrid and parental distributions using a spatial model. In Chapter 2 I analyzed both the combined and separated datasets to determine if previously uncharacterized regions, and the hybrid zone as a whole, take the form of a mosaic, consistent with previous assessments of the southern hybrid zone (Cullingham *et al.* 2012). Differential introgression between the northern and southern zones was observed at three genes, two involved in physiological regulation, all potentially under selection. 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).

Further, results from NewHybrids assignments opposed a tension character for the hybrid zone. In Table 2.2, first generation hybrids and backcrosses are assigned at low to negligible frequencies, while pure lodgepole or jack pines (not listed) made up the remaining ~70% of assignments. Taking into account only those individuals assigned as a hybrid class, the majority of individuals are backcrossed or late generation hybrids. This would suggest that the hybrid zone is both self-sustaining, and variable. If the lodgepole x jack pine hybrid zone was a tension

zone hybrid individuals would experience selective disadvantage (Key 1968), and we would therefore see few late generation hybrids (F2 & F3), which is not the case. 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. Indeed, the presence of steep transitions between pure species would suggest continued hybridization is likely.

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 both in the sampling-based distribution map (Figure 2.3) and the predicted distribution maps (Figures 3.2 and 3.3) where pure lodgepole and jack pine individuals can be found in multiple stands in AB and NWT within the described hybrid zone, as well as within their parental distributions. These results when combined suggest pine species are representative of a mosaic hybrid zone, rather than a clinal or a bounded superiority hybrid zone, and that this mosaic structure is consistent across the entire range of the hybrid zone. The presence of a mosaic hybrid zone may help prevent the establishment of epidemic populations of MPB in hybrid and/or jack pine stands, as the patchy distribution may limit migration (Bone *et al.* 2013) success of the MPB.

This study aimed to characterize the hybrid zone and expand upon research into the distribution of pine species, and has done so with improved sampling across the range, especially in the NWT, YT, eastern AB, SK and MB. We found that the northern and southern zones are statistically more similar than expected, which suggests that parallel selection may be acting

across the hybrid zone in western Canada. Additionally, significant differences in clines at two loci of physiological and phenotypic effect 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.

Using the improved sampling data, I have created a new species distribution map, which will allow for better prediction of the species distributions across the landscape. The last Canada-wide distribution map of pine species (Little 1971) did not account for hybrid distribution except for overlaps between presumed lodgepole and jack pine distributions. The spatial model can be used to accurately predict the species class of pine using environmental data, which can be derived from GPS data. The non-spatial model can be used to predict habitat suitable to hybridization and future pine habitats with predictive climate models. New, more accurate maps will inform both forest management and future research into the pine-MPB system.

As an agriculturally and economically important species in Canada, accurate pine species distribution data are essential to industry and government for resource allocation, reforestation of attacked stands, annual logging assessments, and for MPB control and spread prevention. As the host species for MPB, lodgepole, and now hybrid and jack pine are at risk of future MPB epidemics with climate change and MPB range expansion. The results of this study will allow for the prediction of at-risk pine stands and aid in the implementation of appropriate conservation measures to prevent the continued spread of MPB through the boreal forest. The models may be used for predictive modeling of future distributions of suitable pine habitat, which is especially important for determining MPB spread risk with climate change. Spread risk will be influenced

by both climate and habitat availability for MPB, and climate trends can be modeled along with stand location to predict MPB spread. While the mosaic hybrid zone may prevent MPB spread, we determined from genetic analysis that the southern hybrid zone is at greater risk of mediating host-transfer of MPB from lodgepole to jack pine. This risk is likely to increase with climate change and the increase in winter minimum temperatures.

Literature Cited

1. Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. *In* Proceeding of the Second International Symposium on Information Theory. *Edited by:* B.N. Petrov, F. Csaki, Akademiai Kiado, Budapest. pp. 267-281.
2. Anderson, E.C., and Thompson, E.A. 2002. A model-based method for identifying species hybrids using multilocus genetic data. Genetics **160**: 1217–1229.
3. Arango-Velez, A., Kayal, W.E., Copeland, C.C.J., Zaharia, L.I., Lusebrink, I., and Cooke, J.E.K. 2016. Differences in defence responses of *Pinus contorta* and *Pinus banksiana* to the mountain pine beetle fungal associate *Grosmannia clavigera* are affected by water deficit. Plant Cell Environ. **39**: 726-744
4. Arnold, M.L., and Hodges, S.A. 1995. Are natural hybrids fit or unfit relative to their parents? Trends Ecol. Evol. **10**(2): 67-71.
5. Bassett, M., Lindholm, J., Garza, C., Kvitek, R., and Wilson-Vandenberg, D. 2018. Lingcod (Ophiodon elongatus) habitat associations in California: implications for conservation and management. Environ. Biol. Fish. **101**(1): 2013-213.
6. Bedard, J., Trosch, R., Wu, F., Ling, Q., Florez-Perez, U., Nawaz, F., and Jarvis, P. 2017. Supressors of the chloroplast protein import mutant *tic40* reveal a genetic link between protein import and thylakoid biogenesis. Plant Cell **49:** 1726-1747
7. 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
8. Boecklen, W.J., and Howard, D.J. 1997. Genetic analysis of hybrid zones: numbers of markers and power of resolution. Ecology **78**(8): 2611-2616
9. 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
10. Bradić, M., Costa, J., and Chelo, I.M. 2011. Genotyping with sequenom. Methods Mol. Biol **772**:193–210
11. Buerkle, C.A. 2005. Maximum-likelihood estimation of a hybrid index based on molecular markers. Mol. Ecol. Notes **5:** 684-687
12. Carlson, M.R., Murphy, J.C., Berger, V.G., and Ryrie, L.F. 1999. Genetics of elevational adaptations of lodgepole pine in the interior. J. Sustain. For. **10:** 35-44
13. Cooke, B.J., and Carroll, A.L. 2017. Predicting the risk of mountain pine beetle spread to eastern pine forests: Considering uncertainty in uncertain times. For. Ecol. Manag. **396:** 11–25
14. Critchfield, W.B. 1985. The late quarternary history of lodgepole and jack pines. Can. J. Res. **15:** 740-772
15. Cudmore, T.J., Björklund, N., Carroll, A.L., and Lindgren, B.S. 2010. Climate change and range expansion of an aggressive bark beetle: evidence of higher beetle reproduction in naïve host tree populations. J. Appl. Ecol. **47**(5): 1036–1043
16. Cullingham, C.I., Cooke, J.E.K., Dang, S., Davis, C.S., Cooke, B.J., and Coltman, D.W. 2011. Mountain pine beetle host-range expansion threatens boreal forest. Mol. Ecol. **20**: 2157-2171
17. Cullingham C.I., James, P.M.A., Cooke, J.E.K., and Coltman, D.W. 2012. Characterizing the physical and genetic structure of the lodgepole pine x jack pine hybrid zone: mosaic structure and differential introgression. Evol. Appl. **5**: 879-891
18. Cullingham, C.I., Cooke, J.E.K., Dang, S., and Coltman, D.W. 2013a. A species-diagnostic SNP panel for discriminating lodgepole pine, jack pine, and their interspecific hybrids. Tree Genet. Genom. **9**: 1119-1127
19. Cullingham, C.I., Cooke, J.E.K., and Coltman, D.W. 2013b. Effects of introgression on the genetic populations structure of two ecologically and economically important conifer species lodgepole pine (*Pinus contorta* var. *latifolia*) and jack pine (*Pinus banksiana*). Genome **56**: 1-9
20. Cullingham, C.I., Cooke, J.E.K., and Coltman, D.W. 2014. Cross-species outlier detection reveals different evolutionary pressures between sister species. New Phytol. **204**: 215-229
21. Dancik, B.P., and Yeh, F.C. 1983. Allozyme variability and evolution of lodgepole pine (*Pinus contorta* var. *latifolia*) and jack pine (*P. banksiana*) in Alberta. Can. J. Genet. Cytol. **25**: 57-64
22. Dhar, A., Parrott, L., and Hawkins, C.D.B. 2016. Aftermath of mountain pine beetle outbreak in British Columbia: Stand dynamics, management response and ecosystem resilience. Forests **7**(8)**:** 171

23. Erbilgin, N., Ma, C., Whitehouse, C., Shan, B., Najar, A., and Evenden, M. 2013. Chemical similarity between historical and novel host plants promotes range and host expansion of the mountain pine beetle in a naïve host ecosystem. New Phytol. **201**(3): 940-950
24. Erbilgin, N., Ma, C., Whitehouse, C., Shan, B., Najar, A., and Evenden, M. 2014. Chemical similarity between historical and novel plant hosts promotes range and host expansion of the mountain pine beetle in a naïve host ecosystem. New Phytol. **201**: 940-950
25. Floate, K.D., and Whitham, T.G. 1993. The "hybrid bridge" hypothesis: host shifting via plant hybrid swarms. Am. Nat. **141**:651–662.
26. Fortin, M. and Dale, M.R.T. 2005. Spatial analysis: a guide for ecologists, 2nd edition. Cambridge University Press, London, UK.
27. Fox, J., and Weisberg, S. 2011. An R companion to applied regression, 2nd Edition. SAGE Publications, Thousand Oaks, CA
28. Gabriel, S., Ziaugra, L., and Tabbaa, D. 2009. SNP genotyping using the sequenom MassARRAY iPLEX platform. Curr. Protoc. Hum. Genet. **60**: 2.12.1 – 2.12.18
29. Godbout, J., Jaramillo-Correa, J.P., Beaulieu, J., and Bousquet, J. 2005. A mitochondrial DNA minisatellite reveals the postglacial history of jack pine (*Pinus banksiana*), a broad-range North American conifer. Mol. Ecol. **14**(11): 3497-3512
30. Goldberg, E., and Lande, R. 2007. Species borders and dispersal barriers. Am. Nat. **170**(2): 297-304
31. Gompert, Z., and Buerkle, C.A. 2009. A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. Mol. Ecol. **18**: 1207–1224
32. Gompert, Z., and Buerkle, C.A. 2010. INTROGRESS: a software package for mapping components of isolation in hybrids. Mol. Ecol. Resour. **10:** 378-384
33. Gompert, Z., Mandeville, E.G., and Buerkle, C.A. 2017. Analysis of population genomic data from hybrid zones. Annu. Rev. Ecol. Evol. Syst. **48**: 207-229
34. Guisan, A., and Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. Ecol. Model. **135**: 146-186
35. Harrison, R.G. 1986. Pattern and process in a narrow hybrid zone. Heredity **56**:337–349
36. Harrison, R.G., and Larson, E.L. 2014. Hybridization, introgression and the nature of species boundaries. J. Hered. **105**(1): 795-809
37. Hicks, D.J. 1980. Intrastand distribution patterns of southern appalachian cove forest herbaceous species. Am. Midl. Nat. **104**(2): 209-223
38. Hogg, E.H. 1997. Temporal scaling of moisture and the forest–grassland boundary in western Canada. Agric. For. Meteorol. **84**: 115–122
39. James, P.M.A., Murray, B.W., Hamelin, R.C., Coltman, D.W., and Sperling, F.A.H. 2011. Spatial genetic structure of a symbiotic beetle-fungal system: Toward multi-taxa integrated landscape genetics. PLoS ONE **6**(10): e25359
40. Janes, J.K., Li, Y., Keeling, C.I., Yuen, M.M.S., Boone, C.K., Cooke, J.E.K., Bohlmann, J., Huber, D.P.W., Murray, B.W., Coltman, D.W., and Sperling, F.A.H. 2014. How the mountain pine beetle (*Dendroctonus ponderosae*) breached the canadian rocky mountains. Mol. Biol. and Evol. **31**:7 1803-1815
41. Key, K.H.L. 1968. The concept of stasipatric speciation. Syst. Zool. **17**(1): 14-22
42. Kim, S.G., Lee, J.S., Shin, S., Bae, H.H., Kim, J.T., Son, B.Y., and Baek, S.B. 2016. Developing PCR-based SNP markers for distinguishing korean waxy corn F1 hybrids. Plant Breed. Biotech. **4**(3): 315-323
43. Körner, C., and Diemer, M. 1987. In situ photosynthetic responses to light, temperature and carbon dioxide in herbaceous plants from low and high altitude. Funct. Ecol. **1**: 179–194
44. Körner, C. 2007. The use of 'altitude' in ecological research. Trends Ecol. Evol. **22**(11): 569-574
45. Krehenwinkel, H., and Tautz, D. 2013. Northern range expansion of European populations of wasp spider *Argiope breunnichi* is associated with global warming-correlated genetic admixture and population-specific temperature adaptations. Mol. Ecol. **22**: 2232-2248
46. Lamesch, P., Baradini, T.Z., Li, D., Swarbeck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M., KArthikeyan, A.S., Lee, C.H., Nelson, W.D., Ploetz, L., Singh, S., Wensel, A., and Huala, E. 2012. The Arabidopsis information resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res. **40**: D1202–D1210
47. Liang, L., Hawbaker, J., Chen, Y., Zhu, Z., and Gong, P. 2014. Characterizing recent and projecting future

potential patters of mountain pine beetle outbreaks in the southern Rocky mountains. Appl. Geog. **55:** 165-175
48. 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.
49. Lusebrink, I., Erbilgin, N., and Evenden, M.L. 2013. The lodgepole x jack pine hybrid zone in Alberta, Canada: a stepping stone for the mountain pine beetle on its journey east across the boreal forest? J. Chem. Ecol. **39**: 1209-1220.
50. MacDonald, G.M., and Cwynar, L.C. 1985. A fossil pollen based reconstruction of the late quarternaryhistory of lodgepole pine (*Pinus contorta* ssp. *Latifolia*) in the western interior of Canada. Can. J. Res. **15**: 1039-1044
51. Manka, P., Kormutak, A., and Gomory, D. 2011. Deviations from the hardy-weinberg equilibrium in selected Slovak populations of *Pinus mugo* Turra, *Pinus sylvestris* L. and their putative hybrid swarms. Thaiszia J. Bot. **21**: 167-175
52. May, R.M., Endler, J.A., and McMurtrie, R.E. 1975. Gene frequency clines in the presence of selection opposed by gene flow. Am. Nat. **109**(970): 659-676
53. McLeod, T.K, and MacDonald, G.M. 1997. Postglacial range expansion and population growth of *Picea mariana*, *Picea glauca* and *Pinus banksiana* in the western interior of Canada. J. Biogeogr. **24:** 865-881
54. Meirmans, P.G., Lamothe, M., Périnet, P., and Isabel, N. 2007. Species-specific single nucleotide polymorphism markers for detecting hybridization and introgression in poplar. Can. J. Bot. **85**(11): 1082-1091
55. Mirov, N.T. 1956. Composition of turpentine of lodgepole x jack pine hybrids. Can. J. Bot. **34**: 443-457
56. Moore, W.S. 1977. An evaluation of narrow hybrid zones in vertebrates. Q. Rev. Biol. **52**(3): 263-277
57. Moss, E.H. 1949. Natural pine hybvrids in Alberta. Can. J. Res. **27**: 218-229
58. Neilsen, E.E., Arve Bach, L., and Kotlicki, P. 2006. Hybridlab (version 1.0): a program for generating simulated hybrids from population samples. Mol. Ecol. Notes **6:** 971-973
59. Oliver, M.A. 1990. Kriging: a method of interpolation for geographical information systems. Int. J.of Geogr. Inf. Syst. **4**: 313–332.
60. Peakall, R., and Smouse, P.E. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. Bioinformatics **28**(19): 2537-2539
61. Pollack, J.C., and Dancik, B.P. 1985. Monoterpene and morphological variation and hybridization of *Pinus contorta* and *P. banksiana* in Alberta. Can. J. Bot. **63**(2): 201-210
62. Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics **155:** 945-959
63. Raudnitschka, D., Hensen, I., and Oberprieler, C. 2007. Introgressive hybridization of *Senecio hercynicus* and *S. ovatus* (*compositae, senecioneae*) along an altitudinal gradient in harz national park (Germany). Syst. and Biodivers. **5**(3): 333-344
64. Raymond, M., and Rousset, F. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J. Hered. **86:** 248-249
65. Rigault, P., Boyle, B., Lepage, P., Cooke, J.E.K., Bousquet, J., and MacKay, J.J. 2011. A white spruce gene catalogue resource for conifer genome analyses. Plant Physiol. **157**: 14–28
66. Rweyongeza, D.M., Dhir, N.K., Barnhardt, L.K., Hansen, C., and Yang, R.C. 2007. Population differentiation of the lodgepole pine (*Pinus contorta*) and jack pine (*Pinus banksiana*) complex in Alberta: growth, survival, and responses to climate. Can. J. Bot. **85**(6): 545-556
67. Safranyik, L., and Carroll, A. 2006. The biology and epidemiology of the mountain pine beetle in lodgepole pine forests. *In* The Mountain Pine Beetle: A Synthesis of Biology, Management and Impacts on Lodgepole Pine. *Edited by:* L. Safranyik, and W. Wilson. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC pp. 3–66.
68. Safranyik, L., Carroll, A., Regniere, J., Langor, D.W., Riel, W.G., Shore, T.L., Peter, B., Cooke, B.J., Nealis, V.G., and Taylor, S.W. 2010. Potential for range expansion of mountain pine beetle into the boreal forest of North America. Can. Entomol. **142**(5): 415-442
69. Schweiter, J.A., Martinsen, G.D., and Whitham, T.G. 2002. Cottonwood hybrids gain fitness traits of both parents: a mechanism for their long-term persistence? Am. J. Bot. **89**(6): 981-990
70. Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T. 2005. ROCR: visualizing classifier performance in R. Bioinformatics **21**(20): 3940-3941
71. Stoeckel, S., Grange, J., Fernandez-Maniarres, J.F., Bilger, I., Frascaria-Lacoste, N., and Mariette, S. 2006. Heterozygote excess in a self-incompatible and partially clonal forest tree species – *Prunus avium* L. Mol. Ecol. **15**(8): 2109-2118

72. Stukenbrock, E.H. 2016. The role of hybridization in the evolution and emergence of new fungal plant pathogens. Phytopathology **106**(2):104-112
73. Sykes, M.T., Prentice, I.C., and Cramer, W. 1996. A bioclimatic model for the potential distributions of north European tree species under present and future climates. J. Biogeogr. **23**:203–233
74. Taylor, S.W., and Carroll, A.L. 2003. Disturbance, forest age, and mountain pine beetle outbreak dynamics in BC: a historical perspective. *In* Mountain Pine Beetle Symposium: Challenges and Solutions, Kelowna, British Columbia, October 30-31, 2003. *Edited by*: T.L. Shore, J.E. Brooks, and J.E. Stone, Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC, pp. 41-51
75. Vercruyssen, L., Verkest, A., Gonzalez, N., Heyndrickx, K.S., Eeckhout, D., Han, S.K., Jégu, T., Archacki, R., Van Leene, J., Andriankaja, M., De Bodt, S., Abeel, T., Coppens, F., Dhondt, S., De Milde, L., Vermeersch, M., Maleux, K., Gevaert, K., Jerzmanowski, A., Benhamed, M., Wagner, D., Vandepoele, K., De Jaeger, G., and Inzé, D. 2014. ANGUSTIFOLIA3 binds to SWI/SNF chromatin remodeling complexes to regulate transcription during Arabidopsis leaf development. Plant Cell. **26**(1): 210-29.
76. Walsh, J., Rowe, R.J., Olsen, B.J., Shriver, W.G., and Kovach, A.I. 2016. Genotype-environment associations support a mosaic hybrid zone between two tidal marsh birds. Ecol. Evol. **6**(1): 279–294
77. Wang, X.R., and Szmidt, A.E. 1994 Hybridization and chloroplast DNA variation in a *Pinus* species complex from Asia. Evolution **48**: 1020–1031
78. Wang, T., Hamann, A., Spittlehouse, D.L., and Carroll, C. 2016. Locally downscaled and spatially customizable climate data for historical and future periods for North America. *PLoS One* **11**: e0156720
79. Wheeler, N.C., and Guries, R.P. 1987. A quantitative measure of introgression between lodgepole and jack pines. Can. J. Bot. **65:** 1876-1885.
80. Williams, W., Friedman, J., Gaskin, J., and Norton, A. 2014. Hybridization of an invasive shrub affects tolerance and resistance to defoliation by a biological control agent. Evol. Appl. **7**(3): 381–393
81. Wood, L., Hartley, I., and Watson, P. 2009. Determining hybridization in jack pine and lodgepole pine from british Columbia. Wood Fiber Sci. **41**(4): 386-395
82. Yang, R.C., Yeh, F.C., and Ye, T.Z. 2007. Multilocus structure in the *Pinus contorta – Pinus banksiana* complex. Can. J. Bot. **85**:774–784.
83. Yeatman, C.W. and Teich, A.H. 1969. Genetics and breeding of jack and lodgepole pines in Canada. For. Chron. **45**: 428-433
84. Yemshanov, D., McKenney, D.W., and Pedlar, J.H. 2012. Mapping forest composition from the Canadian National Forest Inventory and land cover classification maps. Environ. Monit. Assess.**184**: 4655–4669. doi:10.1007/s10661-011-2293-2
85. Zavarin, E., Critchfield, W.B., and Snajberk, K. 1969. Turpentine composition of *Pinus contorta* × *Pinus banksiana* hybrids and hybrid derivatives. Can. J. Bot. **47**(9): 1443-1453
86. Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., and Smith, G. 2009. Mixed effects models and extensions in ecology with R. Springer-Verlag, New York, NY

Appendix

Table A1 Summary of samples included in each chapter. Stands are the individual sampling sites in each province or territory.

	Chapter 2		Chapter 3	
Location	Stands	Individuals	Stands	Individuals
British Columbia	19	254	19	246
Alberta	24	416	24	393
Saskatchewan	4	78	4	44
Manitoba	1	10	1	10
Ontario	4	46	4	46
Yukon Territory	5	39	5	38
Northwest Territory	5	96	9	151
Total	62	939	66	928

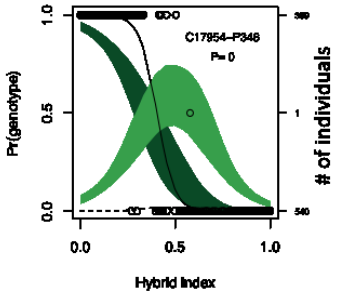
Table A2 Original variance inflation factor (VIF) values used for model validation to reduce correlation between predictor variables. VIFs are from the original logistic regression containing all predictor variables.

Predictor	VIF
Latitude	2.1814
Longitude	2.4675
Elevation	2.8546
Drought	2.1131
MAT	5029.1
MWMT	1276.6
MCMT	9848.8
TD	13683
MAP	477.28
MSP	180.78
AHM	38.576
SHM	41.035
DD0	2172.1
DD5	454.47
DD18	6847.2
NFFD	37.001
bFFP	547.08
eFFP	282.66
FFP	1303.6
PAS	217.36
EMT	35.183
EXT	38.816
Eref	61.727
CMD	71.726
RH	17.893

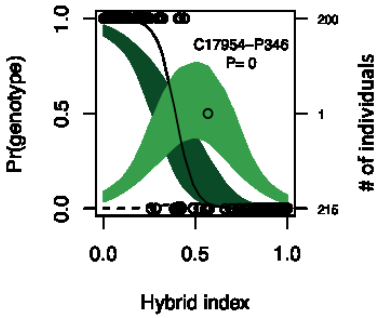
Figure A1 Genomic clines resulting from multinomial regressions for all 29 loci. Each locus has three clines in a single row, first for all samples ("All"), second for samples above 56°N latitude ("High"), and third for samples below 56°N latitude ("Low"). Thick bands represent the 95% confidence interval of the homozygous and heterozygous (dark and light respectively) genotypes, while the solid and dashed line denote the genomic clines of the same. Open circles are the individual samples included in the regression. Numbers on the right axis are the numbers of individuals included for each genotype.

C17954-P346

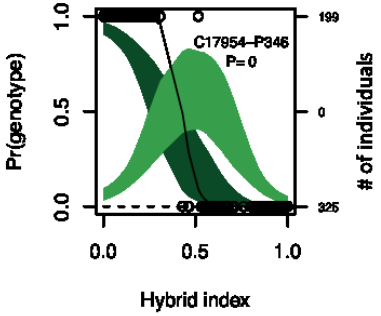
All



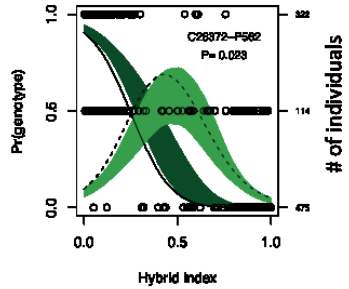
High

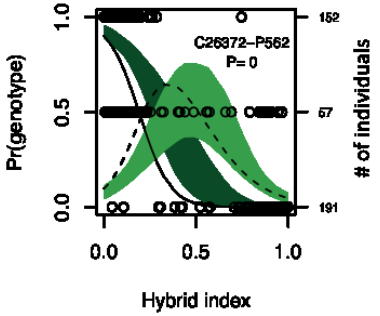


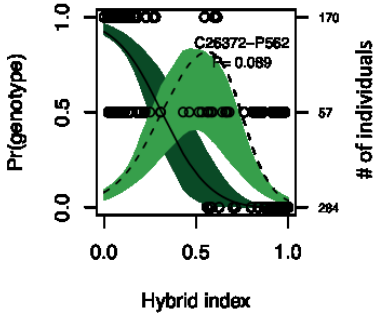
Low



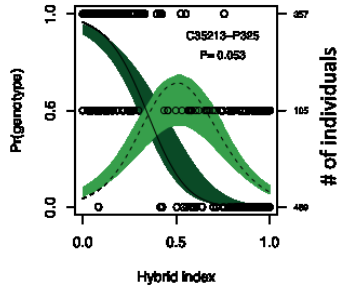
C26372-P562

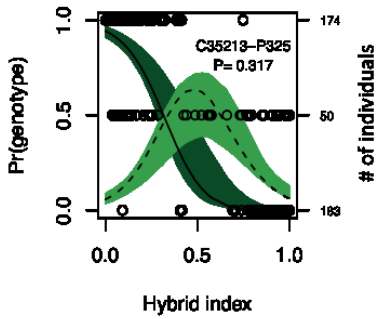


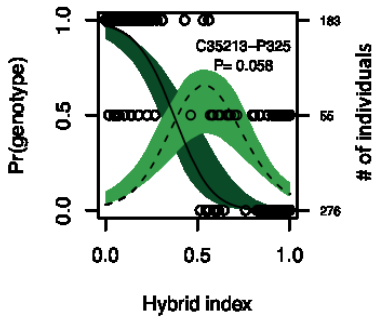




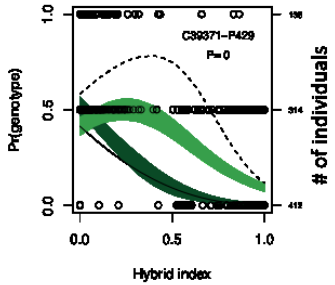
C35213-P325



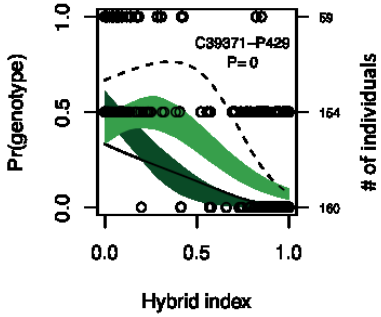




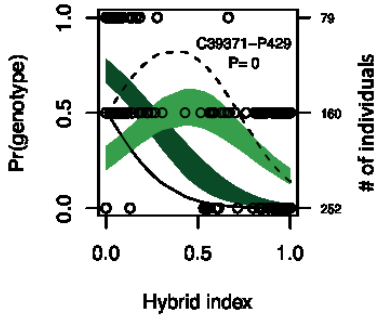
C39371-P429 All



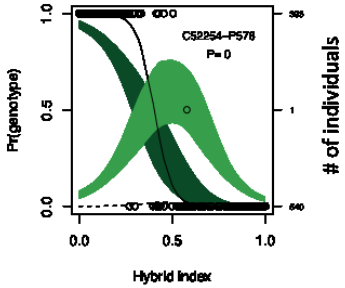
High

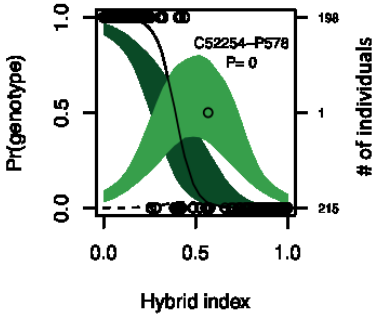


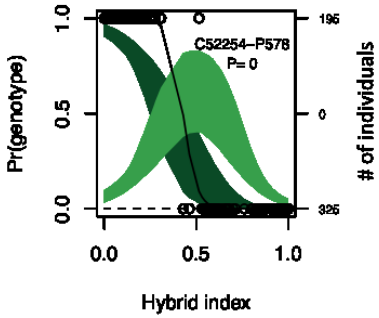
Low



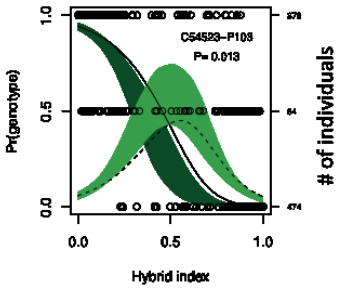
C52254-P578

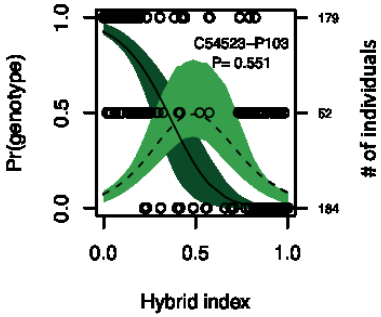


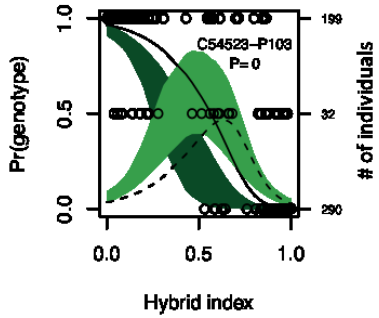




C54523-P103

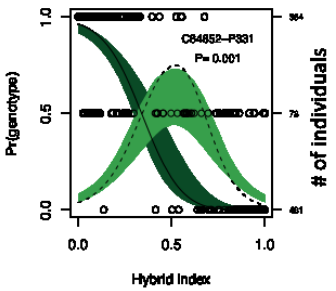




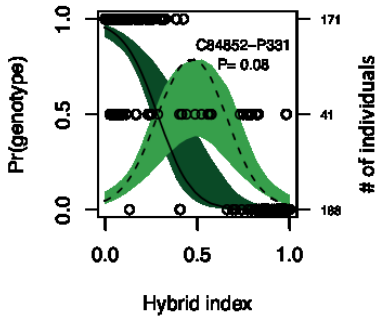


C84852-P331

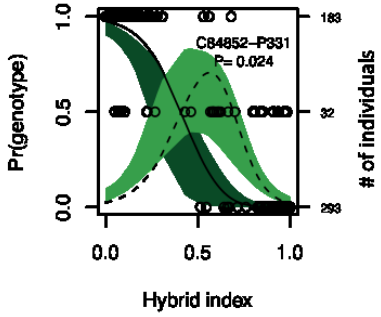
All



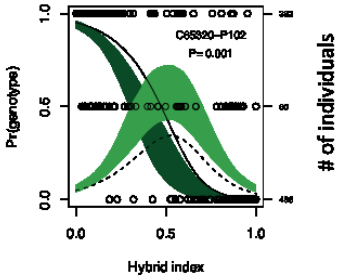
High

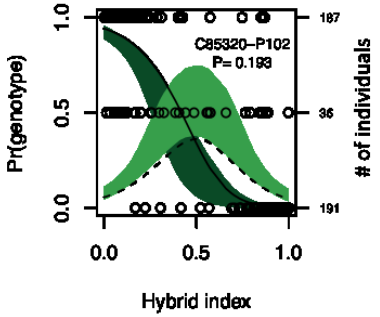


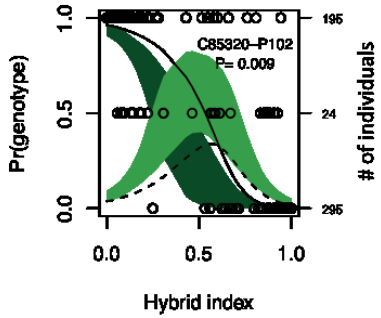
Low



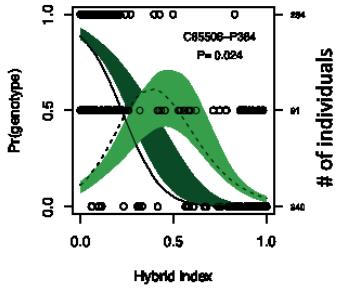
C85320-P102

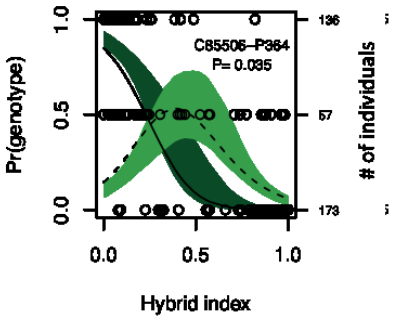


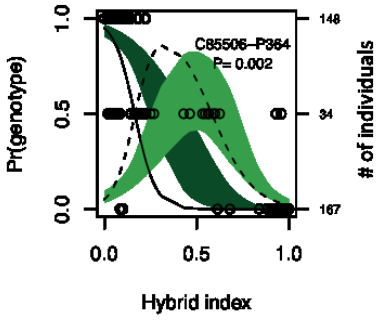




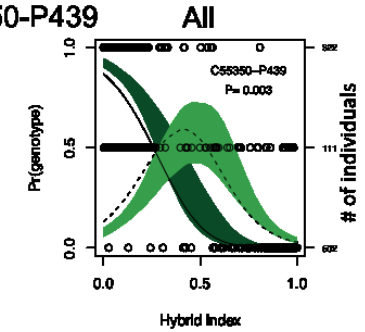
C85506-P364

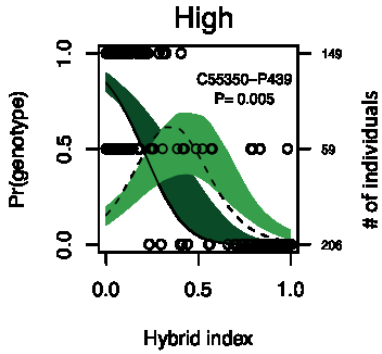


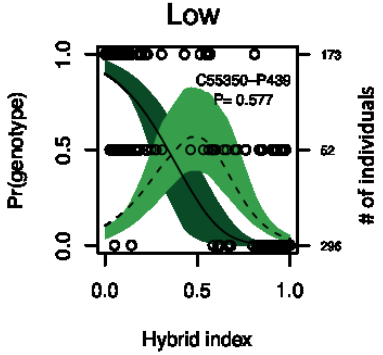




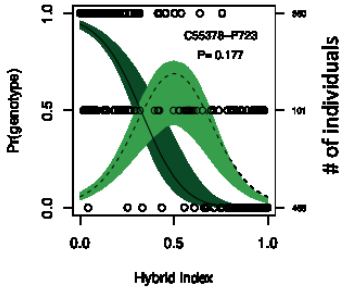
C55350-P439

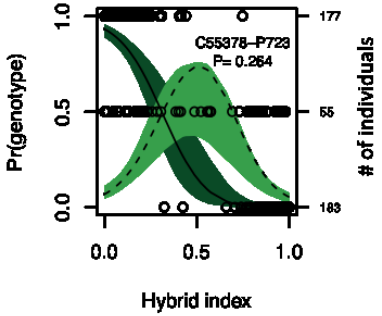


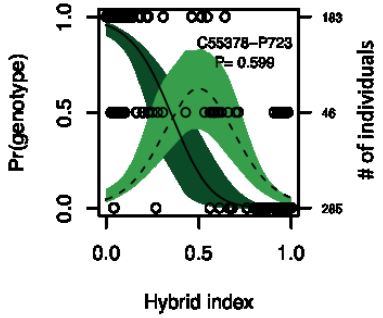




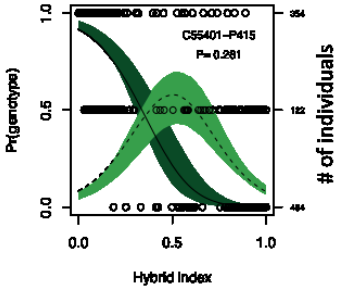
C55378-P723

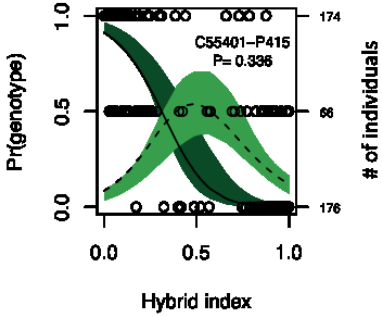


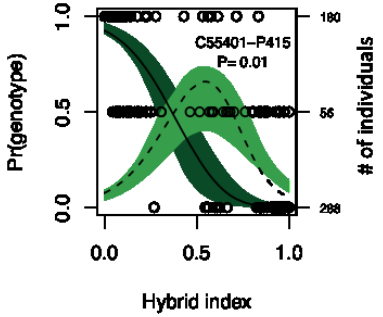




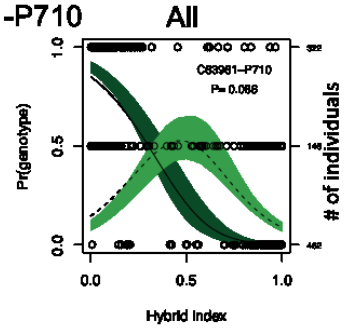
C55401-P415

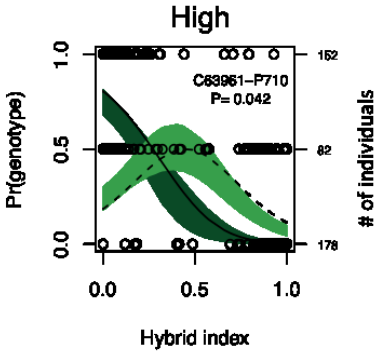


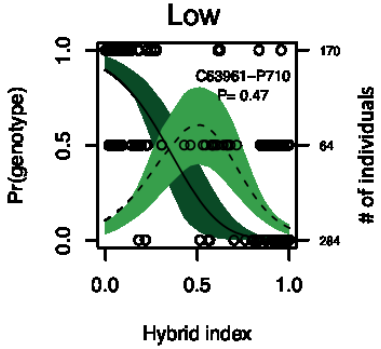




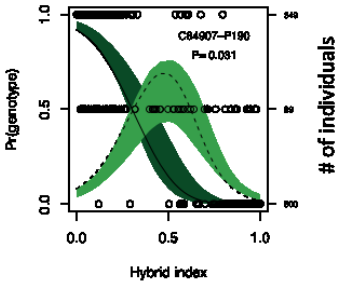
C63961-P710

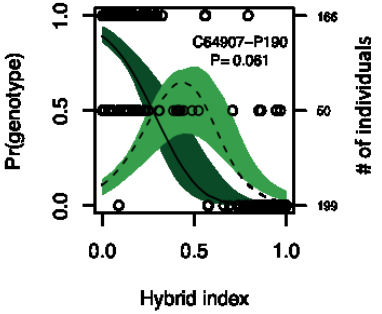


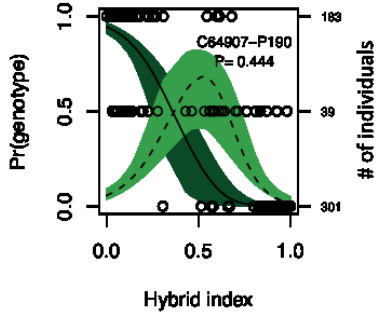




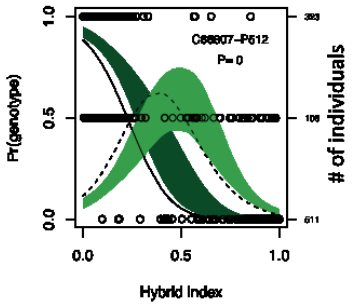
C64907-P190

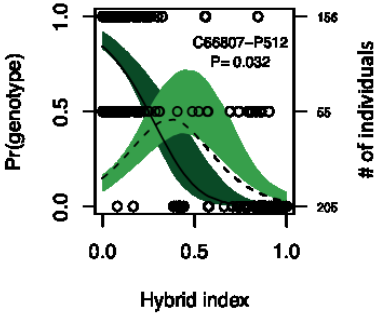


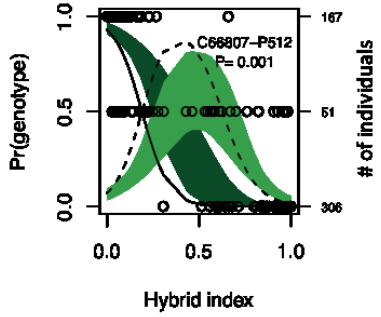




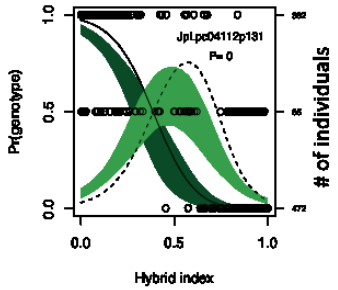
C66807-P512



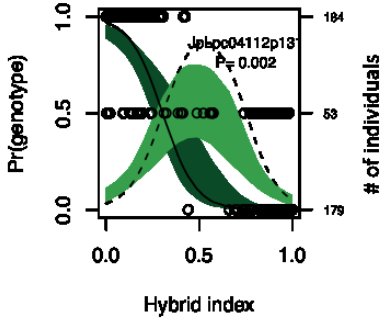




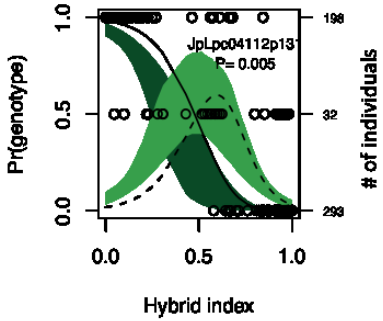
JpLpc04112p131 All



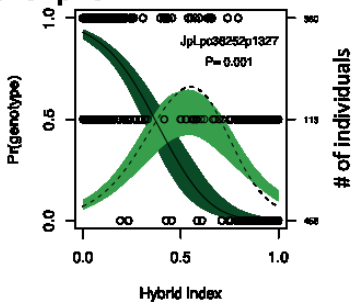
High

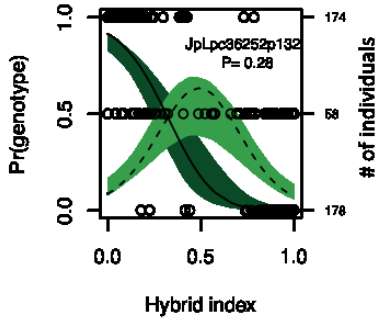


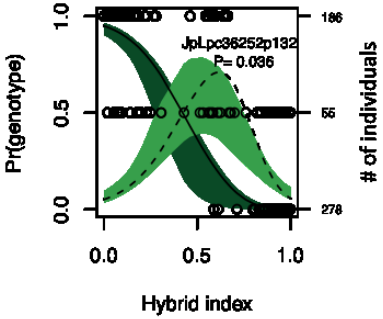
Low



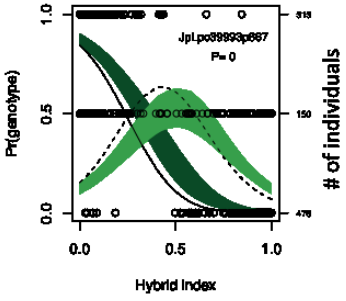
JpLpc36252p1327

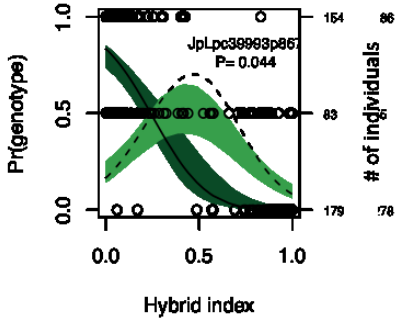


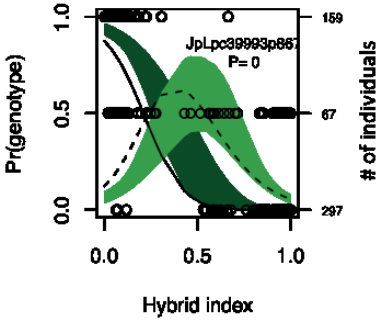




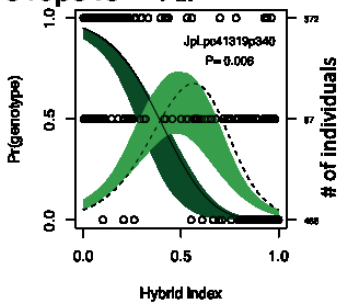
JpLpc39993p867



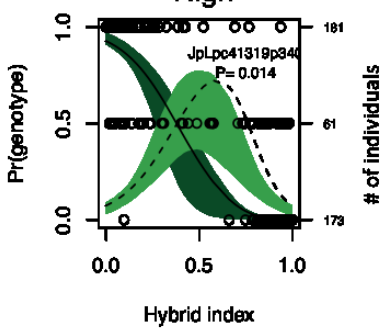




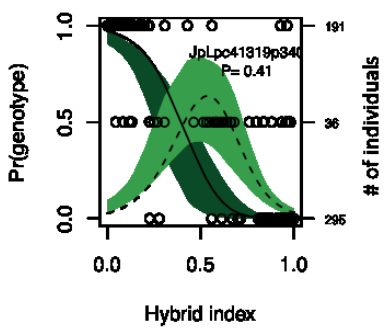
JpLpc41319p340 All



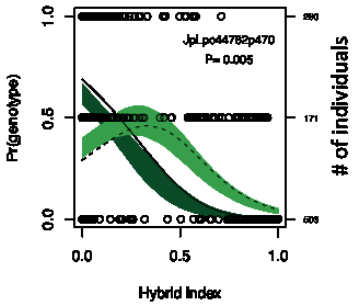
High

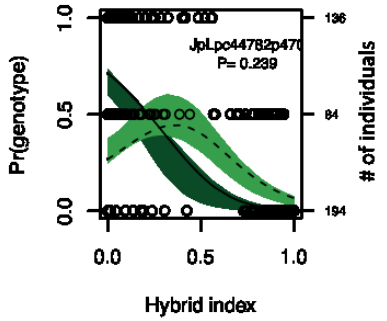


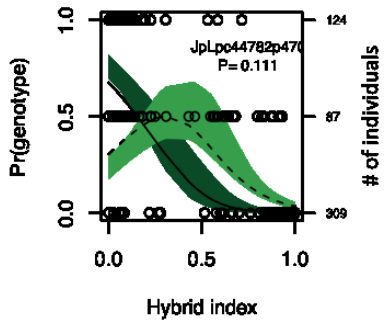
Low



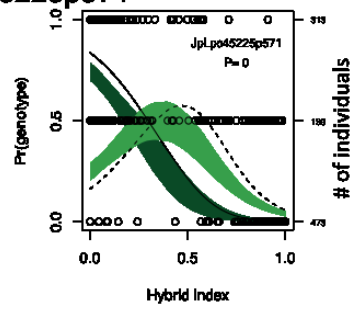
JpLpc44782p470

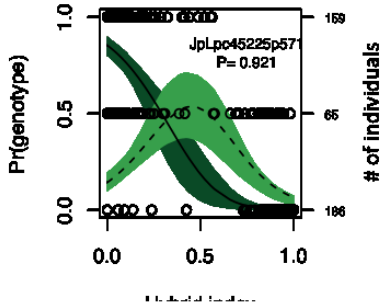


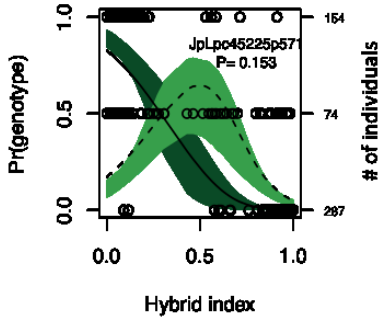




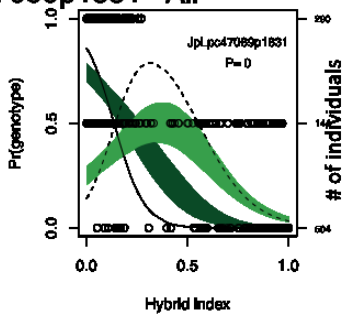
JpLpc45225p571



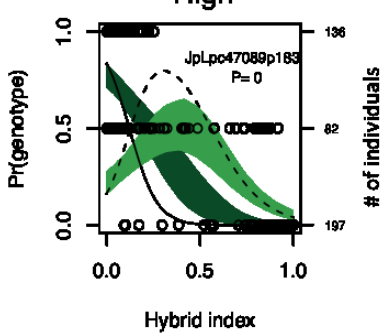




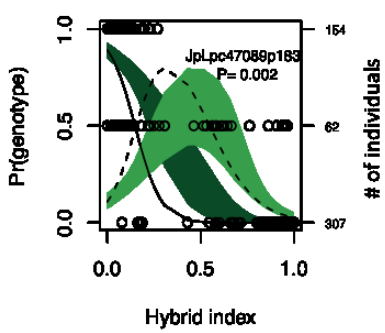
JpLpc47089p1831 All



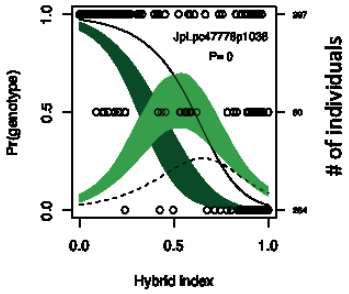
High

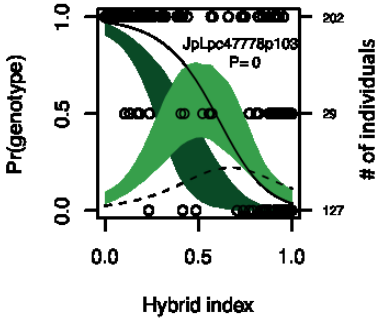


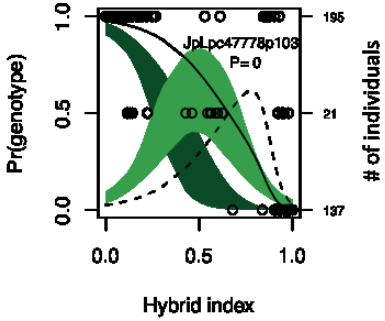
Low



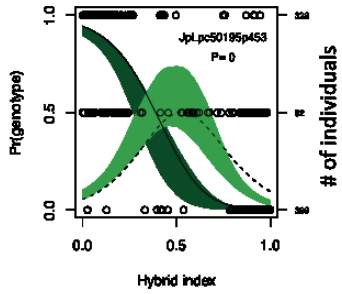
JpLpc47778p1036

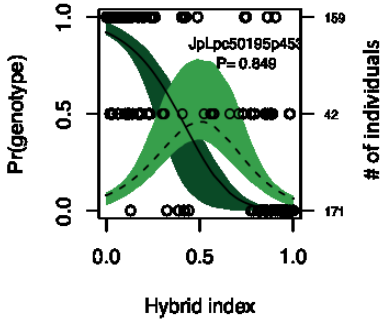


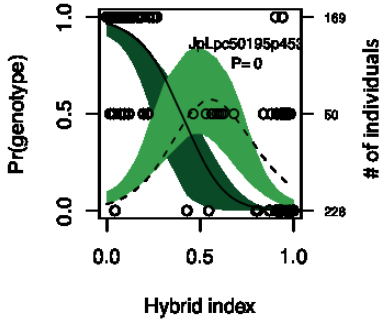




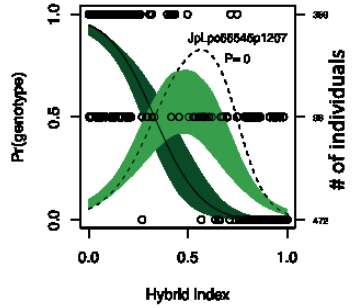
JpLpc50195p453



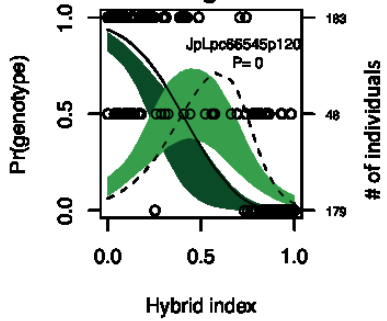




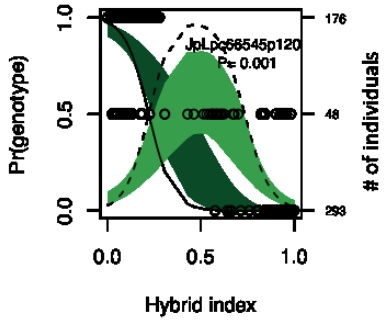
JpLpc66545p1207 All



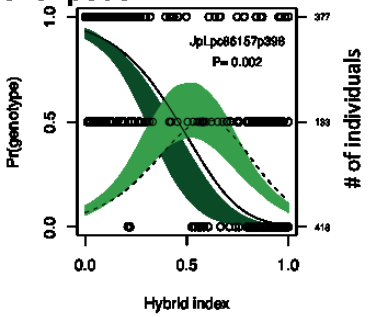
High

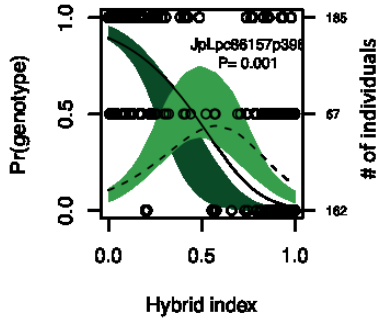


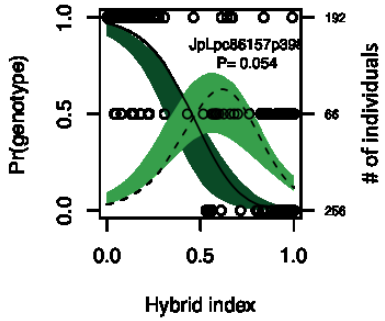
Low



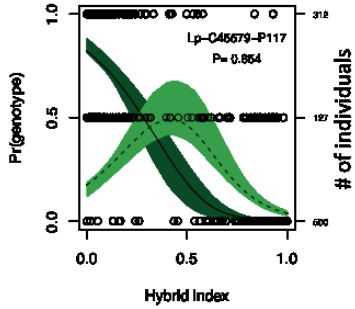
JpLpc86157p398

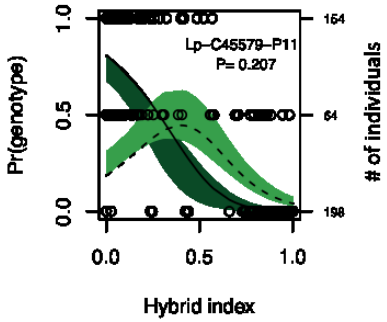


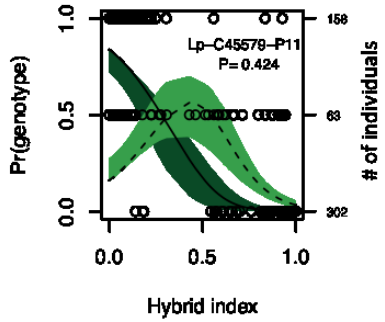




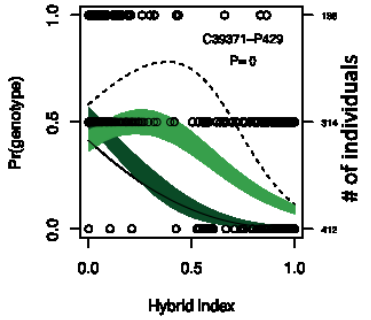
Lp-C45579-P117

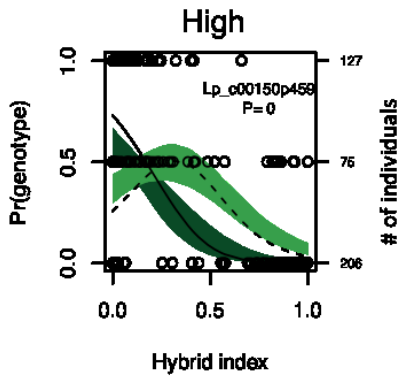


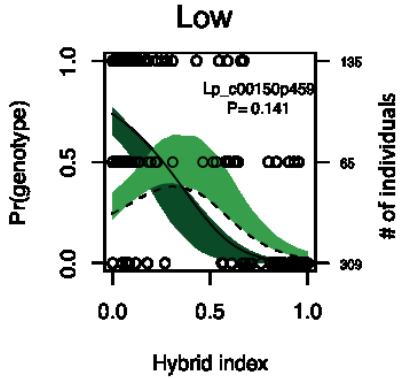
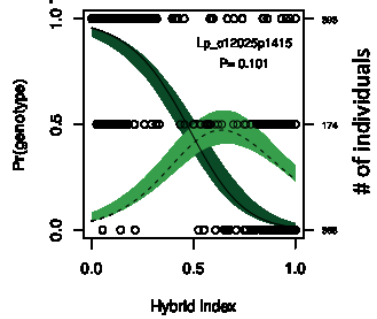
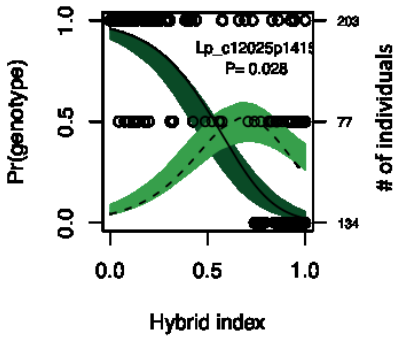
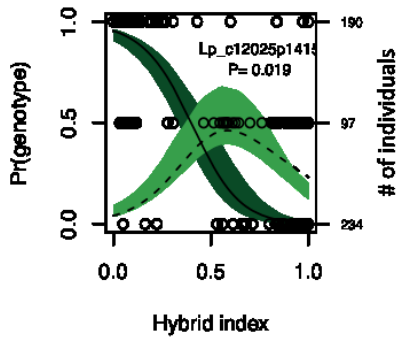




C39371-P429







Lp_c12025p1415