Genetic architecture underlying response to the fungal pathogen *Dothistroma septosporum* in lodgepole pine, jack pine, and their hybrids

Mengmeng Lu[1,2], Nicolas Feau[3,4], Brandon Lind[3], Dragana Obreht Vidakovic[3], Pooja Singh[1,5,6]

Sally N. Aitken[3], Richard C. Hamelin[3,7], and Sam Yeaman[1]

[1] Department of Biological Sciences, University of Calgary, 507 Campus Drive NW, Calgary, AB T2N 1N4, Canada

[2] Department of Biological Sciences, University of Notre Dame, 100 Galvin Life Sciences, Notre Dame, IN 46556, USA

[3] Department of Forest and Conservation Sciences, University of British Columbia, 3041-2424 Main Mall, Vancouver, BC V6T 1Z4, Canada

[4] Canadian Forest Service, Pacific Forestry Centre, Natural Resources Canada, Victoria, BC V8Z 1M5, Canada

[5] Aquatic Ecology & Evolution Division, Institute of Ecology and Evolution, University of Bern, Bern CH-3012, Switzerland

[6] Center for Ecology, Evolution & Biogeochemistry, Swiss Federal Institute of Aquatic Science and Technology (EAWAG), Kastanienbaum CH-6047, Switzerland

[7] Institut de Biologie Intégrative et des Systèmes, Université Laval, Pavillon Charles-Eugène-Marchand 1030, avenue de la Médecine, Québec, Québec G1V 0A6, Canada

Correspondance

Mengmeng Lu, Department of Biological Sciences, University of Calgary, 507 Campus Drive NW, Calgary, AB T2N 1N4, Canada

Email : mengmeng.lu@ucalgary.ca

Funding information

This research was supported by the CoAdapTree Project (241REF), with funding from Genome Canada, Genome BC, Genome Alberta, Genome Québec, the BC Ministry of Forests, Lands, Natural Resource Operations and Rural Development, and many other sponsors http://adaptree.forestry.ubc.ca/sponsors/, http://coadaptree.forestry.ubc.ca/sponsors/, and seed

contributors http://adaptree.forestry.ubc.ca/seed-contributors/. Dr. Sam Yeaman is also funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) and Alberta Innovates.

Abstract

In recent decades, *Dothistroma* needle blight (DNB), a pine tree disease caused by the fungal pathogen *Dothistroma septosporum,* has severely damaged lodgepole pine (*Pinus contorta* Dougl. ex Loud. var. *latifolia*) in British Columbia, Canada. The pathogen has already shown signs of host shift eastward to lodgepole × jack hybrid pine populations, and possibly into pure jack pine (*Pinus banksiana* Lamb.). However, we have little knowledge about mechanisms of resistance to *D. septosporum*, especially the underlying genetic basis in pines. In this study, we conducted controlled inoculations to induce infection by *D. septosporum* and performed a genome-wide case-control association study with pooled sequencing (pool-seq) data to dissect the genetic architecture underlying response in lodgepole pine, jack pine, and their hybrids. We identified candidate genes associated with *D. septosporum* response in lodgepole pine and in lodgepole × jack hybrid pine populations. We also assessed genetic structure in hybrid populations and inferred that introgression may affect the distribution of genetic variation involved in *D. septosporum* response in the study populations. These results can be used to develop genomic tools to evaluate DNB risk, guide forest management strategies, and potentially select for resistant genotypes.

KEY WORDS
Conifer, fungal pathogen resistance, pool-GWAS, divergent selection, introgression, hybrid zone

1 | INTRODUCTION

Lodgepole pine (*Pinus contorta* Dougl. ex Loud. var. *latifolia*) mainly inhabits montane regions of western North America (Figure 1), throughout the interior of British Columbia and the Rocky Mountains of Alberta (Lotan & Critchfield, 1990). Jack pine (*Pinus banksiana* Lamb.) is widely distributed across boreal forests in eastern North America, extending in Canada from Northwest Territories to Nova Scotia (Rudolph & Laidly, 1990). Lodgepole and jack pine are closely related species, both belonging to the genus *Pinus*, with an estimated divergence time of ~8 million years before present (Hao et al., 2015). Lodgepole and jack pine evolved in allopatry but hybridize in areas of contact, primarily in central and northwestern Alberta (Burns et al., 2019; Rudolph & Yeatman, 1982). Following secondary contact of these two species, additional patchy regions of introgression have developed (Cullingham et al., 2012). Natural hybrids display a wide range of phenotypes that are intermediate or more typical of one parent, depending on the level of introgression.

Pests and pathogens constantly threaten forest trees, and their impacts are changing as climates are altered. *Dothistroma* needle blight (DNB), which is caused by the fungal pathogen *Dothistroma septosporum,* has caused increased damage to lodgepole pine in recent decades in Western Canada (Dale et al., 2011; Woods, 2003). While *D. septosporum* tends to be more common in moist western forests closer to the Pacific Ocean, in 2012 and 2013, *Dothistroma*-like symptoms were observed in drier lodgepole and lodgepole × jack hybrid pines in northern Alberta, with further tests confirming *D. septosporum* was the cause of the disease observed (Ramsfield et al., 2021). A controlled inoculation experiment demonstrated that lodgepole pine, jack pine, and their hybrids are susceptible to *D. septosporum*, with jack pine showing a higher disease severity than lodgepole or hybrid pine. This implies that *D. septosporum* has expanded its host range to spread eastward into Alberta (Feau et al., 2021). Although there is no report showing DNB infection in natural jack pine forests, their higher susceptibility to *D. septosporum* raises serious concern. Dissecting the genetic architecture underlying responses to *D. septosporum* in these pines is therefore critical to understanding and managing the impact of this pathogen.

Using RNA-seq analysis on experimentally infected seedlings, Lu et al. (2021) identified 43 differentially expressed genes from tolerant samples, which were present in the plant-pathogen interaction pathway, as well as nine candidate resistance (R) genes that contained sites under positive selection, indicating both constitutive and induced defense have developed in lodgepole pine. However, the genome-wide genetic architecture underlying fungal pathogen response and the evolutionary forces that likely impact the gene structure are largely unknown, which presents a hurdle to the development of tools for forest management.

Understanding the genetic architecture of pathogen resistance and identifying candidate R genes constitute a relevant approach to improve tree resilience to forest pathogens. This could help genetic selection by screening populations for resistant genotypes. Knowledge about genes and alleles that confer disease resistance or tolerance can eventually be used to develop genomic tools for rapidly selecting resistant genotypes (Isabel et al., 2019). Current breeding practices for forest trees involve recurrent selection, which includes repeated cycles of selection, breeding, and testing, to search for genotypes with better disease resistance, and deployment of these selected tree genotypes (White, 2004). This takes years of effort to design and implement. Increasingly, genome editing is used to develop crops resistant to disease (van Esse et al., 2020). Given the long generation time in conifers and the frequent threats from diseases and insects, identifying R genes that can be targeted for selection or editing is a useful step in developing resistant genotypes for reforestation.

To understand the genetic basis of disease response in tree species, candidate genes have been identified using genetic markers, mainly SNPs, and methods like quantitative trait locus (QTL) mapping and genome-wide association analysis (GWAS). Polygenic disease responses, comprising numerous loci of small effect, appear to be the rule. Large numbers of SNPs were associated with the pitch canker resistance in loblolly pine (De La Torre et al., 2019; Lu et al., 2017; Quesada et al., 2010), with the white pine blister rust resistance in sugar pine (Weiss et al., 2020), and with the *Heterobasidion* root rot resistance in Norway spruce (Capador-Barreto et al., 2021).

Most GWAS studies on forest trees use individual-based methods, with both phenotypic and genotypic data collected on each sampled individual. Case-control methods have been developed to study disease where individuals are pooled into healthy vs. infected groups, but such approaches typically still use individual-based sequencing to call genotypes before pooling (Wu et al., 2010). When multiple replicates are available, it is possible to combine pool-seq and the case-control design: Endler et al. (2016) analyzed the genetic differences underlying abdominal pigmentation variation among *Drosophila* populations, while Stocks et al. (2019) identified SNPs associated with low versus high ash dieback damage in European ash tree (*Fraxinus excelsior*) using this approach. In a study on conifer genotyping, Lind et al. (2022) found that allele frequencies estimated from pooled DNA sequencing (pool-seq) samples were highly correlated with frequencies estimated from individual sequencing. Recently, Singh et al. (2024) used a case-control pool-seq approach similar to that deployed here to search for signatures of resistance to Swiss needle cast in Douglas-fir.

The objectives of our study were to: 1) develop a GWAS case control approach using pool-seq samples to identify loci with consistent associations with *D. septosporum* response across replicates; and 2) assess genetic structure in hybrid trees and infer introgression that may affect the genetic basis of *D. septosporum* responses. This study provides a first step towards identifying candidate genes to develop tools for screening trees resistant to DNB infection.

2 | METHODS

2.1 | Plant materials

Seeds were obtained from 40 natural populations across Western Canada (Figure 1), including 25 lodgepole pine populations from British Columbia (BC_pl) and three from Alberta (AB_pl), seven lodgepole × jack hybrid pine populations from Alberta (AB_px), and five jack pine populations from Alberta (AB_pj). When the seeds were collected in the wild, they were assigned to pure lodgepole pine, pure jack pine, or lodgepole × jack hybrid pine, based on their location and tree morphology. Cullingham et al. (2011) used microsatellite markers to identify the ancestry of the collected seeds. Seedlings were grown for their first year in a greenhouse at UBC, Vancouver, BC.

Resistance was phenotyped by inoculating 100 individuals per population with one of two *D. septosporum* isolates (D1 & D2, 50 individuals of each isolate). D1 was isolated from an infected lodgepole × jack hybrid pine in Alberta, Canada, while D2 was isolated from infected lodgepole pine in the Kispiox Valley region close to Smithers in Northwestern British Columbia, Canada (Capron et al., 2021). The controlled inoculation experiment was performed as described in Feau et al. (2021). Briefly, the one-year old seedlings were placed in a completely randomized experimental design in a growth chamber with the condition of 16 hours daylight at 20 ℃, 8 hours a night at 12 ℃, and minimum relative humidity of 80 %. *D. septosporum* conidia were harvested from colonies grown for 10~16 days on *Dothistroma* sporulation medium plates and then suspended in sterile distilled water. Each seedling was sprayed twice at 10-day interval with a standard inoculum of approximately 3 mL of 1.6×10^6 conidia mL^{-1}, using a household trigger action atomizer. Once the needles were dry (~45 minutes), the seedlings were wrapped in

transparent plastic bags and kept in the growth chamber for 48 h. A mister spraying tap water was activated hourly for 3 minutes during the trial to maintain needle wetness. Control seedlings were inoculated with sterile water.

Twelve weeks after inoculation, seedlings were rated for proportion of necrotic needles with red bands and/or fruiting bodies after 15 weeks since the beginning of the experiment according to a disease severity scale of 1 to 5 (1 = least, 5 = most disease severity). Following disease rating, the ten most- and ten least-infected individuals per population (five populations for each *D. septosporum* isolate) were identified and retained for DNA extraction. Genomic DNA of each of these individuals was extracted as described in Lind et al. (2022), using the Nucleospin 96 Plant II Core kit (Macherey–Nagel GmbH & Co. KG, Germany) on an Eppendorf epMotion 5075 liquid-handling platform. The DNA of five most-infected and five least-infected individuals by each of the two *D. septosporum* isolates were combined in equimolar amounts to compose four pooled libraries per population, two susceptible (D1-S, D2-S) and two resistant libraries (D1-R, D2-R). Hence,160 pooled DNA libraries were generated.

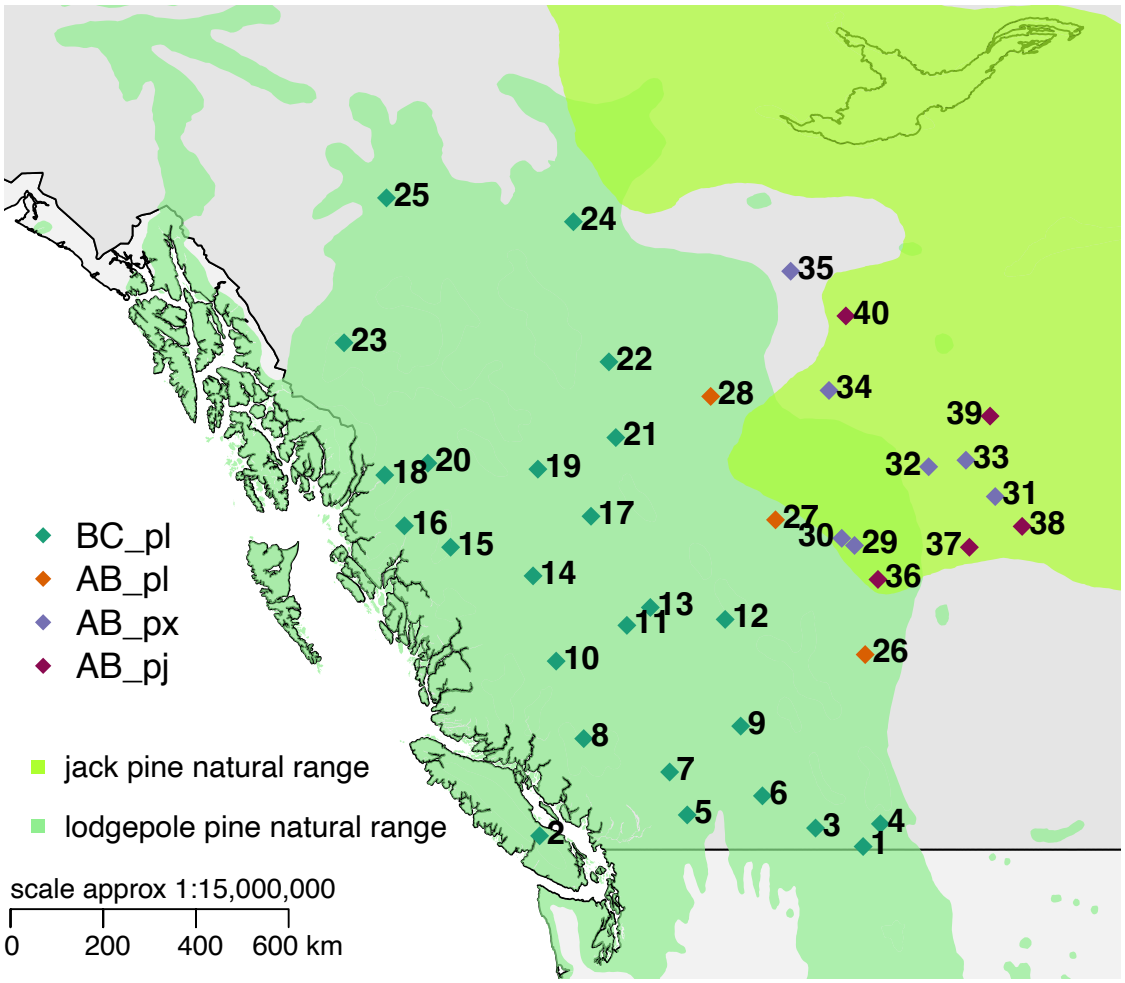


FIGURE 1 Sampling locations of the 40 pine populations analyzed in this study, plotted within the natural range of lodgepole and jack pine in Canada and USA. BC_pl (green diamonds) —

lodgepole pine populations from British Columbia (N=25); AB_pl (orange diamonds) — lodgepole pine populations from Alberta (N=3); AB_px (purple diamonds) — lodgepole × jack hybrid pine populations from Alberta (N=7); AB_pj (brown diamonds) — jack pine populations from Alberta (N=5).

2.2 | Sequence capture and pool-seq genotyping

The capture probes were designed based on an existing lodgepole pine sequence capture array (Suren et al., 2016) by removing the probes that did not yield successful genotyping in Yeaman et al. (2016) and adding new probes derived from newly assembled *D. septosporum*-induced genes. Probe sequences of the existing capture array were aligned to the reference genome using GMAP v2019-03-15 (Wu & Watanabe, 2005). Probes that covered the genomic regions with called SNPs in the dataset from Yeaman et al. (2016) were retained, otherwise the probes were discarded. Since there is no available lodgepole pine reference genome, a masked loblolly pine reference genome Pita.2_01.masked3k2.fa (https://treegenesdb.org/FTP/Genomes/Pita/v2.01/genome/) (Neale et al., 2014) was used instead. New probes targeting *D. septosporum*-induced genes were designed based on a lodgepole pine reference transcriptome assembled using the RNA-seq data of *D. septosporum* infected lodgepole pine samples (Lu et al., 2021). To avoid duplicates, only those *D. septosporum*-induced genes with low homology to the retained probe sequences were used to design the new probes. To do so, the retained probe sequences were aligned to this transcriptome using blastn v2.9.0 with an E-value of 1e-10. A total of 8,778 *D. septosporum*-induced genes did not have any aligned probe sequences. These non-duplicate *D. septosporum*-induced genes were subsequently aligned to the reference genome to predict the exon-intron boundaries using GMAP v2019-03-15. Exon sequences from these induced genes with a length of at least 100 bp were combined with the previously designed working probe sequences, and this combined set of sequences were submitted to Roche NimbleGen (Roche Sequencing Solutions, Inc., CA USA) for Custom SeqCap EZ probe design (design name: 180321_lodgepole_v2_EZ). This updated lodgepole pine sequence capture array has a capture space of 44 Mbp, containing roughly 35,467 assembled genes.

The capture libraries for each of the 160 pools were constructed following NimbleGen SeqCap EZ Library SR User's Guide and as described in Lind et al. (2022). Then the R and S libraries (two libraries per capture, R1+S1 or R2+S2, indexed with different barcodes) per population and per isolate were combined for sequence capture and enrichment. Sequencing was performed using the Illumina NovaSeq 6000 S4 PE 150 platform in Centre d'expertise et de services Génome Québec. Our in-house pool-seq pipeline (Lind, 2021b) was employed to align the reads to the reference genome and call SNPs. For raw SNPs, only bi-allelic loci in regions without annotated repetitive elements or potentially paralogous genes were retained. The annotated repetitive elements were acquired from the loblolly pine genome annotation (Wegrzyn et al., 2014). The potentially paralogous genes were identified as described in Lind et al. (2022) using haploid megagametophyte sequences, for the heterozygous SNP calls for haploid sequences are likely to represent misalignments of paralogs. Afterwards, the SNP loci with depth (DP) <10, DP >400, global minor allele frequency <0.05, or >25% missing data were also removed.

2.3 | Genetic structure analyses

The genetic structure among the 160 pooled samples, which represented four pooled libraries (D1-R, D1-S, D2-R, D2-S) for each of the 40 populations, was detected by using principal component analysis (PCA), reconstructing phylogenetic tree, and by assessing correlation of allele frequencies and population differentiation (F_{ST}) between samples using unlinked SNPs. To reduce the impact of linkage disequilibrium on estimation, the SNP set was thinned using vcftools (Danecek et al., 2011), so that no two sites were within 100,000 bp. PCA and genetic distance were calculated using the major allele depth and the R package "adegenet" (Jombart, 2008). An unrooted phylogenetic tree was constructed using the distance output and the Neighbour-Joining algorithm, which was implemented by the R package "ape" (Paradis & Schliep, 2019). Correlation coefficients of major allele frequencies between pairwise samples were calculated and plotted using the R package "corrplot" (Wei & Simko, 2017). F_{ST} was estimated between pairwise samples using the R package "poolfstat" (Hivert et al., 2018).

2.4 | Dissection of genetic architecture underlying *D. septosporum* response

The alleles and candidate genes associated with *D. septosporum* response were detected using the pool-GWAS method. A GWAS case control approach was developed using pool-seq samples to identify allele frequency differences between susceptible (case) and resistant (control) pines inoculated with *D. septosporum*. This method was based on Cochran-Mantel-Haenszel (CMH) method and implemented in our in-house pipeline (Lind, 2021a). While applications of this test commonly use sequencing read depth as an estimation of allele count (Futschik & Schlötterer, 2010; Schlötterer et al., 2014), this will only be accurate when the haploid size of the sample is much greater than the depth of coverage (so that most reads arise from uniquely sampled haplotypes). As the pool size used in the present study (diploid size of 2N=10) is often smaller than the sequencing depth of coverage, read counts were converted into allele frequencies and then multiplied by the ploidy to yield an approximate count for each pool that represents the real replication level. We conducted a simulation to compare the false positive rate from the CMH test using uncorrected or corrected allele counts across a range of number of demes, individuals, and depths. Results showed that the false positive rate of the CMH method is high, especially when the sequencing depth is much higher than the pool size (Figure S1a). Using corrected allele counts, the allele frequencies multiplied by the ploidy (N=10) tends to lower the false positive rate (Figure S1b). Thus, the corrected allele counts were used for CMH test.

As the genetic structure analysis showed that lodgepole × jack hybrid pine and jack pine populations tend to cluster, jack pine populations and lodgepole × jack hybrid pine populations were combined as jack-hybrid populations for downstream analyses. Pool-GWAS was performed separately within lodgepole and jack-hybrid pine populations. R and S samples inoculated with D1 or D2 were contrasted jointly and separately. To identify candidate genes, SNPs in the top 1% of $-\log_{10}(p)$ values from pool-GWAS test were classified as outliers. As linkage disequilibrium can amplify signatures of selection, we identified candidate resistance genes as those with a large number of outlier SNPs relative to the average genome-wide expectation, as represented by an index based on the binomial distribution as per Yeaman et al. (2016). The identified loci on the enriched genes were realigned to the 12 lodgepole pine linkage groups (MacLachlan et al., 2021) and Manhattan plots were drawn. The lodgepole pine

transcriptome (Lu et al., 2021) was used to annotate the reference genome. The aligned transcripts to the reference genome had a minimum alignment identity of 90 % and a minimum alignment coverage of 85 % using GMAP v2019-03-15.

2.5 | Linkage disequilibrium (LD) between loci

A non-linear model (Hill & Weir, 1988) was used to estimate the decay of LD with physical distance. Pairwise correlation (r^2) of allele frequencies between loci on the same scaffold were calculated. The r^2 values and the physical distances between loci on all scaffolds (genome-wide) or on scaffolds containing the identified top candidate genes (significant) were used to fit the non-linear model as described by Marroni et al. (2011). The LD decay graph was plotted using R (R Core Team, 2018).

2.6 | Introgression

To study patterns of introgression affecting different regions of the genome and among different populations, the mean pairwise F_{ST} was calculated for each locus using the R package "poolfstat". Loci with high F_{ST} values have high genetic differentiation, which may represent species barriers or genomic regions under divergent selection, while loci with low F_{ST} values imply introgression. The F_{ST} values were calculated between the five pure jack pine populations (p36 to p40) and six pure lodgepole populations. The F_{ST} values between jack and lodgepole pine were averaged to estimate the genetic similarity of studied jack pine and lodgepole pine. To evaluate whether regions associated with *D. septosporum* response had atypical patterns of introgression, we compared the F_{ST} values of the top 1% *D. septosporum* response outliers identified from lodgepole pine (N = 3,358) and those from jack-hybrid pine (N = 3,041), with F_{ST} values of randomly drawn SNPs (N = 3,400).

Similarly, F_{ST} between the jack-hybrid populations (p29 to p40) and six pure lodgepole populations were averaged to estimate the genetic similarity of studied jack-hybrid and lodgepole pine. Different genomic regions were represented by different SNP sets, including *D. septosporum* response outliers identified from lodgepole pine (N = 3,358) and outliers identified from jack-hybrid pine (N = 3,041), as well as the unlinked SNP set (N = 31,716), which was used for detecting genetic structure and representing random genomic region. To detect the regression pattern of F_{ST} with longitude, the genetic similarity of jack-hybrid and lodgepole pine was regressed to the longitude of these jack-hybrid populations. The regression slopes between the three SNPs sets were compared using analysis of covariance method implemented by the R package "lsmeans" (Lenth, 2016; R Core Team, 2018).

3 | RESULTS

3.1 | Patterns of population structure

After filtering, a total of 364,691 SNP loci were retained for downstream analyses, and most loci had a MAF between 0.05 and 0.1 (Figure S2). A thinned set of 31,716 unlinked SNPs was used for studying genetic structure among the 160 pooled samples (Figure 2). The most prominent patterns of structure were associated with species (Figure 2a & b), with lodgepole pine samples

readily distinguishable from jack and hybrid pine samples. Hybrid samples clustered more closely with jack than lodgepole pine. Compared to those sampled from the eastern hybrid zone, western hybrid pines tend to be more similar to lodgepole pine (samples were arranged according to longitude from left to right in Figure 2 c & d). Given this clear pattern of population structure, we conducted GWAS in lodgepole pine and jack-hybrid pine populations separately.

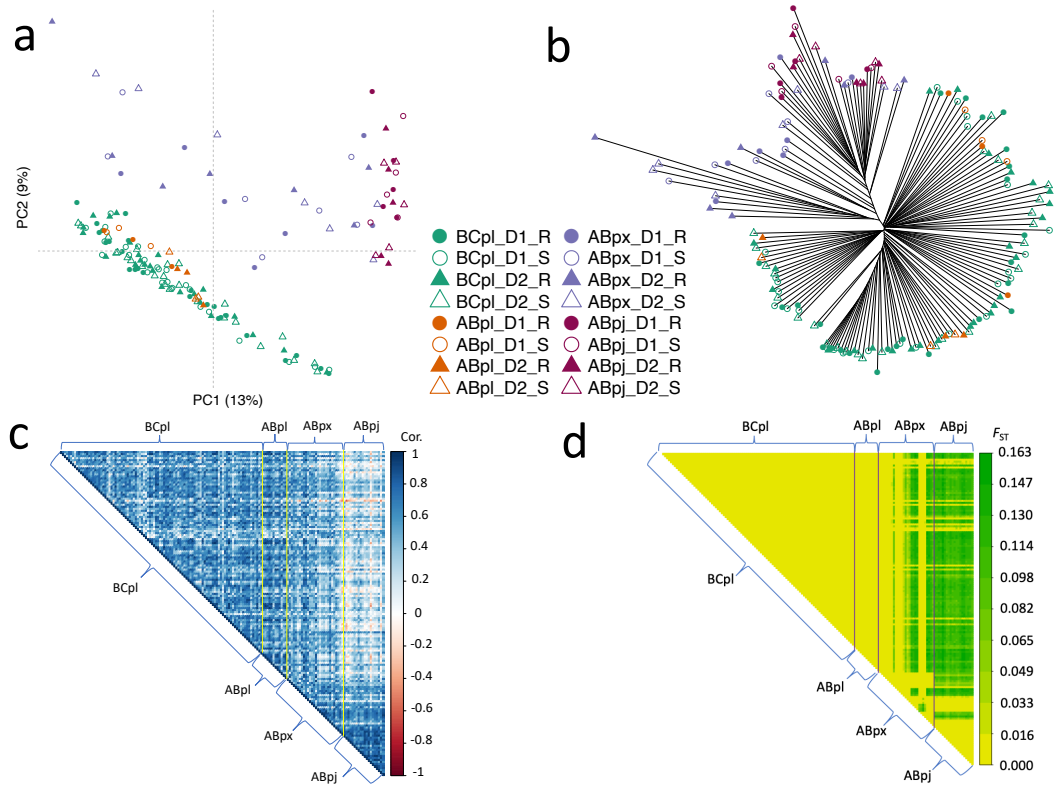


FIGURE 2 Genetic structure among the 160 pooled samples. a) Principal component analysis. b) Unrooted phylogenetic tree. c) Correlation plot of the major allele frequencies between samples. d) Pairwise F_{ST} values between pools. BC_pl — lodgepole pine populations from British Columbia; AB_pl — lodgepole pine populations from Alberta; AB_px — lodgepole × jack hybrid pine populations from Alberta; AB_pj — jack pine populations from Alberta. D1 — *D. septosporum* isolate 1; D2 — *D. septosporum* isolate 2. R — resistant tree; S — susceptible tree. In c) and d), each column represents a pooled sample. The samples were arranged from left to right according to the longitude of sampling location, and four pooled libraries per population (D1-R, D1-S, D2-R, D2-S) were arranged together.

3.2 | Signatures of association to infection response

As the significance threshold is quite a stringent limit, we explored whether some genes had more SNPs with extreme *p*-values than expected by chance. Genes with an enrichment of outliers were identified using a binomial approach. Using this index to rank candidate genes, out of the 100 top-ranked candidate genes in each test (Table S1), three were identified within both the lodgepole and jack-hybrid pine tests. This is significantly more than the hypergeometric

356 expectation, based on the probability of overlap for two draws of 100 genes from 20,026 genes.
357 A previous study anchored 10,093 scaffolds on the lodgepole pine linkage map (MacLachlan et
358 al., 2021), so we plotted SNPs and highlighted some enriched plant-pathogen interaction
359 candidate genes on the 12 linkage groups (Figure 3). The three genes identified in both species
360 encode ABC transporter, F-box Kelch-repeat protein, and serine/threonine protein kinase (F-box
361 protein and serine protein kinase were highlighted in Figure 3, ABC transporter genes were not
362 included in the linkage groups, so they are not included in the figure). We also checked the
363 enrichment of differentially expressed genes (DEGs) in the top 100 ranked genes. Out of 631
364 DEGs identified in the previous study (Lu et al., 2021), four were also candidate genes in the
365 present study. Fisher's Exact test showed there is no significant enrichment of DEGs in the
366 candidate genes.
367

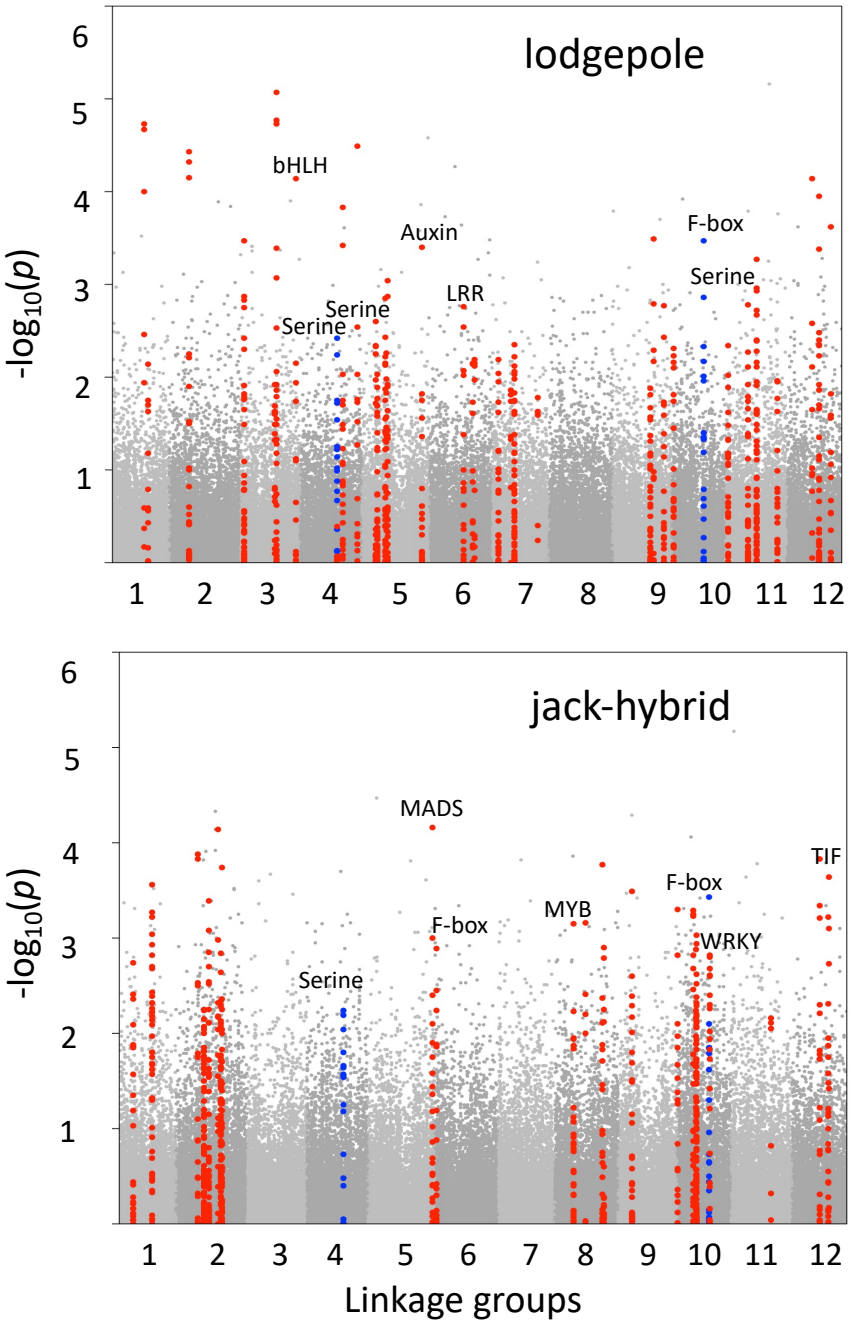
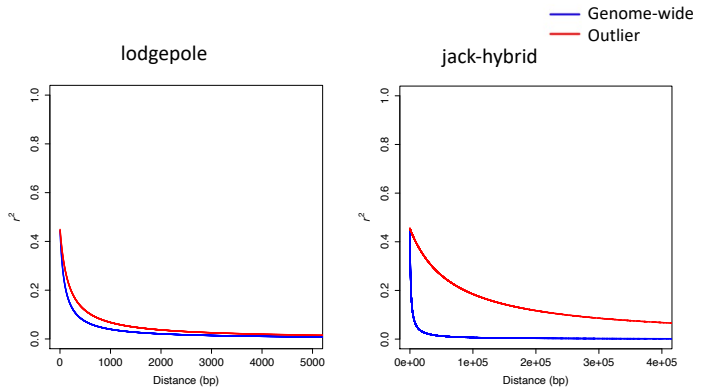


FIGURE 3 Manhattan plot showing SNPs associated with *D. septosporum* response within lodgepole and jack-hybrid pine populations. The x axis represents 12 linkage groups. The y axis represents significance of associations identified from pool-GWAS. Candidate genes were identified using a binomial test for enrichment of outlier SNPs per gene. Candidate genes and their surrounding 50kbp regions are highlighted in red. Of the three top-ranked candidate genes identified within both the lodgepole and jack-hybrid pine populations, one was not present in the linkage map, while the other two are highlighted in blue ("Serine" on linkage group 4 and "F-box" on linkage group 10). Labels in the graphs represent genes encoding: bHLH — transcription factors with basic helix-loop-helix domain; Serine — serine/threonine-protein

379 kinase; Auxin — auxin response factor; LRR — nucleotide-binding site leucine-rich repeat
380 protein; F-box — F-box Kelch-repeat protein; MADS — transcription factor with MADS-box
381 domain; MYB — genes encoding MYB transcription factor; WRKY — transcription factor
382 WRKY; TIF — translation initiation factor.
383
384
385 Loci on scaffolds with outliers (the SNPs with top 1 % of -$\log_{10}(p)$ values from pool-GWAS test)
386 tend to exhibit a more gradual decay in LD compared with the genome-wide loci on all scaffolds
387 (Figure 4), with jack-hybrid pine outliers having a much slower decay rate than lodgepole pine.
388 In lodgepole pine, the distance at which LD decays to half of its maximum value is 96 bp for
389 genome-wide loci, and 174 bp for outlier loci. In jack-hybrid pine, the half LD decay distance is
390 1,357 bp for genome-wide loci, and 68,215 bp for outlier loci.



391
392
393 FIGURE 4 Linkage disequilibrium (LD) decay for genome-wide loci and outlier loci identified
394 from lodgepole and jack-hybrid pine populations. The r^2 values (pairwise correlation of allele
395 frequencies between loci) and the physical distances between loci were used to fit the non-linear
396 model as described by Marroni et al. (2011).
397
398
399 **3.3 | Introgression**
400 By calculating F_{ST} between pairs of jack pine and lodgepole populations, we found that the *D.*
401 *septosporum* response outliers, which were identified separately in lodgepole or jack-hybrid
402 populations, have higher F_{ST} values than randomly drawn SNPs (Wilcoxon test p-value < 0.01,
403 Figure 5a & 5b), with jack-hybrid outliers showing higher F_{ST} values than lodgepole outliers
404 (Wilcoxon test p-value < 0.01). These results suggest that the outliers associated with disease
405 response constitute or are linked to genomic regions for divergent selection or species barriers.
406 F_{ST} estimates between jack-hybrid and pure lodgepole pine populations showed a pattern of
407 increased differentiation with distance eastwards (Figure 5c). Yet the differences between
408 regression slopes on different gene sets (lodgepole outliers, jack-hybrid outliers, and unlinked
409 SNPs) were not significant (p-values of pairwise comparisons > 0.3).
410
411

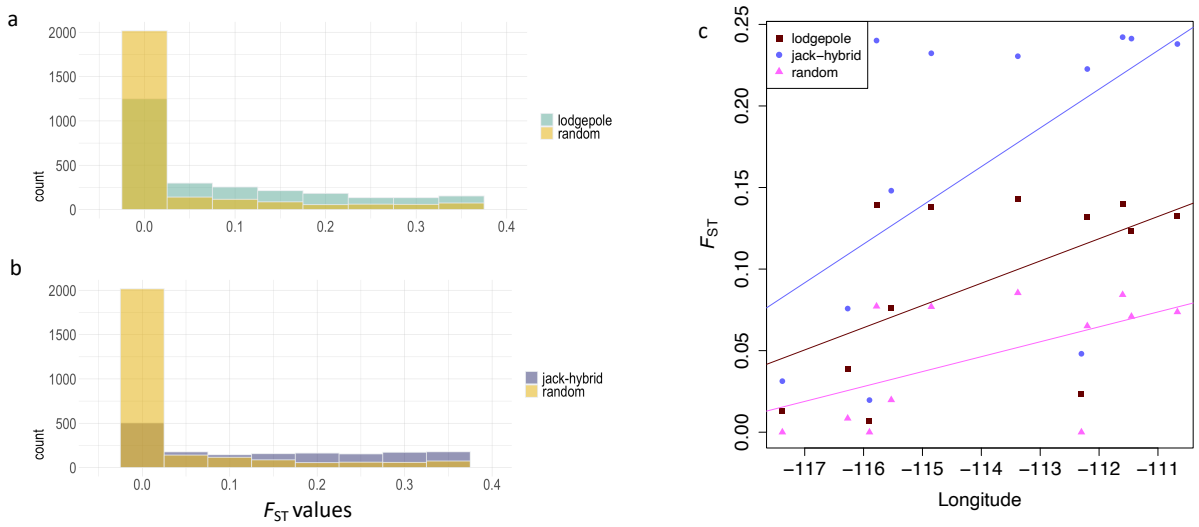


FIGURE 5 F_{ST} values of *D. septosporum* response outliers (a & b) and regression patterns of F_{ST} values with longitude for comparisons between each of the jack-hybrid populations and a group of pure lodgepole pine populations in different SNP sets (c). For a & b, F_{ST} values of *D. septosporum* response outliers were identified from lodgepole (N=3,358) and jack-hybrid (N=3,041) populations, and randomly drawn SNPs (N=3,400). The F_{ST} values were calculated between the jack pine populations (p36 to p40) and six pure lodgepole populations. Outliers identified from both species have higher F_{ST} values than randomly drawn SNPs (p-value < 0.01). For c, The F_{ST} values were calculated for different SNP sets, including lodgepole outliers (N=3,358, brown), jack-hybrid outliers (N=3,041, slateblue), and unlinked SNPs (N=31,716, pink). Differences between slopes are not significant (ANCOVA, p-value > 0.3).

4 | DISCUSSION

In recent decades, the fungal pathogen *D. septosporum* has become an increasing threat to lodgepole pine in BC (Feau et al., 2021; Welsh et al., 2009). Although *Dothistroma* needle blight (DNB) has not been observed in natural stands of jack pine in AB, Ramsfield et al. (2021) confirmed jack pine's susceptibility to infection by *D. septosporum* using a trial under natural conditions. The potential of climate change to allow this pathogen to spread across the wide natural distribution of jack pine raises serious concerns. The hybrid zone of contact between the sister species lodgepole and jack pine may act as a bridge, allowing this fungal pathogen host shift (Taylor et al., 2015). The genetic basis of pathogen resistance in each species is therefore an important factor affecting the host shift. To study the genetics of resistance to DNB, we characterized the genetic architecture underlying *D. septosporum* response in lodgepole pine, jack pine, and their hybrids. We identified candidate genes associated with *D. septosporum* response and inferred the impact of divergent selection and introgression on the genetic structure underlying *D. septosporum* response in the studied populations. We detected largely non-overlapping sets of candidate genes in lodgepole and jack-hybrid pine in response to *D. septosporum*, which might be shaped by different evolutionary forces of selection and introgression. Alternatively, the non-overlapping sets of candidate genes might reflect that many non-causal genes exist in the top 1 % GWAS outliers of each species, given that we did not test

445 for significance at the gene-level. In either case, it seems clear that resistance to DNB is a highly
446 polygenic trait with alleles of small effect.
447
448 In hybrid zones, alleles in some genomic regions are able to introgress across species boundaries,
449 whereas alleles that constitute species barriers or are under divergent selection will remain
450 differentiated (Harrison & Larson, 2014). Since we analyzed pool-seq data, it is not possible to
451 analyze parental origin information, thus we inferred the ancestry using clustering. The
452 clustering patterns in Figure 2 show that most hybrid pine trees in our study had greater jack pine
453 than lodgepole pine ancestry. Western hybrid pines tend to be more genetically similar to
454 lodgepole pine, while eastern hybrid pines tend to be more genetically dissimilar to lodgepole
455 pine. This is a typical characteristic for ongoing introgression and hybridization during range
456 expansion in the contact zone. Such observed pattern differs from previous studies that reported
457 a mosaic of hybridization and introgression pattern in the lodgepole pine × jack pine hybrid zone
458 (Burns et al., 2019; Cullingham et al., 2012). Ecologically-based divergent selection arises when
459 species have divergent habitat preferences (Rundle & Nosil, 2005). Additionally, Cullingham et
460 al. (2011) suggested differential genetic basis of local adaptation identified in lodgepole and jack
461 pine. It seems that divergent selection operates in the studied lodgepole and jack-hybrid pine
462 population, which may affect the frequency of disease-resistant alleles within these species via
463 introgression.
464
465 We found significantly high F_{ST} values between jack and lodgepole pine populations for the
466 candidate genes associated with *D. septosporum* infection response. This can occur if these
467 candidate genes tend to reside in genomic regions with restricted gene flow, driven by linkage
468 with genes for divergent selection or species barriers. Bechsgaard et al. (2017) showed that R
469 genes can adaptively introgress between closely related species. Since lodgepole pine has a long
470 coevolutionary history with *D. septosporum* (Capron et al., 2021; Welsh et al., 2009), the
471 introgression of lodgepole pine genes into hybrids may increase resistance to *D. septosporum*.
472 However, variation in F_{ST} with longitude (Figure 5c) did not differ significantly between
473 background and candidate loci, so we cannot draw strong conclusions about adaptive
474 introgression here.
475
476 This study is the first report of genome-wide scanning of *D. septosporum* response genes in
477 lodgepole pine, jack pine, and their hybrids. We identified lodgepole pine candidate genes
478 involved in the *D. septosporum* response, but these genes require cloning and characterization to
479 confirm the function. Plants rely on two strategies, PAMP-triggered immunity (PTI) and
480 effector-triggered immunity (ETI), to detect pathogens (Dodds & Rathjen, 2010). PTI and ETI
481 share signaling pathways and reinforce each other (Chang et al., 2022). When attacked by
482 infectious pathogens, PTI responds as basal resistance, followed by ETI. Pathogens can deliver
483 effector proteins into host cell to suppress PTI, and plant nucleotide-binding leucine-rich repeat
484 (NB-LRR) receptors can recognize pathogen effectors by either direct or indirect mechanisms.
485 Plant breeders have long used these NB-LRR receptor resistance genes (R genes) for preventing
486 or reducing disease in crops. The long coevolutionary history between lodgepole pine and *D.*
487 *septosporum* (Capron et al., 2021; Welsh et al., 2009) likely resulted in some tolerance or
488 resistance mechanism against DNB in lodgepole pine, while potential new hosts like jack pine
489 may not have coevolved such specific defense. Lu et al. (2021) previously identified genes
490 involved in PTI and ETI and R genes with positive selection signals in lodgepole pine,

suggesting lodgepole pine adaptation to *D. septosporum* and specific interaction between pathogen effector and host receptors. In the present study, we discovered a few top candidate genes encoding LRR and serine threonine kinases in lodgepole pine populations (Figure 3, Table S1). Since most of the identified associated markers are likely not causal but rather are in LD with the causal loci, the observed fast LD decay rate in lodgepole pine implies promising fine mapping resolutions as the identified markers are likely to be proximal to the causative genes. On the contrary, the long LD blocks surrounding jack-hybrid pine candidate genes, which is typical of hybrid zones, limit the resolution of association mapping, thus it is difficult to locate the causative genes for follow-up studies.

5 | CONCLUSION

The recent outbreaks of DNB in lodgepole pine and its ability to infect jack pine pose a pressing need for understanding the genetic architecture of resistance to *D. septosporum*. The present study provides an array of candidate genes associated with *D. septosporum* response using pool-GWAS. In the studied lodgepole pine × jack pine hybrid zone, both divergent selection and introgression operate on the hybrid trees, but the adaptive introgression is yet to be found. This knowledge can be used to develop genomic tools to minimize DNB risk and guide forest management strategies. Further testing and validation of these associations using a SNP array is currently underway.

AUTHOR CONTRIBUTIONS

Mengmeng Lu and Sam Yeaman conceived analyses and wrote the manuscript. Mengmeng Lu performed computational analyses. Nicolas Feau and Richard C. Hamelin designed pathogen inoculation and phenotyping protocol. Nicolas Feau performed the pathogen inoculation and phenotyping in the greenhouse. Brandon Lind and Pooja Singh designed pool-GWAS analyses pipeline. Dragana Obreht Vidakov performed sequence capture work. Sally N. Aitken, Richard C. Hamelin, and Sam Yeaman obtained funding. All authors edited the manuscript.

ACKNOWLEDGMENTS

We thank Centre d'expertise et de services Génome Québec for sequencing service, University of Calgary Information Technologies for system support, the Research Oversight Committee of CoAdapTree Project for suggestions and help.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

FUNDING INFORMATION

This research was supported by the CoAdapTree Project (241REF), with funding from Genome Canada, Genome BC, Genome Alberta, Genome Québec, the BC Ministry of Forests, Lands, Natural Resource Operations and Rural Development, and many other sponsors http://adaptree.forestry.ubc.ca/sponsors/, http://coadaptree.forestry.ubc.ca/sponsors/, and seed contributors http://adaptree.forestry.ubc.ca/seed-contributors/. Dr. Sam Yeaman is also funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) and Alberta Innovates.

DATA AVAILABILITY STATEMENT

The raw pool-seq data were deposited in NCBI SRA (accession number: PRJNA602898; http://www.ncbi.nlm.nih.gov/sra). The assembled transcriptomes and their annotation were deposited in Dryad https://doi.org/10.5061/dryad.prr4xgxtk .The scripts used for analyses were deposited in https://github.com/Mengmeng-Lu/Genetic-architecture-underlying-response-to-the-fungal-pathogen-Dothistroma-septosporum-in-Pinus-con

ORCID

Mengmeng Lu https://orcid.org/0000-0001-5023-3759

Nicolas Feau https://orcid.org/0000-0001-5925-9867

Brandon Lind https://orcid.org/0000-0002-8560-5417

Pooja Singh https://orcid.org/0000-0001-6576-400X

Sally N. Aitken https://orcid.org/0000-0002-2228-3625

Richard C. Hamelin https://orcid.org/0000-0003-4006-532X

Sam Yeaman https://orcid.org/0000-0002-1706-8699

REFERENCES

Bechsgaard, J., Jorgensen, T. H., & Schierup, M. H. (2017). Evidence for Adaptive Introgression of Disease Resistance Genes Among Closely Related *Arabidopsis* Species. *G3 (Bethesda, Md.), 7*(8), 2677-2683. https://doi.org/10.1534/g3.117.043984

Burns, I., James, P. M. A., Coltman, D. W., & Cullingham, C. I. (2019). Spatial and genetic structure of the lodgepole × jack pine hybrid zone. *Canadian Journal of Forest Research, 49*(7), 844-853. https://doi.org/10.1139/cjfr-2018-0428

Capador-Barreto, H. D., Bernhardsson, C., Milesi, P., Vos, I., Lundén, K., Wu, H. X., . . . Elfstrand, M. (2021). Killing two enemies with one stone? Genomics of resistance to two sympatric pathogens in Norway spruce. *Molecular Ecology, 30*(18), 4433-4447. https://doi.org/10.1111/mec.16058

Capron, A., Feau, N., Heinzelmann, R., Barnes, I., Benowicz, A., Bradshaw, R. E., . . . Hamelin, R. C. (2021). Signatures of Post-Glacial Genetic Isolation and Human-Driven Migration in the *Dothistroma* Needle Blight Pathogen in Western Canada. *Phytopathology®, 111*(1), 116-127. https://doi.org/10.1094/phyto-08-20-0350-fi

Chang, M., Chen, H., Liu, F., & Fu, Z. Q. (2022). PTI and ETI: convergent pathways with diverse elicitors. *Trends in plant science, 27*(2), 113-115. https://doi.org/10.1016/j.tplants.2021.11.013

Cullingham, C. I., Cooke, J. E., Dang, S., Davis, C. S., Cooke, B. J., & Coltman, D. W. (2011). Mountain pine beetle host-range expansion threatens the boreal forest. *Molecular Ecology, 20*(10), 2157-2171. https://doi.org/10.1111/j.1365-294X.2011.05086.x

Cullingham, C. I., James, P. M., Cooke, J. E., & Coltman, D. W. (2012). Characterizing the physical and genetic structure of the lodgepole pine × jack pine hybrid zone: mosaic structure and differential introgression. *Evolutionary Applications, 5*(8), 879-891. https://doi.org/10.1111/j.1752-4571.2012.00266.x

Dale, A. L., Lewis, K. J., & Murray, B. W. (2011). Sexual reproduction and gene flow in the pine pathogen Dothistroma septosporum in British Columbia. *Phytopathology, 101*(1), 68-76. https://doi.org/10.1094/phyto-04-10-0121

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158. https://doi.org/10.1093/bioinformatics/btr330

De La Torre, A. R., Puiu, D., Crepeau, M. W., Stevens, K., Salzberg, S. L., Langley, C. H., & Neale, D. B. (2019). Genomic architecture of complex traits in loblolly pine. *New Phytologist, 221*(4), 1789-1801. https://doi.org/110.1111/nph.15535

Dodds, P. N. & Rathjen, J. P. (2010). Plant immunity: towards an integrated view of plant–pathogen interactions. *Nature review genetics, 11*(8), 539-548. https://doi.org/10.1038/nrg2812

Endler, L., Betancourt, A. J., Nolte, V., & Schlötterer, C. (2016). Reconciling Differences in Pool-GWAS Between Populations: A Case Study of Female Abdominal Pigmentation in *Drosophila melanogaster*. *Genetics, 202*(2), 843-855. https://doi.org/10.1534/genetics.115.183376

Feau, N., Ramsfield, T. D., Myrholm, C. L., Tomm, B., Cerezke, H. F., Benowicz, A., . . . Hamelin, R. C. (2021). DNA-barcoding identification of *Dothistroma septosporum* on *Pinus contorta* var. *latifolia*, *P. banksiana* and their hybrid in northern Alberta, Canada. *Canadian Journal of Plant Pathology, 43*(3), 472-479. https://doi.org/10.1080/07060661.2020.1829065

Futschik, A. & Schlötterer, C. (2010). The next generation of molecular markers from massively parallel sequencing of pooled DNA Samples. *Genetics, 186*(1), 207-218. https://doi.org/10.1534/genetics.110.114397

Hao, Z.-Z., Liu, Y.-Y., Nazaire, M., Wei, X.-X., & Wang, X.-Q. (2015). Molecular phylogenetics and evolutionary history of sect. Quinquefoliae (Pinus): Implications for Northern Hemisphere biogeography. *Molecular Phylogenetics and Evolution, 87*, 65-79. https://doi.org/10.1016/j.ympev.2015.03.013

Harrison, R. G. & Larson, E. L. (2014). Hybridization, Introgression, and the Nature of Species Boundaries. *Journal of Heredity, 105*(S1), 795-809. https://doi.org/10.1093/jhered/esu033

Hill, W. G. & Weir, B. S. (1988). Variances and covariances of squared linkage disequilibria in finite populations. *Theoretical population biology, 33*(1), 54-78. https://doi.org/10.1016/0040-5809(88)90004-4

Hivert, V., Leblois, R., Petit, E. J., Gautier, M., & Vitalis, R. (2018). Measuring Genetic Differentiation from Pool-seq Data. *Genetics, 210*(1), 315-330. https://doi.org/10.1534/genetics.118.300900

Isabel, N., Holliday, J. A., & Aitken, S. N. (2019). Forest genomics: Advancing climate adaptation, forest health, productivity, and conservation. *Evolutionary Applications, 13*(1), 3-10. https://doi.org/10.1111/eva.12902

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics, 24*(11), 1403-1405. https://doi.org/10.1093/bioinformatics/btn129

Lenth, R. V. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software, 69*(1), 1 - 33. https://doi.org/10.18637/jss.v069.i01

Lind, B. M. (2021a). GitHub.com/brandonlind/cmh_test: preprint release (Version 1.0.0). Zenodo. http://doi.org/10.5281/zenodo.5083798

Lind, B. M. (2021b). GitHub.com/CoAdapTree/varscan_pipeline: Publication release (Version 1.0.0). Zenodo. https://doi.org/10.5281/zenodo.5083302

Lind, B. M., Lu, M., Obreht Vidakovic, D., Singh, P., Booker, T. R., Aitken, S. N., & Yeaman, S. (2022). Haploid, diploid, and pooled exome capture recapitulate features of biology and paralogy in two non-model tree species. *Molecular Ecology Resources, 22*(1), 225-238. https://doi.org/10.1111/1755-0998.13474

Lotan, J. E. & Critchfield, W. B. (1990). Lodgepole pine. In R. M. Burns & B. H. Honkala (Eds.), *Silvics of North America* (Vol. Volume 1: Conifers, pp. 302-315). Washington, D.C.: USDA Forest Service.

Lu, M., Feau, N., Obreht Vidakovic, D., Ukrainetz, N., Wong, B., Aitken, S. N., . . . Yeaman, S. (2021). Comparative gene expression analysis reveals mechanism of *Pinus contorta* response to the fungal pathogen *Dothistroma septosporum*. *Molecular plant-microbe interactions: MPMI, 34*(4), 397-409. https://doi.org/10.1094/mpmi-10-20-0282-r

Lu, M., Krutovsky, K. V., Nelson, C. D., West, J. B., Reilly, N. A., & Loopstra, C. A. (2017). Association genetics of growth and adaptive traits in loblolly pine (Pinus taeda L.) using whole-exome-discovered polymorphisms. *Tree Genetics & Genomes, 13*(3), 57. https://doi.org/10.1007/s11295-017-1140-1

MacLachlan, I. R., McDonald, T. K., Lind, B. M., Rieseberg, L. H., Yeaman, S., & Aitken, S. N. (2021). Genome-wide shifts in climate-related variation underpin responses to selective breeding in a widespread conifer. *Proceedings of the National Academy of Sciences, 118*(10), e2016900118. https://doi.org/10.1073/pnas.2016900118

Marroni, F., Pinosio, S., Zaina, G., Fogolari, F., Felice, N., Cattonaro, F., & Morgante, M. (2011). Nucleotide diversity and linkage disequilibrium in *Populus nigra* cinnamyl alcohol dehydrogenase (CAD4) gene. *Tree Genetics & Genomes, 7*(5), 1011-1023. https://doi.org/10.1007/s11295-011-0391-5

Neale, D. B., Wegrzyn, J. L., Stevens, K. A., Zimin, A. V., Puiu, D., Crepeau, M. W., . . . Langley, C. H. (2014). Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biology, 15*(3), R59. https://doi.org/10.1186/gb-2014-15-3-r59

Paradis, E. & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics, 35*(3), 526-528. https://doi.org/10.1093/bioinformatics/bty633

Quesada, T., Gopal, V., Cumbie, W. P., Eckert, A. J., Wegrzyn, J. L., Neale, D. B., . . . Davis, J. M. (2010). Association mapping of quantitative disease resistance in a natural population of loblolly pine (*Pinus taeda* L.). *Genetics, 186*(2), 677-686. https://doi.org/10.1534/genetics.110.117549

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Ramsfield, T., Myrholm, C., & Tomm, B. (2021). Natural infection of *Pinus contorta* var. *latifolia*, *Pinus banksiana* and their hybrid by *Dothistroma septosporum* in Alberta, Canada. *Forest Pathology, 51*(5), e12717. https://doi.org/10.1111/efp.12717

Rudolph, T. D. & Laidly, P. R. (1990). *Pinus banksiana* Lamb. Jack Pine. In R. M. Burns & B. H. Honkala (Eds.), *Silvics of North America* (Vol. Volume 1: Conifers. Agriculture Handbook 654, pp. 280-293). Washington, D.C.: USDA Forest Service.

Rudolph, T. D. & Yeatman, C. W. (1982). Genetics of Jack Pine. In *Research Paper WO-38* (pp. 35-37). Washington, D.C.: USDA Forest Service.

Rundle, H. D. & Nosil, P. (2005). Ecological speciation. *Ecology Letters, 8*(3), 336-352. https://doi.org/10.1111/j.1461-0248.2004.00715.x

Schlötterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics, 15*(11), 749-763. https://doi.org/10.1038/nrg3803

Singh, P., StClair, J. B., Lind, B., Cronn, R., Wilhelmi, N. P., Feau, N., . . . Yeaman, S. (2024). Genetic architecture of disease resistance and tolerance in Douglas-fir trees. *New Phytologist*. https://doi.org/10.32942/osf.io/s5rd7

Stocks, J. J., Metheringham, C. L., Plumb, W. J., Lee, S. J., Kelly, L. J., Nichols, R. A., & Buggs, R. J. A. (2019). Genomic basis of European ash tree resistance to ash dieback fungus. *Nature Ecology & Evolution, 3*(12), 1686-1696. https://doi.org/10.1038/s41559-019-1036-6

Suren, H., Hodgins, K. A., Yeaman, S., Nurkowski, K. A., Smets, P., Rieseberg, L. H., . . . Holliday, J. A. (2016). Exome capture from the spruce and pine giga-genomes. *Molecular Ecology Resources, 16*(5), 1136-1146. https://doi.org/10.1111/1755-0998.12570

Taylor, S. A., Larson, E. L., & Harrison, R. G. (2015). Hybrid zones: windows on climate change. *Trends in Ecology & Evolution, 30*(7), 398-406. https://doi.org/10.1016/j.tree.2015.04.010

van Esse, H. P., Reuber, T. L., & van der Does, D. (2020). Genetic modification to improve disease resistance in crops. *New Phytologist, 225*(1), 70-86. https://doi.org/10.1111/nph.15967

Wegrzyn, J. L., Liechty, J. D., Stevens, K. A., Wu, L. S., Loopstra, C. A., Vasquez-Gross, H. A., . . . Neale, D. B. (2014). Unique features of the loblolly pine (*Pinus taeda* L.) megagenome

revealed through sequence annotation. *Genetics, 196*(3), 891-909. https://doi.org/10.1534/genetics.113.159996

Wei, T. & Simko, V. (2017). R package "corrplot": Visulaization of a correlation matrix. *https://github.com/taiyun/corrplot*.

Weiss, M., Sniezko, R. A., Puiu, D., Crepeau, M. W., Stevens, K., Salzberg, S. L., . . . De La Torre, A. R. (2020). Genomic basis of white pine blister rust quantitative disease resistance and its relationship with qualitative resistance. *The Plant Journal, 104*(2), 365-376. https://doi.org/10.1111/tpj.14928

Welsh, C., Lewis, K., & Woods, A. (2009). The outbreak history of *Dothistroma* needle blight: an emerging forest disease in northwestern British Columbia, Canada. *Canadian Journal of Forest Research, 39*(12), 2505-2519. https://doi.org/10.1139/X09-159

White, T. L. (2004). TREE BREEDING, PRINCIPLES | Breeding Theory and Genetic Testing. In J. Burley (Ed.), *Encyclopedia of Forest Sciences* (pp. 1551-1561). University of Oxford, Oxford, UK: Elsevier.

Woods, A. J. (2003). Species diversity and forest health in northwest British Columbia. *The Forestry Chronicle, 79*(5), 892-897. https://doi.org/10.5558/tfc79892-5

Wu, M. C., Kraft, P., Epstein, M. P., Taylor, D. M., Chanock, S. J., Hunter, D. J., & Lin, X. (2010). Powerful SNP-set analysis for case-control genome-wide association studies. *American Journal of Human Genetics, 86*(6), 929-942. https://doi.org/10.1016/j.ajhg.2010.05.002

Wu, T. D. & Watanabe, C. K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics, 21*(9), 1859-1875. https://doi.org/10.1093/bioinformatics/bti310

Yeaman, S., Hodgins, K. A., Lotterhos, K. E., Suren, H., Nadeau, S., Degner, J. C., . . . Aitken, S. N. (2016). Convergent local adaptation to climate in distantly related conifers. *Science, 353*(6306), 1431-1433. https://doi.org/10.1126/science.aaf7812