Canadian
Science
Publishing

PLANT GENOMIC RESOURCES

Comparative genomics of the chitinase gene family in lodgepole and jack pines: contrasting responses to biotic threats and landscape level investigation of genetic differentiation

Rhiannon M. Peery, Chandra H. McAllister, Catherine I. Cullingham, Elizabeth L. Mahon, Adriana Arango-Velez, and Janice E.K. Cooke

Abstract: The sister species, lodgepole pine (*Pinus contorta* var. *latifolia* Engelm.) and jack pine (*Pinus banksiana* Lamb), face pressures from a multitude of biotic agents, including mountain pine beetle (*Dendroctonus ponderosae* Hopkins, 1902) and their pathogenic fungal associates (e.g., *Grosmannia clavigera* (Rob.-Jeffr. & R.W. Davidson) Zipfel, Z.W. de Beer & M.J. Wingf.), mistletoe (*Arceuthobium americanum* Nutt. ex A.Gray), and the pathogen causing western gall rust (*Cronartium harknessii* E. Meinecke). Here, we report new stem tissue transcriptome resources developed for lodgepole and jack pines subjected to these biotic stresses. The annotated transcriptomes were compared to determine species-specific responses to the necrotrophic *G. clavigera* and the biotrophic *Cronartium harknessii*. We focused on chitinases, a family that includes members with well-documented roles in defense. Putative chitinase family members were identified via annotation, sequence similarity to reference chitinase genes, phylogenetic analyses, and in silico motif characterization. RNA-Seq revealed marked differences in the responses of lodgepole and jack pine chitinases to *G. clavigera* and *Cronartium harknessii*. The potential for adaptive variation in chitinases was investigated by assessing the level of genetic differentiation between and within lodgepole and jack pines using single nucleotide polymorphisms within chitinases. These analyses illustrate the potential of combining transcriptomic and genotyping resources to investigate genotype–phenotype correlations for non-model species.

Key words: defense genes, western gall rust, *Grosmannia clavigera*, necrotroph, biotroph, conifer.

Résumé : Les espèces sœurs du pin lodgepole (*Pinus contorta* var. *latifolia* Engelm.) et du pin gris (*Pinus banksiana* Lamb) subissent les pressions d'une multitude d'agents biotiques, dont le dendroctone du pin (*Dendroctonus ponderosae* Hopkins, 1902) et leurs champignons pathogènes associés (p. ex. *Grosmannia clavigera* (Rob.- Jeffr. & R.W. Davidson) Zipfel, Z.W. de Beer & M.J. Wingf.), le gui (*Arceuthobium americanum* Nutt. ex A.Gray) et l'agent pathogène responsable de la rouille-tumeur autonome (*Cronartium harknessii* E. Meinecke). Les chercheuses présentent ici de nouvelles ressources transcriptomiques du tissu de la tige développées pour le pin lodgepole et le pin gris soumis à ces stress biotiques. Les transcriptomes annotés ont été comparés pour déterminer les réponses au nécrotrophe *G. clavigera* et au biotrophe *Cronartium harknessii* spécifiques à l'espèce. Elles se sont concentrées sur les chitinases, une famille qui comprend des membres ayant des rôles bien documentés dans la défense. Les membres présumés de la famille des chitinases ont été identifiés par l'annotation, la similarité des séquences avec les gènes de référence de chitinases, des analyses phylogénétiques et la caractérisation de motifs in silico. Le séquençage d'ARN a révélé des différences marquées dans les réponses des chitinases du pin lodgepole et du pin gris à *G. clavigera* et *Cronartium harknessii*. Le potentiel de variation adaptative des chitinases a été étudié en évaluant le niveau de différenciation génétique entre et au sein des pins lodgepole et

pins gris en utilisant des polymorphismes de nucléotides uniques au sein des chitinases. Ces analyses illustrent le potentiel de la combinaison des ressources de la transcriptomique et du génotypage pour étudier les corrélations génotype–phénotype pour les espèces non modèles. [Traduit par la Rédaction]

Mots-clés : gènes de défense, rouille-tumeur autonome, *Grosmannia clavigera*, nécrotrophe, biotrophe, conifère.

Introduction

Conifers are generally long-living, dominant species within the environs that they inhabit. Conifer taxa are targeted by a plethora of pests and pathogens, which has led to the evolution of an array of different physical and chemical defenses (Tiffin and Moeller 2006). An early and effective response to attack by these biotic threats is critical for containment of pest and pathogen invasion and involves a complex combination of constitutive and induced defenses. Pathogenesis-related (PR) proteins are found in low or negligible levels in healthy plant tissue but can accumulate as an induced response in high concentrations during the onset of pathogen challenge (van Loon et al. 2006; Laluk and Mengiste 2010). In annual species, chitinases are PR proteins that have functionally validated roles in defense against biotic threats and are often used as molecular reporters or "sentinels" of plant defense (van Loon et al. 2006; Laluk and Mengiste 2010). Chitinases have similar roles in conifer defense against pests and pathogens (Hietala et al. 2004; Islam et al. 2011; Veluthakkal and Dasgupta 2012; Kolosova et al. 2014). Norway spruce (*Picea abies* (L.) H.Karst.) and slash pine (*Pinus elliottii* Engelm.) resistant phenotypes have early induction of chitinases in response to pathogen challenge of, (Davis et al. 2002; Hietala et al. 2004) and overexpression of, chitinases enhanced resistance to pathogens in black spruce (Noël et al. 2005) and western white pine (*Pinus monticola* Douglas ex D. Don) (Liu et al. 2011). Given increasing pest and pathogen pressure on forest systems, and the potential for chitinases to serve as biomarkers of resistance, a more thorough understanding of the role that chitinases play in conifer defense is warranted (Lovett et al. 2006; Liu et al. 2011; Gauthier et al. 2014).

Chitinases constitute a large gene family belonging to the glycosyl hydrolase (GH) class of enzymes (Davies and Henrissat 1995). Chitinases fall into two separate GH groups, GH-18 and GH-19, which likely have independent evolutionary origins (Hossain et al. 2010). Chitinases are further classified into seven biochemical classes (I–VII) based on the presence or absence of conserved chitin-binding (CBD) and chitinolytic domains, and the variable hinge domain (Collinge et al. 1993; Neuhaus et al. 1996; Neuhaus 1999; Grover 2012). Class I, II, IV, and VII chitinases are members of GH family 19 and share high sequence similarity with one another (Neuhaus 1999); whereas class III and V chitinases fall into GH family 18 (van Aalten et al. 2001). Of the proposed chitinase classes, class VII has not been well studied and its status as an independent class is still unclear (González et al.

2015). Six of the seven classes of chitinase have been identified within conifers to date through biochemical characterization, in silico sequence characterization, and phylogenetic analysis (Islam et al. 2011; Kolosova and Bohlmann 2012; Grover 2012). Class VI chitinases have not previously been described in conifers, although they have been identified in several angiosperm species (Grover 2012). It has been well documented that some chitinases belonging to classes II and IV act to catalyze the hydrolytic cleavage of β-1-4 linked *n*-acetyl glucosamine units of chitin found in arthropod exoskeletons and fungal cell walls and are associated with plant defense (Collinge et al. 1993; Jøhnk et al. 2005; Veluthakkal and Dasgupta 2012; Kolosova et al. 2014). Diverse functions have been reported for other chitinases, including lysozyme activity (Collinge et al. 1993), antifreeze activity (Griffith et al. 1997), plant cell wall synthesis (Sánchez-Rodríguez et al. 2012), vegetative storage proteins (Avice et al. 2003), and more. Therefore, it cannot be assumed that all sequences annotated as chitinases function in plant defense.

Lodgepole pine (*Pinus contorta* var. *latifolia* Engelm.) is a dominant species of western North American forests. Millions of hectares of lodgepole pine forests have been lost over the past two decades due to an ongoing epidemic of mountain pine beetle (MPB; *Dendroctonus ponderosae* Hopkins, 1902c) (Safranyik and Carroll 2006; Safranyik et al. 2010; Cooke and Carroll 2017). Over the course of the current outbreak, MPB has undergone significant range expansion, including eastward spread into jack pine (*Pinus banksiana* Lamb), a boreal forest species (Cullingham et al. 2011). While lodgepole pine shares a coevolutionary history with MPB (Kelley and Farrell 1998), jack pine is considered to be a new host (Cullingham et al. 2011). Mountain pine beetle vectors a number of fungal associates, including the blue-stain fungus *Grosmannia clavigera* (Rob.-Jeffr. & R.W.Davidson) Zipfel, Z.W.Beer & M.J.Wingf. (Safranyik and Carroll 2006; Adams et al. 2013; Arango-Velez et al. 2016). *Grosmannia clavigera* is a lesion-inducing pathogen that invokes jasmonic acid (JA), but not salicylic acid (SA), production by its host, providing evidence that this fungus is a necrotrophic pathogen (Arango-Velez et al. 2016). In addition to serving as a source of nutrition for MPB (Six 2012), these fungal associates may also play a role in modifying host defenses (Lieutier et al. 2009) and are often used as proxies for MPB attack (Raffa and Berryman 1982).

Lodgepole pine and jack pine are also pressured by *Cronartium harknessii* E. Meinecke, the causative agent of western gall rust (WGR). This stem rust is distributed

throughout Canada and forms galls on stems and branches that result in reduced growth, and can cause mortality in the case of stem galls on juvenile pines (van der Kamp and Spence 1987; Woods et al. 2000). The rust fungi (Pucciniales, Basidiomycota) are obligate biotrophs (Duplessis et al. 2011). Lodgepole pine is particularly susceptible to WGR, and the extensive loss of mature lodgepole pine trees as a result of the recent MPB outbreak has resulted in an age-structure shift towards juveniles that are more susceptible to WGR infection (Heineman et al. 2010).

Lodgepole and jack pine show markedly different responses to inoculation with *G. clavigera* (Lusebrink et al. 2011; Arango-Velez et al. 2016), suggesting that lodgepole pine–*G. clavigera* interactions may have been shaped by their shared coevolutionary history. Similarly, lodgepole pine is more susceptible to WGR than jack pine (Yang et al. 1999), suggesting jack pine may have more effective defenses against this fungal pathogen (Wu et al. 1996). Addressing the implications of pest/pathogen interactions requires robust indicators for the defense response that can be examined at the landscape scale. Multiple chitinases have been implicated in lodgepole pine and jack pine defense against *G. clavigera* (Arango-Velez et al. 2014; Kolosova et al. 2014) and other pathogens (Nsolomo and Woodward 1994). Given their well-characterized status and their use as markers of defense responses in annual plants, chitinases represent an excellent tool with which to test the hypothesis that lodgepole and jack pines' defenses against MPB and their fungal associates, as well as their responses to *Cronartium harknessii*, differ as a result of their evolutionary histories.

As a consequence of the size and complexity of chitinase gene families, together with the diverse functions and roles of members of these families, genomic approaches are a logical means by which to address this hypothesis. Building genomic-scale knowledge of processes within conifers is challenging because the genome sizes are large, ranging from 22 to 31 Gb in *Pinus* species (Neale et al. 2017), and gene flow among individuals is high (Yeaman and Jarvis 2006). Consequently, transcriptome resources remain the most feasible means by which to conduct genomic-scale investigations in these species. There are genomic resources available for lodgepole pine including transcriptomes (Parchman et al. 2010; Hall et al. 2013; Hodgins et al. 2016), single nucleotide polymorphism (SNP) resources (Cullingham et al. 2013a; Yeaman et al. 2016; Burns et al. 2019), and microsatellite assays (Cullingham et al. 2011). Jack pine has fewer genomic-level resources that are largely limited to comparative papers with lodgepole pine, including transcriptome (Hall et al. 2013), SNP (Cullingham et al. 2013a; Burns et al. 2019), and microsatellite resources (Cullingham et al. 2011). Several of the existing lodgepole pine and jack pine resources are modest in scale, and not all these resources were designed for comparative analyses.

In this study, we sought to develop comprehensive genomic-scale resources suitable for conducting comparative analyses of the responses of lodgepole and jack pines to biotic threats affecting these sister species at the landscape level. To this end, we generated and annotated comprehensive master transcriptomes for lodgepole pine and jack pine suitable for comparative analyses. As a key step towards enabling phenotype–genotype connections, these master transcriptomes were complemented by high-throughput SNP data for both lodgepole pine and jack pine from a 50 K Axiom SNP array designed for lodgepole pine. These resources were then used to test the hypothesis that chitinases contribute to species-level differences in defense responses of lodgepole and jack pines against the necrotroph *G. clavigera* and the biotroph *Cronartium harknessii*. Specifically, we (*i*) identified chitinase gene family members in lodgepole and jack pine, (*ii*) characterized these family members through in silico motif identification and phylogenetic analysis, (*iii*) identified family members responding to pathogen inoculation, and (*iv*) analyzed whether variation in chitinase genes show spatial variation in lodgepole pine and jack pine across their ranges. The integrated nature of the multiple genome-level resources that we developed in this study facilitates future genomic-scale investigations into defense responses and other adaptive traits of these sister species.

Materials and methods

Plant material for transcriptome sequencing

Three independent experiments were conducted on lodgepole pine and jack pine to generate tissues for transcriptome sequencing (Table 1). Each of the three experiments included a treatment in which trees were inoculated with stem-colonizing fungal pathogens. Two experiments included additional abiotic and biotic variables. In all cases, harvested stem material was flash-frozen in liquid nitrogen and kept at $-80\,°C$ until processed.

Experiment 1 evaluated the effect of nitrogen availability on lodgepole seedling defense responses to *G. clavigera*. One-year old seedlings from a provenance near Hinton, AB, Canada were obtained from PRT Growing Services and potted in 3.78 L pots as described in Arango-Velez et al. (2016). Seedlings were grown in controlled-environment chambers with $19\,°C$ constant temperature, 15 h day/9 h night photoperiod, 200–250 µmol photosynthetically active radiation (PAR), and 20%–35% relative humidity. A fully randomized, complete block design was used for the experiment. Starting two weeks after potting, plants were fertilized weekly with commercial N:P:K formulations: two applications of 0.5 g/L 15:30:15 (N:P:K), followed by two applications of 0.5 g/L 20:20:20 (N:P:K). Thereafter and until the end of the experiment, plants were fertilized weekly with approximately 400 mL of Hocking's complete nutrient solution (Hocking 1971) containing either 0.3 mmol/L NH_4NO_3 or 10 mmol/L NH_4NO_3. These treatments were sufficient to lead to significant differences in total N content in

Table 1. Summary of experiments conducted to generate lodgepole and jack pine tissues for transcriptome sequencing.

	Experiment 1	Experiment 2	Experiment 3
Species	Lodgepole pine	Lodgepole and jack pine	Jack pine
Plant age	Second growth cycle	First growth cycle	ca. 64 years old
Treatments	Low nitrogen fertilization, mock-inoculated	Mock-inoculated	No visible *Arceuthobium americanum*, uninoculated
	High nitrogen fertilization, mock-inoculated	*Cronartium harknessii*-inoculated	No visible *Arceuthobium americanum*, *G. clavigera*-inoculated
	Low nitrogen fertilization, *G. clavigera*-inoculated		Significant natural *Arceuthobium americanum* colonization, uninoculated
	High nitrogen fertilization, *G. clavigera*-inoculated		Significant natural *Arceuthobium americanum* colonization, *G. clavigera*-inoculated
Tissue(s)	Phloem and xylem	Whole stems	Phloem
Harvest time point(s)	7 dpi	7 dpi and 21 dpi	42 dpi
No. individuals per biological replicate	1	5	1
No. biological replicates	4	4	4
No. libraries	32	32	16

Note: dpi, days post inoculation. In each of the three experiments, stem tissues were harvested from individual plants challenged with stem-infecting pathogens (*Arceuthobium americanum*, *Cronartium harknessii*, and (or) *Grosmannia clavigera*) to maximize discovery of stem-expressed defense-associated sequences in the resulting transcriptomes.

needles (1.55% vs 2.05%, $p < 0.05$). Five weeks after starting fertilization treatments, the woody stem of each tree was either mock-inoculated with 10 μL of water or inoculated with 10 μL of a ca. 140 spores μL^{-1} *G. clavigera* isolate M001-03-03-07-UC04DL09 (Roe et al. 2010, 2011) suspended in Milli-Q water (Merck Millipore, Burlington, MA, United States) according to Arango-Velez et al. (2016). At 7 days post-inoculation (7 dpi), plants were harvested by collecting bark (hereafter referred to as phloem) and wood (hereafter referred to as xylem) from the portion of the stem that was subjected to treatment.

Experiment 2 compared responses of lodgepole and jack pine seedlings to *Cronartium harknessii* inoculation. Two provenances were used for the study: lodgepole pine from Shelter Creek, AB, Canada and jack pine from Stoney Mountain Road, AB, Canada. Seeds were surface sterilized using bleach (1% (*v/v*) sodium hypochlorite) according to Groome et al. (1991), and were moist chilled between sterilized layers of Kimpads in clamshell containers at 4 °C for three weeks to break dormancy. Following the moist chilling, seeds were surface sterilized again with 1% (*v/v*) sodium hypochlorite prior to sowing in 2:1 (*v/v*) peat moss–vermiculite within 220 mL cavity styroblocks (Beaver Plastics, AB, Canada). Seedlings were grown in controlled-environment growth chambers with 22 °C day/16 °C night temperatures, an 18 h day photoperiod, ca. 250 μmol PAR, and ca. 80% relative humidity. In the absence of a mist chamber, styroblocks were placed in sealed, tented plastic bags for the first two weeks following sowing to create elevated humidity conditions that maximize and synchronize germination. After 8 weeks, half of the seedlings for each species were inoculated using the torn needle method (Myrholm and Hiratsuka 1993) with *Cronartium harknessii* spores isolated from galls

collected from naturally infected mature trees located near Hinton, AB, Canada. Treated plants were inoculated with a 3:1 (*v/v*) spore to talc ratio and controls were mock inoculated using only talcum powder. After inoculation/mock inoculation, blocks were again covered with sealed, tented plastic bags for two days while being kept in the dark and at a constant temperature of 16 °C to promote fungal growth. After two days, bags were removed and blocks were returned to the pre-inoculation growth room settings. Seedlings were harvested at 7 and 21 dpi.

Experiment 3 compared responses of mature jack pine that were either uninfected or heavily infected with American dwarf mistletoe (*Arceuthobium americanum* Nutt. ex A.Gray) to inoculation with *G. clavigera*. This full factorial experiment was conducted at the Alberta Tree Improvement and Seed Centre (54°53′N, 112°14′E), located about 116 km northeast of Edmonton, AB, Canada. Jack pine used for this experiment were selected from an even-aged (ca. 64 years old), naturally regenerated stand with a relatively high proportion of trees naturally infected with *Arceuthobium americanum*. Genetic analysis using microsatellites (Cullingham et al. 2012) confirmed that these trees were not lodgepole–jack pine hybrids. A total of 40 experimental trees were selected based on either visible absence of *Arceuthobium americanum* infection or visible *Arceuthobium americanum* colonization affecting more than one third of the tree crown. Experimental trees otherwise showed no signs of insect, disease, or mechanical damage. Ten *Arceuthobium americanum*-free and 10 *Arceuthobium americanum*-colonized trees were inoculated July 11–12, 2011 using *G. clavigera* isolate M001-03-03-07-UC04DL09 as described in Arango-Velez et al. (2014, 2016). In brief, a cordless drill was used to make two rings

Table 2. Sampling locations of provenances included in population genetic analyses of single nucleotide polymorphism loci including latitude and longitude, number of individuals assessed (n), per provenance averaged Q value used to determine species, observed heterozygosity (H_O), within provenance heterozygosity (H_S), total heterozygosity (H_T), and inbreeding coefficient (G_{IS}) of lodgepole and jack pine provenances assessed with 21 907 SNP loci.

Species	Provenance location	Latitude	Longitude	n	Q	H_O	H_S	H_T	G_{IS}
Lodgepole pine	Minto, YT	62.6	−136.83	15	0.991	0.267	0.328	0.328	0.187
	Baldy Hughes, BC	53.67	−122.95	6	0.986	0.302	0.332	0.332	0.091
	Edgewood, BC	49.92	−118.18	7	0.983	0.325	0.330	0.330	0.017
	Cypress Hills, AB	50.00	−110.00	7	0.974	0.296	0.318	0.318	0.069
Jack pine	Creighton, SK	54.85	−102.42	1	0.013	NA	NA	NA	NA
	London, ON	52.00	−92.50	10	0.003	0.113	0.124	0.124	0.093
	Weagamow Lake, ON	43.00	−81.00	5	0.002	0.124	0.123	0.123	−0.008
	Depres Lake, NB	46.65	−65.57	13	0.001	0.122	0.124	0.124	0.016

of holes approximately 8 mm diameter, 2–3 cm apart, to the depth of the cambial zone, around the circumference of the main trunk at breast height and at ca. 40 cm higher than breast height. Each hole was filled with a plug of *G. clavigera* mycelium grown on 2% malt extract agar, then plugged with a small piece of sterilized dowel rod and wrapped with Parafilm. At 42 dpi, trees were felled, bark was removed from bolts containing the *G. clavigera*-inoculated zone or equivalent for control trees, and the secondary phloem was peeled from the bark.

**Plant material for SNP genotyping**

SNP genotyping was carried out using materials collected from lodgepole and jack pine seedlings representing eight provenances spanning the species ranges from British Columbia to New Brunswick (Table 2; Fig. 1), provided by the National Tree Seed Centre, Natural Resources Canada. Seeds were surface sterilized prior to moist chilling as described for Experiment 2 with the following modifications: germination at 25 °C with ca. 75% humidity under a 12 h day photoperiod. Megagametophytes were removed from germinated seedlings once cotyledons had elongated but megagametophytes were still attached to the seedlings.

RNA isolation, library generation, and Illumina sequencing

All tissues were ground to a fine powder in liquid nitrogen before RNA extraction. RNA was isolated from mature tree samples via a large-scale hexadecyl–trimethyl ammonium bromide protocol outlined in Chang et al. (1993), and from seedling samples using a small-scale modification of this protocol as described in Pavy et al. (2008). RNA quantity and quality were measured with an Infinite M200 PRO NanoQuant (Tecan, Männedorf, Switzerland) and 2100 Bioanalyzer (Agilent, Mississauga, ON, Canada). Four biological replicates were prepared for each of the 20 different combinations of species, treatments, time points, and tissues, for a total of 80 samples.

Stranded, indexed RNA-Seq libraries were generated for each of these 80 samples using the TruSeq Stranded mRNA LT Sample Prep kit (Illumina, San Diego, CA, USA) by means of the standard procedure outlined by

Fig. 1. Sample locations of lodgepole pine and jack pine provenances included in single nucleotide polymorphism analyses overlaid with species ranges (within Canada). Number of individuals per location and exact coordinates used to mark sampling locations are provided in Table 2. Location information was downloaded from the National Tree Seed Centre (https://www.nrcan.gc.ca/science-data/ research-centres-labs/forestry-research-centres/atlantic-forestry-centre/national-tree-seed-centre/13449#collection). Figure was completed using ArcMap v10.5 (ESRI) and species ranges were approximated from Burns et al. (2019) and corrected for pine distribution in Canada (Yemshanov et al. 2012).
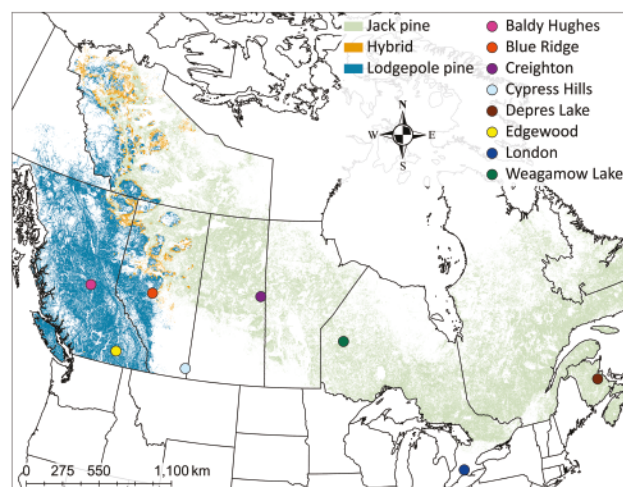
the manufacturer. Agencourt Ampure beads (Beckman Coulter, Indianapolis, IN, USA) and Superscript II Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA) were substituted for the equivalent components in the TruSeq kits. Prior to pooling, quantification was carried out using a Qubit Fluorometric Quantitation (Invitrogen), and quality and fragment size were assessed via the 2100 Bioanalyzer. Pooled samples were sent to the University of Alberta Molecular Biology Services Unit for further processing and sequencing. Sequencing was carried out on an Illumina NextSeq 500 using NextSeq 500/550 High Output v2 cartridges (Illumina).

Fig. 2. Assembly (A) and annotation (B) workflows for generating the control and master transcriptomes for both lodgepole pine and jack pine. Stages are within boxes and programs used are in italics below.
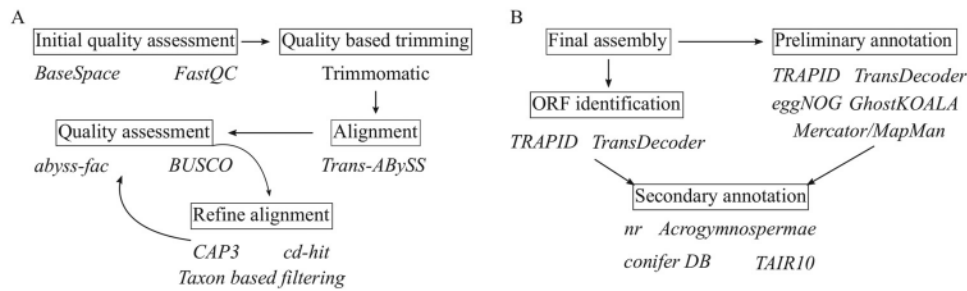
Transcriptome assembly and annotation

All samples were processed following the workflow for transcriptome assembly and annotation outlined in Fig. 2. Sequences were binned, adapters were removed, and fastq files were generated within Illumina's BaseSpace environment (https://basespace.illumina.com/home/index). Four lanes from each run were evaluated in FastQC v0.11.8 (Andrews et al. 2018) to quickly check data quality. Reads were trimmed by lane using Trimmomatic v0.36 (Bolger et al. 2014) with the following parameters: ILLUMINACLIP:/AdapterIndex.fa:2:30:10 SLIDINGWINDOW:4:15 MINLEN:51. All trimming was done within Compute Canada's HPC environment (www.computecanada.ca).

Data for all 80 individuals were combined in silico to generate the master transcriptomes, while the 44 pathogen-challenged samples were removed to generate control transcriptomes from the remaining 36 individuals. Assemblies were iterative, beginning with assembly using several k-mers for each individual, followed sequentially by merging treatment groups, experiments, and species. All assemblies were generated using Compute Canada's HPC environments. Assemblies were generated with Trans-ABySS v1.5.5 (Robertson et al. 2010). Values for k-mers were optimized to balance total number of contigs with assembly quality, as determined by the statistics reported in ABySS v1.5.2 (Simpson et al. 2009). Assemblies for each individual were run for k-mer values 25, 29, 33, and 36, and we used a modified *assemble.sh* script written by Ka Ming Nip (https://github.com/bcgsc/). The programs cd-hit v4.6.8 (Li and Godzik 2006) and CAP3 (Huang 1999) were used to collapse the over-assembled transcriptomes at 99% similarity followed by 95% similarity. The merged and collapsed transcriptomes were then filtered for fungal sequences and green plants with custom databases generated from filtering the non-redundant database (accessed 25 January 2018). Contigs less than 250 bp were removed using Geneious v10.2 (Biomatters Inc., NJ, USA), followed by an additional collapsing of the transcriptomes at 95% similarity with CAP3 (Huang 1999). At each stage, transcriptome quality was assessed with the abyss-fac command in ABySS v1.5.2 (Simpson et al. 2009) and inferred through retention and quality of orthologous

genes, which was assessed using BUSCO v2.0.1 (Simão et al. 2015).

Control transcriptomes were submitted to the Mercator4 v1.0 web portal (https://www.plabipd.de/portal/mercator4) to generate MAPMAN annotations (Thimm et al. 2004; Lohse et al. 2014), and master transcriptomes were submitted to a variety of annotation tools (Fig. 2B) to generate a functional annotation of these data. We identified the longest open reading frames (ORFs) in each contig of the assembly with TransDecoder v3.0.1 (Haas et al. 2013) and TRAPID (Van Bel et al. 2013). Annotation of transcriptome contigs and identified ORFs was preformed to generate GO, COGs, KEGG, and MAPMAN categories. These were identified with TRAPID (Van Bel et al. 2013), eggNOG (Huerta-Cepas et al. 2016, 2017), ghostKOALA (Kanehisa et al. 2016), and Mercator4 v1.0 (Thimm et al. 2004; Lohse et al. 2014), respectively. Open reading frames identified by TransDecoder and TRAPID searches were used to query databases generated from: (*i*) Acrogymnospermae identical proteins database (accessed 18 February 2018; GenBank search "txid1437180 [Organism:exp] NOT unknown"); (*ii*) all annotated conifer transcriptome data within the TreeGenes database (Wegrzyn et al. 2008), ConGenIE database, and the conifer EST database on GenBank (accessed 18 April 2017); (*iii*) protein sequence data from TAIR10 (The Arabidopsis Information Resource); and (*iv*) the GenBank non-redundant database restricted to green plant protein reference sequences with an e value of -5. Venn diagrams comparing number of contigs annotated were generated with VennDiagram v1.6.0 (Chen and Boutros 2011) and gridExtra v2.3 (Baptiste 2017) packages in R v3.5.2 (R Core Team 2018) and RStudio v1.1.463 (RStudio Team 2018).

Comparative transcriptomics

The reciprocal best hit (RBH) method, which uses BLAST algorithms to identify putative orthologs based on sequence similarity (Altenhoff and Dessimoz 2009; Camacho et al. 2009), was used to compare transcriptome content between the lodgepole and jack pine master transcriptomes to conifers with more than 100 000 contigs published on TreeGenes. These RBH pairwise comparisons, simplified with a python script

(Cock et al. 2015), included white spruce (*Picea glauca* (Moench) Voss; Warren et al. 2015), whitebark pine (*Pinus albicaulis* Engelm.; Baker et al. 2018); Chinese red pine (*Pinus massoniana* Lamb.; Fan et al. 2014), Mexican weeping pine (*Pinus patula* Schiede ex Schltdl. & Cham.; Visser et al. 2015), and Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco; Little et al. 2016). These transcriptomes were generated from either needle (white spruce and whitebark pine), stem (Mexican weeping pine), whole seedling (Chinese red pine), or megagametophyte tissues (Douglas-fir). The previously published lodgepole and jack pine transcriptomes (Hall et al. 2013) were also compared. Commonalities and differences in putative orthologous contigs from pairwise comparisons were visualized using Venn diagrams generated in the same fashion as described for annotation comparisons.

Transcriptome-wide chitinase identification and phylogenetic analysis

To identify putative chitinase contigs within the master transcriptome, we generated custom BLAST databases with reference chitinase sequences. Databases were generated from reference sequences using the white spruce gene catalog (Rigault et al. 2011), Norway spruce (González et al. 2015; Sundell et al. 2015), lodgepole pine and jack pine chitinase sequences (GenBank accession numbers in Supplementary data Table S1),[1] and any sequences identified as chitinase or chitinase-like on GenBank (as of 29 June 2018). This approach was designed to capture all potential chitinase gene family sequences within the master transcriptomes. All putative chitinase contigs from GH-18 and GH-19 were then translated into deduced amino acid sequences and aligned as two separate groups in Geneious using gap penalty of -2.7 and the MAFFT plug-in (Katoh 2002) with representative chitinase genes included (Supplementary data Table S1).[1] Contigs identified by BLAST searches were not initially filtered for false positives, short, or poor-quality matches; therefore, we heavily pruned the data matrix at the alignment stage of phylogenetic analysis. Poorly aligned sequences and sequences with fewer than 100 amino acids were discarded. The final alignments were submitted to the CIPRES Science Gateway to generate a maximum likelihood, bootstrapped phylogeny using RAxML v8.2.12 algorithm (Stamatakis 2014) and the WAG substitution model. WAG was chosen because it was the best fit model determined by the suggested method and Perl script provided with RAxML on GitHub (https://github.com/stamatak/standard-RAxML).

To determine the selection pressure on newly identified pine chitinase genes, we subsampled the chitinase alignment to generate per class alignments. By calculating selection pressure on consensus sequences generated through building the master transcriptome, the ts/tv (κ) and dN/dS (ω) ratios are conservative estimates of the true amount of selection within each chitinase gene class. These alignments were pruned to only those sequences that could be unambiguously categorized (Supplementary data Table S9)[1] and were full length. Outgroup orthologs were mined from *Pinus* species included in RBH comparisons and aligned in frame. The resultant alignments were submitted to IQ-TREE 2 (parameters: -B, 5000; -m, TEST; Minh et al. 2020) to generate a phylogeny for input to the codeml script within the PAML4 package (parameters: runmode = 0, seqtype = 1, CodonFreq = 2, clock = 0, model = 0, NSsites = 0, cleandata = 1; Yang 2007).

Chitinase expression profiling

Read mapping of trimmed reads to the master transcriptome was carried out using Bowtie 2 v2.3.4.1 (Langmead and Salzberg 2012) on the Compute Canada HPC cluster using a custom workflow (https://github.com/rmpeery/dataProcessing). Contigs identified as chitinase through phylogenetic analysis were used to generate simplified, SAF formatted annotation files; these were submitted along with bam files to featureCounts v1.6.2 (Liao et al. 2014) to generate count tables. These tables were input into edgeR v3.22.5 (Robinson and Oshlack 2010; Robinson et al. 2010; McCarthy et al. 2012; Zhou et al. 2014) to calculate $\log_2$-fold-changes and p values. Transcript abundance comparisons were always generated between treatment (pathogen challenge) and control (mock inoculations), with other experimental factors such as nitrogen treatment or tissue type included as additive traits in the model. Comparisons were considered significant at $p \leq 0.05$ and a minimum $\log_2$-fold-change of 1. Heatmaps were generated using ggplot2 v3.1.0.9 (Wickham 2016), R v3.5.0 (R Core Team 2018), and RStudio v1.1.383 (RStudio Team 2017).

High-throughput SNP genotyping and SNP annotation

Single nucleotide polymorphism genotyping was carried out at GeneSeek (Neogen; Lincoln, NE, USA) on seedlings described in the "Plant material for population genetics" section. Genotypes were generated using the AdapTree pine Affymetrix Axiom 50 K SNP array (Suren et al. 2016; MacLachlan et al. 2018; Mahony et al. 2020). Genotypes were called within Axiom Analysis Suite v3.1 (Thermo Fisher Scientific, Waltham, MA, USA). Parameters for sample quality and genotype calling results were optimized as: pass value ≥92; minimum call rate ≥92; average call rate ≥93; and call rate cut off ≥90. Single nucleotide polymorphism loci categorized in the "poly high resolution" cluster type, meaning those clusters that were perfectly bi-allelic, were selected, then assessed using three quality controls to generate data with the highest standards of quality (Laurie et al. 2010). First, because paralogous genes are prevalent within pine genomes (Cui et al. 2006), we examined the haploid tissue and filtered out loci with high heterozygosity assuming

---

[1]Supplementary data are available with the article at https://doi.org/10.1139/cjb-2020-0125.

those were from paralogs. Second, we looked for differences between individuals exhibiting an overabundance of unique alleles compared to individuals within the same provenance 2.7.1 (Kamvar et al. 2014, 2015). Finally, we used STRUCTURE (Pritchard et al. 2000) to validate expected species identity as lodgepole pine, jack pine, or a hybrid of lodgepole and jack pines. Species identity cut off values follow recommended Q values from Cullingham et al. (2013b). If a sample did not type as the expected species, it was excluded from further analysis.

Single nucleotide polymorphism loci were annotated using the *Pinus taeda* assembly v1.01 (Zimin et al. 2014). An in-house PERL script was used to separate relevant contigs from the genome assembly multifasta file (https://github.com/rmpeery/dataProcessing). A Python script was used to extract 500 bp on either side of the SNP location from each of these contigs (Kozik 2010). Single nucleotide polymorphism loci were then annotated by sequence homology using BLAST+ (Altschul et al. 1990; Camacho et al. 2009) and custom databases were generated for transcriptome annotation as described in the "Transcriptome-wide chitinase identification and phylogenetic analysis" section of the methods.

Identification of chitinase SNP and population level genetic diversity

The subset of SNP putatively within chitinase contigs were identified by comparing the sequence flanking the SNP locus to the verified chitinase contigs using BLASTN. Matches were filtered for a percent identity of ≥98% to the contig sequences and a single mismatch between probe and contig sequences. These stringent cut off values were applied to minimize false positive identifications of SNP within transcriptome contigs.

If a locus is under selection, there is often a change in the F_{ST} value of that locus (Lewontin and Krakauer 1973). To determine if chitinase SNP loci are under selection, we grouped the mined chitinase SNP (20 loci) to increase power in our analysis and determine the between species F_{ST} value. To generate a null distribution for comparison, we subset 20 loci randomly from the complete dataset (less the chitinase SNP = 21 894 loci) 100 times and calculated the between species F_{ST} values for each subset. These data were summarized as a histogram plot using the ggplot2 package v3.2.1 (Wickham 2016). Additionally, we calculated the per locus F_{ST} values among all loci to see where the chitinase loci fall within this greater distribution to determine if any chitinase loci were potential outliers. This analysis is also visualized with a histogram. We visualized the relationships among allele frequencies within and between lodgepole and jack pines via principal component analysis (PCA) in R using the packages adegenet v2.1.1 (Jombart and Ahmed 2011) and hierfstat v0.04-22 (Goudet and Jombart 2015) and using the ggplot2 package v3.2.1 (Wickham 2016).

Results

Generation of lodgepole pine and jack pine transcriptomes

To create robust, high-quality and high-coverage transcriptomes suitable for comparative genomic studies, cDNA libraries of 48 lodgepole and 32 jack pine individuals were sequenced from three independent experiments (BioProject PRJNA524866). In total, four biological replicates for each of 20 different treatments were used to generate transcriptomes. These 20 treatments comprised control trees and trees challenged with either the necrotrophic pathogen *G. clavigera* or the biotrophic pathogen *Cronartium harknessii*; in some cases, trees were simultaneously challenged with other biotic and abiotic stresses (Table 1). The master transcriptome included all 80 individuals, which were selected to represent diverse stem-expressed defense-associated transcriptome contigs within the resultant assembly. We also constructed a control transcriptome for both species, comprising the 24 lodgepole and 12 jack pine individuals that were not subjected to the biotic stresses of *G. clavigera*, *Cronartium harknessii*, or *Arceuthobium americanum*. Read generation ranged from just over 18.5 million pairs in the smallest library to almost 115.5 million in the largest library (Supplementary data Table S2).[1] Overall, the libraries were of high quality, supported by the number of paired reads and Embryophyta (land plants) conserved orthologous groups (COGs) retained after all quality filters (Table 3). The master transcriptomes retained 582 and 608 COGs in lodgepole pine and jack pine, and the control transcriptomes retained 575 and 621, respectively.

Many of the COGs identified in the initial, unfiltered assemblies were present as duplicates and the total number of contigs generated were in the millions (Table 3). This reflects the genetics of the individual pine trees included — each is a unique, highly heterozygous individual with no family or provenance structure. Combining these individuals with a lot of inherent variability led to splitting homologous genes due to high allelic variation rather than collapsing the contigs into one unit. Therefore, to improve the utility of these master transcriptomes for subsequent analyses such as RNA-Seq, contigs were also merged based on percent similarity. Additionally, we filtered out contigs that were not identified as belonging to Viridiplantae (green algae and plants) through BLASTP searches (Supplementary data Tables S3–S6).[1] These efforts minimized redundancy while retaining the maximum number of COGs (Supplementary data Table S6).[1] The master assemblies for jack and lodgepole pines comprise 473 406 and 375 632 contigs, respectively (Table 3). Both control transcriptomes have fewer total contigs at 227 072 (jack pine) and 253 913 (lodgepole pine).

The master and control transcriptomes were both annotated using GO, GenBank nr database, Mercator/MAPMAN, COG/KOG, and KEGG. Annotations for a given

Table 3. Assembly quality was assessed using BUSCO (conserved orthologous groups) and N50 (length weighted median) for the initial and final control and master transcriptomes of both jack and lodgepole pines (above solid line).

| Transcriptome | Stage | BUSCO | | | | ABySS | | Tissue |
| | | Complete | | | | | | |
		Single copy	Duplicated	Fragmented	Missing	N50	No. of contigs	
Lodgepole control	Initial	80	1170	30	160	1788	2 175 250	Stem
	Final	575	482	76	307	1936	253 913	Stem
Lodgepole master	Initial	56	1193	37	154	1878	7 446 695	Stem
	Final	582	564	68	226	2001	375 632	Stem
Jack control	Initial	48	1198	37	157	1819	3 598 773	Stem
	Final	621	437	73	309	1931	227 072	Stem
Jack master	Initial	38	1206	39	157	1851	10 170 000	Stem
	Final	608	537	6	232	1868	473 406	Stem
Picea glauca	Final	308	591	177	364	1703	455 504	Needle
Pinus albicaulis	Final	47	1157	45	191	2026	357 872	Needle
Pinus banksiana	Final	913	292	44	191	1849	158 666	Stem
Pinus contorta	Final	767	192	134	347	1503	146 731	Stem
Pinus massoniana	Final	259	971	34	176	1758	264 406	Seedling
Pinus patula	Final	1112	130	30	168	1988	105 454	Stem
Pseudotsuga menziesii	Final	457	780	42	161	1825	331 725	Megagametophyte

Note: Quality was assessed using the same parameters for transcriptomes included in the reciprocal best-hit analysis (below solid line; Fig. 4) and previously published lodgepole and jack pine transcriptomes. Details on assemblies for individual lodgepole and jack pine libraries at each stage of assembly including steps before the initial assembly stage are reported in Supplementary data Tables S1–S5.[1]

contig were considered more robust if the same or very similar annotation information was obtained from more than one database (Fig. 3). The GenBank nonredundant database produced the most uniquely annotated contigs, 139 351 (lodgepole pine) and 98 623 (jack pine), while KEGG produced the smallest number (437 and 999) of uniquely annotated contigs for lodgepole and jack pine. Overall, only 8.6% of contigs in lodgepole and 13.6% of the jack pine master transcriptomes did not contain some form of annotation information. Results from the plant focused Mercator4 annotation showed that all major categories (bins) are represented in the master and control transcriptomes (Table 4; Supplementary data Table S7),[1] and there were more Mercator4 subcategories represented in the master transcriptomes than in the control transcriptomes. Similar to using COGs as a measure of transcriptome completeness, the number of Mercator4 functional categories represented can be used to infer the completeness of the transcriptomes. Of all the subcategories considered, contigs were annotated for 62.2%–90% for all four transcriptomes generated.

One of the key objectives of this study was to produce transcriptomes for comparative genomics. To identify putatively orthologous contigs between lodgepole and jack pine transcriptomes, we used a pairwise RBH method that finds contigs in which the best match in Species A is also the best match in Species B (available at https://doi.org/10.7939/DVN/A7BWB8). Proportionally, the two master transcriptomes were determined to be about 20% orthologous (84 446 RBH). The lodgepole control and master transcriptomes contained 175 880 RBH, meaning approximately 38% of the master transcriptome was

unique. The control and master transcriptomes of jack pine yielded similar results (141 107 RBH), indicating that approximately 31% of the master transcriptome represented unique contigs.

When both jack pine and lodgepole pine master transcriptomes were compared to contigs of these five Pinaceae transcriptomes using RBH, considerable variation was found in the total numbers of contigs shared with either lodgepole pine (Fig. 4A) or jack pine (Fig. 4B). Mexican weeping pine is the closest relative examined in the comparative RBH study and we found that 3.4%–4.2% of lodgepole and jack pines contigs were shared (Supplementary data Table S8)[1] between species, respectively. Chinese red pine shared the most putative orthologs with lodgepole (8.2%) and jack pine (6.4%). Collectively there were over 2000 contigs in common among all pairwise comparisons. The contigs common to all Pinaceae included in this analysis, i.e., the core set of genes, fell within all functional categories represented in MAPMAN bins.

Identification, in silico characterization, and expression analysis of chitinases

Candidate chitinase contigs were identified through mining annotations of the master transcriptome, along with sequence similarity determined by BLASTP to (*i*) published chitinase genes in *Arabidopsis* and white spruce (Rigault et al. 2011; González et al. 2015), and (*ii*) sequences fetched from the GenBank protein database using a query for chitinase and filtered to contain results only from Species:plants, and Source database:RefSeq. This approach intentionally captured all potential chitinase contigs, including related glycosyl hydrolases; therefore, the list

Fig. 3. Total number of contigs annotated with several tools (see Fig. 2B) in master transcriptomes of lodgepole pine (A) and jack pine (B). Venn diagrams show commonalities among annotations of lodgepole pine and jack pine using gene ontology (GO) categories, the GenBank non-redundant sequences database limited to Viridiplantae sequences (nr), the Kyoto Encyclopedia of Genes and Genomes (KEGG), conserved orthologous groups (COG/KOG), and the plant specific MAPMAN database plaBi. [Colour online.]
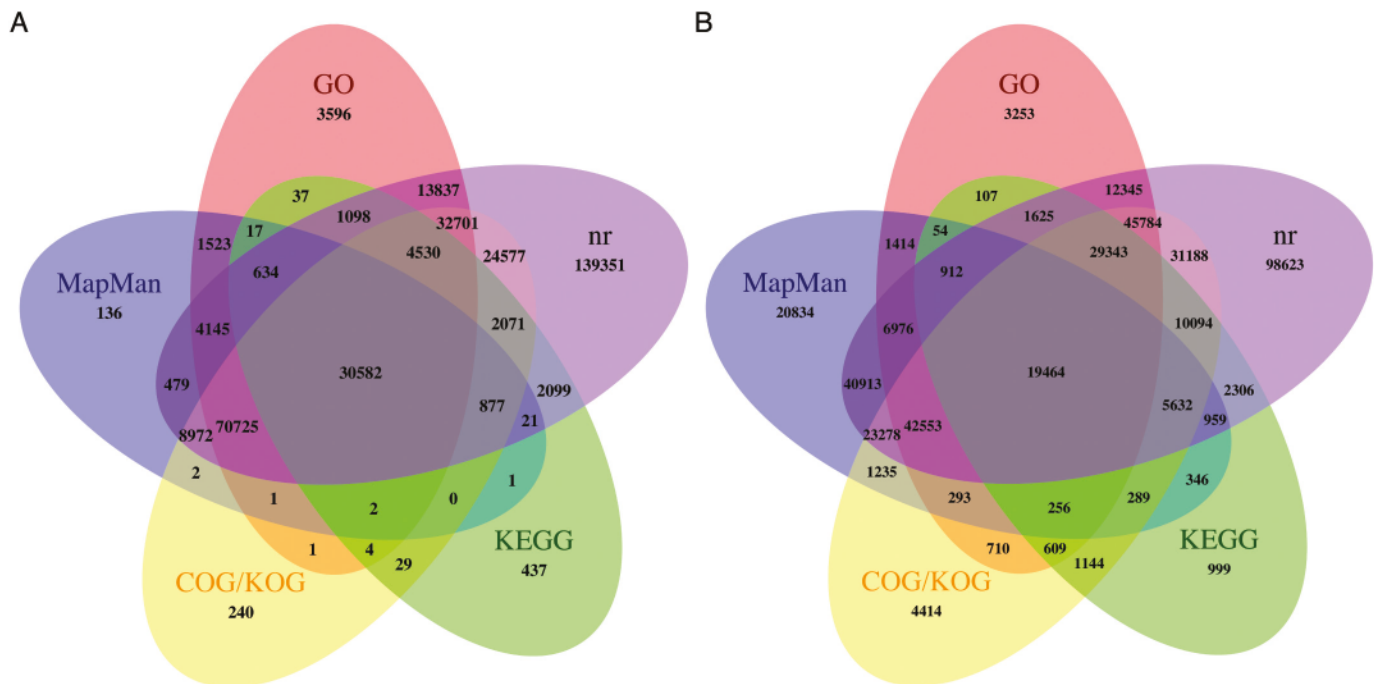
of putative chitinases was pruned through sequence alignment, assessment of homology via preliminary phylogenies, and parametric phylogenetic inference using maximum likelihood. These analyses yielded a final dataset containing 134 lodgepole pine chitinases and 94 jack pine chitinases (Table 5; Supplementary data Table S9).[1] The final alignment of GH-18 comprised 61 sequences (database sequences are listed in Supplementary data Table S1)[1] and 434 aligned characters. The GH-19 alignment comprised 261 sequences and 479 aligned characters. Alignment of GH-18 and GH-19 chitinases was split because they are hypothesized to have different evolutionary origins (Hossain et al. 2010) and we found no homologous sites shared between the two groups. These are strong indicators that they should not be combined for rigorous phylogenetic analysis. The master transcriptomes of both lodgepole and jack pine contained contigs that corresponded to chitinase classes I–V and VII. These classifications are based both on position within monophyletic groups within the phylogenetic tree (Fig. 5) and the presence or absence of diagnostic canonical motifs (Supplementary data Table S9).[1] As these canonical motifs are diagnostic (Grover 2012; González et al. 2015), we investigated their presence or absence in every chitinase contig from the master transcriptomes (Supplementary data Table S9)[1] and used these to validate chitinase designations. This was especially critical for groups I and II, which were not monophyletic in our analysis. From the

RBH analysis, we found 29 verified chitinase contigs that are putative orthologs between the two species. Lastly, we investigated the selection pressure on each class of chitinase gene and found that all chitinase genes are under purifying selection (Table 5 and Supplementary data Table S101; alignments available at https://doi.org/10.7939/DVN/A7BWB8).

To examine pathogen responsiveness of these chitinases, the RNA-Seq data were used to generate expression profiles for the 134 lodgepole pine and 94 jack pine chitinases that were identified from the above analyses (Supplementary data Table S8).[1] Most chitinases had detectable levels of expression in at least one treatment ($p \le 0.05$; Supplementary data Fig. S11).[1] Within the *G. clavigera* experiments, lodgepole pine had nine contigs below detectable levels of expression and jack pine had 36. In the *Cronartium harknessii* experiments, there were 11 lodgepole and 11 jack pine chitinase contigs with undetectable levels of expression.

Within the expression profile, we found distinct differences in *G. clavigera*-responsive, differentially expressed (DE) chitinases between species ($p \le 0.05$; Fig. 6). At least one contig from classes I–V and VII was significantly DE in *G. clavigera*-infected stem tissues relative to control lodgepole pine, while only classes IV and VII were significantly DE in jack pine. Within lodgepole pine, 33.6% of identified chitinase contigs were significantly DE between control and *G. clavigera*-challenged individuals, while in jack pine

Table 4. Number of functional categories (defined by MAPMAN) represented within each transcriptome and the total number of subcategories per larger functional category.

Mercator bin code	Functional category	Total possible	Transcriptome			
			Jack control	Jack master	Lodgepole control	Lodgepole master
1	Photosynthesis	226	159	156	157	152
2	Cellular respiration	136	123	133	119	131
3	Carbohydrate metabolism	92	86	90	86	89
4	Amino acid metabolism	135	125	131	123	131
5	Lipid metabolism	173	161	169	156	166
6	Nucleotide metabolism	53	49	51	46	51
7	Coenzyme metabolism	158	141	153	144	150
8	Polyamine metabolism	12	9	9	9	9
9	Secondary metabolism	93	61	63	58	63
10	Redox homeostasis	47	44	45	43	45
11	Phytohormones	140	110	112	111	117
12	Chromatin organization	113	103	105	106	107
13	Cell cycle	258	226	234	221	237
14	DNA damage response	67	62	65	64	65
15	RNA biosynthesis	296	266	275	264	275
16	RNA processing	328	275	300	272	304
17	Protein biosynthesis	328	296	303	279	296
18	Protein modification	299	267	289	274	290
19	Protein degradation	187	166	181	160	182
20	Cytoskeleton	107	88	99	90	97
21	Cell wall	126	96	103	100	104
22	Vesicle trafficking	212	194	206	193	203
23	Protein translocation	135	123	130	120	125
24	Solute transport	174	164	167	163	165
25	Nutrient uptake	52	38	38	37	39
26	External stimuli response	111	69	75	69	76
27	Multi-process regulation	38	31	33	31	33
35	Not assigned	2	2	2	2	2
50	Enzyme classification	50	46	46	43	45

Note: Each column for the transcriptome reports the number of subcategories for which at least one transcriptome contig was assigned.

18.1% of contigs were significantly DE. Lodgepole pine showed significant upregulation and downregulation of expression within all chitinase classes. This pattern is different than what was seen in jack pine where only upregulation occurred and was limited to classes IV and VII. This result suggests that the GH-18 chitinases we identified were not involved in jack pine response to *G. clavigera*.

Chitinases from classes I, III–V and VII were significantly DE in response to *Cronartium harknessii* inoculation ($p \leq 0.05$; Fig. 6). Lodgepole pine and jack pine showed contrasting DE chitinase profiles in response to *Cronartium harknessii*. Conversely, 8.2% of lodgepole pine chitinase contigs were significantly DE in response to *Cronartium harknessii* inoculation, 30.9% of jack pine chitinase contigs were significantly DE. In lodgepole pine, when *Cronartium harknessii*-challenged replicates were compared to control replicates downregulated and upregulated classes were mutually exclusive. The downregulated contigs were from classes I, V, and VII, while the upregulated contigs were from classes III and IV. The same comparison in jack pine resulted in a mixed effect.

There were upregulated and downregulated contigs in each represented class except class III (and II since there were no significant changes in those contigs).

Single nucleotide polymorphism genotyping, identification of chitinase SNP, and chitinase genetic diversity

To examine genetic differentiation in individuals, as opposed to pooled data, we mined the large-scale SNP genotyping dataset for chitinase specific loci. A total of 69 lodgepole and jack pine individuals were successfully genotyped (Table 2; available at https://doi.org/10.7939/DVN/A7BWB8). The initial dataset contained 22 003 biallelic SNP. Ninety-six of these SNP were removed because they were suspected to be located within paralogs (Supplementary data Fig. S21).[1] No individuals had an abundance of unique alleles in private allele analyses, suggesting that there was no cross contamination of samples. Lodgepole and jack pine can form hybrids (Cullingham et al. 2012); therefore, we verified species designations (Table 2). Five individuals showed a degree of introgression that exceeded our criteria for designation as

Fig. 4. Summary of putative orthologous contigs in conifer transcriptomes identified via the reciprocal best-hits method. Pairwise comparisons were done between lodgepole pine (A) and jack pine (B) master transcriptomes to published transcriptomes with ≥100 000 contigs in their final assembly. Species abbreviations (clockwise from the top of each fig.): PAGL, *Picea glauca* (Warren et al. 2015); PSME, *Pseudotsuga menziesii* (Little et al. 2016); PIPT, *Pinus patula* (Visser et al. 2015); PIMA, *Pinus massoniana* (Fan et al. 2014); PIAL, *Pinus albicaulis* (Baker et al. 2018). [Colour online.]
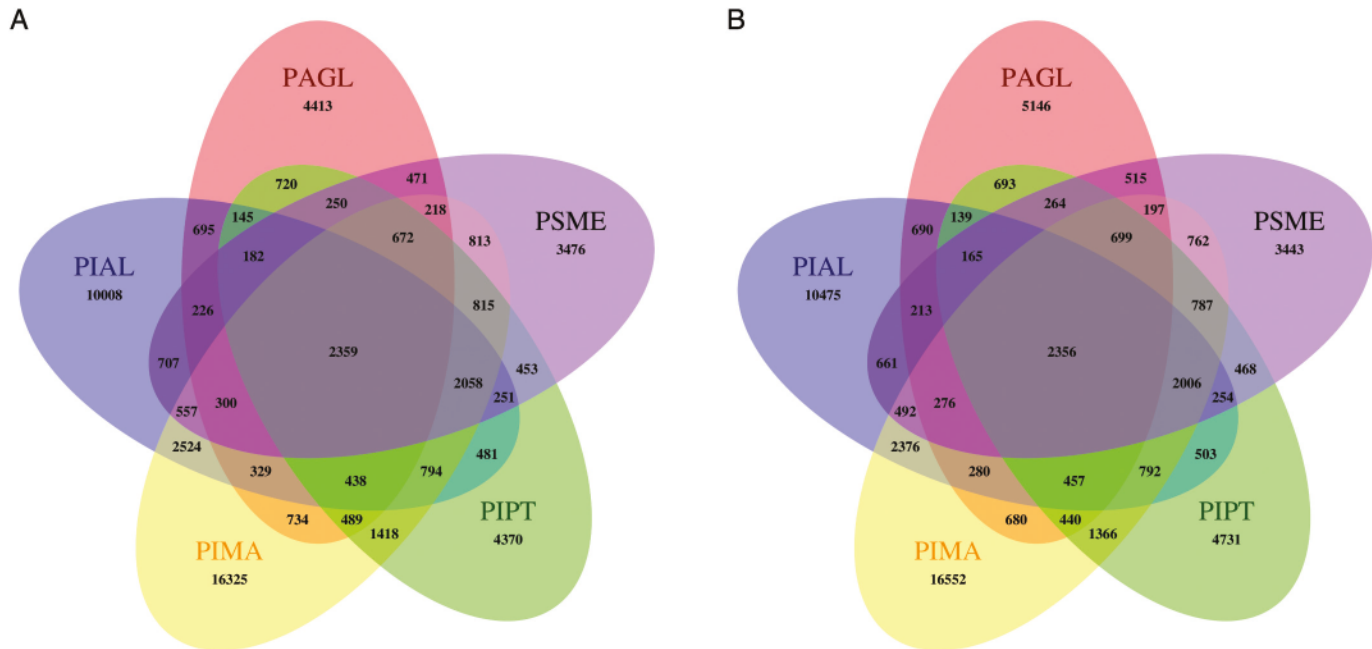
Table 5. Total number of chitinase contigs, the number of unambiguously identified contigs with full length coding sequences (CDS), and measures of selection (κ and ω) per chitinase class within lodgepole pine and jack pine validated through in silico motif identification and phylogenetic analysis.

| Class	Master transcriptome		Validated CDS		κ (ts/tv)	ω (dN/dS)
	Lodgepole pine	Jack pine	Lodgepole pine	Jack pine		
GH-18	21	19	13	9		
III	9	7	8	6	1.8580	0.2824
V	12	12	5	3	1.6361	0.2831
GH-19	113	75	68	54		
I + II	25	19				
I			11	5	2.0506	0.2150
II			6	5	1.2650	0.4313
IV	34	26	19	15	1.4290	0.3043
VII	54	30	32	20	1.8573	0.3801

pure lodgepole or pure jack pine based on a Q value between 0.1 and 0.9 (Cullingham et al. 2012, 2013b) and were removed. As a result of applying these quality control filters, the final dataset that was used for subsequent population genetic analyses consisted of 64 individuals and 21 907 SNP.

We identified 20 putative chitinase SNP loci, 11 of which were within aligned exons and nine were in noncoding flanking sequences. Eight of the 20 SNP loci were located within 10 master transcriptome contigs identified as significantly DE in response to at least one pathogen. Six of the 10 contigs were from lodgepole pine (Fig. 6A),

while the other four were from jack pine (Fig. 6B). Four lodgepole pine SNP loci were located within contigs that were significantly DE when challenged with *G. clavigera*, while the other two were significantly DE when challenged with *Cronartium harknessii*. Three jack pine SNP loci were located within contigs that were only expressed when challenged with *Cronartium harknessii*, while the other was located within a contig that was significantly DE in response to both *G. clavigera* and *Cronartium harknessii*.

Observed heterozygosity among provenances ranged from 0.2967 to 0.325 in lodgepole pine and 0.133 to 0.124 in jack pine. Genetic differentiation (F_{ST}) between

Fig. 5. Chitinase gene family GH-18 (left) and GH-19 (right) cladograms derived from maximum likelihood inference, containing all chitinase transcripts from the two master transcriptomes with conifer and angiosperm outgroups. Accession information for sequences included in phylogenetic analysis (not produced by this study) are available in Supplementary data Table S1.[1] Relationships among contigs and reference sequences were estimated using RAxML on the CIPRES Science Gateway with the WAG substitution model. Number of bootstrap replicates were determined by RAxML with the autoMRE flag. Nodes with less than 50% bootstrap support (BS) were collapsed and BS values for interior nodes are given.

GH-18

GH-19
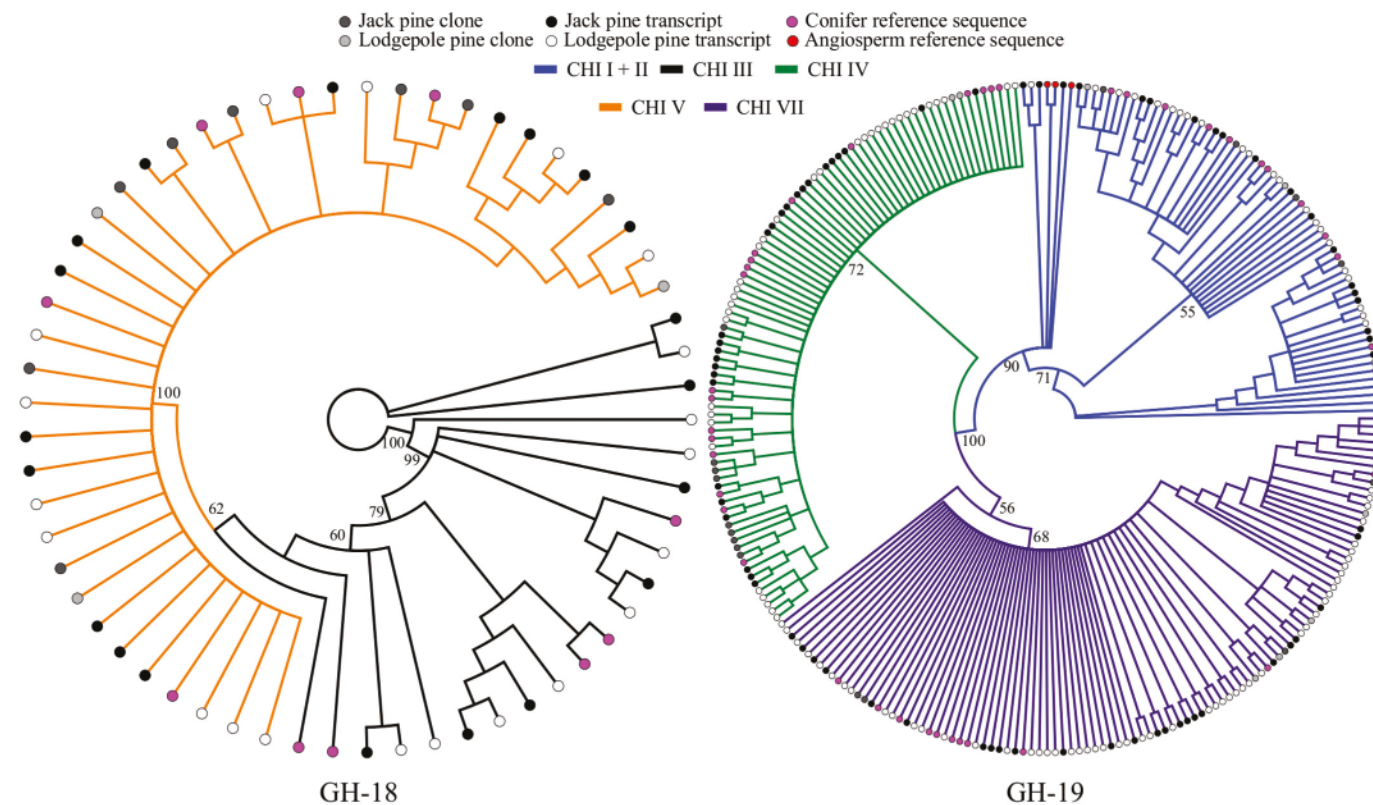
species was 0.281. Single nucleotide polymorphisms located within chitinase loci had an F_{ST} of 0.126. This F_{ST} value was lower than the overall F_{ST} between species, and lower than the mean F_{ST} (0.193) and lowest 5% of F_{ST} values (0.127) for 100 replicates of 20 randomly selected loci from within the remaining 21 894 SNP (Fig. 7). The PCA of chitinase SNP showed a clear distinction between species, but not provenances (Supplementary data Fig. S31).[1] The amount of observed variation within the chitinase SNP in both the PCA and F_{ST} analyses was lower than what was expected by chance (Fig. 7). When per locus F_{ST} values were considered, chitinase SNP were determined not to be outliers, but rather falling within the range of all other loci (Supplementary data Fig. S41).[1]

Discussion

Lodgepole pine and jack pine are sister species with distinct but overlapping geographic ranges (Cullingham et al. 2012; Burns et al. 2019). These species occupy quantifiably different ecological niches, even in the mosaic hybrid zone where the ranges of these species overlap (Cullingham et al. 2012). Consequently, we expect that different environmental pressures, including biotic and abiotic stresses, act as drivers of adaptation. This is particularly true when we consider

scenarios where (*i*) a pathogen shares a long coevolutionary history with one of the pine species but not the other (*G. davigera*), and (*ii*) a pathogen shares a long coevolutionary history with both pine species (*Cronartium harknessii*). These two scenarios of host and pathogen interaction can lead to differences in patterns of landscape level genetic differentiation and different responses in gene expression. In the first scenario, represented with *G. clavigera* experiments, differences in host responses were expected. Responses of a coevolved host to a pathogen may be specific to that pathogen, or alternatively may only be susceptible to certain pathogen genotypes (Lambrechts 2010). We might also observe that the noncoevolved host exhibits a smaller amplitude, delayed, or generalized response. In this case, we predict large F_{ST} values in genes involved in the pathogen response and evidence for balancing selection in the coevolved species that would not be present in the species that has an independent evolutionary history (Allen et al. 2004). Additionally, we predict contrasting expression responses between the two species in response to pathogen challenge. This pattern is the opposite of what is predicted when both hosts share a coevolutionary history, exemplified by *Cronartium harknessii* experiments. In this second scenario, both hosts should have similar genetic responses

Fig. 6. RNA-Seq expression profiling of significantly differentially expressed ($p \leq 0.05$) lodgepole pine (A) and jack pine (B) chitinase transcripts in response to inoculation with either *G. clavigera*, a necrotrophic pathogen, or *Cronartium harknessii*, a biotrophic pathogen. In all cases, $\log_2$-fold-change values, with additional experimental variables as additive in the model, are derived from comparisons of at least four replicate libraries representing inoculated samples compared to their counterpart replicate control samples. The y-axis is ordered by chitinase class determined by phylogenetic and canonical motif identification within master transcriptomes. Superscript "a" next to a class identifier designates contigs with corresponding single nucleotide polymorphism data. Class identifiers with adjacent symbols §, †, ¥, Ω, or ~ indicate contigs identified as orthologous pairs between the two species. Non-significant comparisons are depicted with grey.
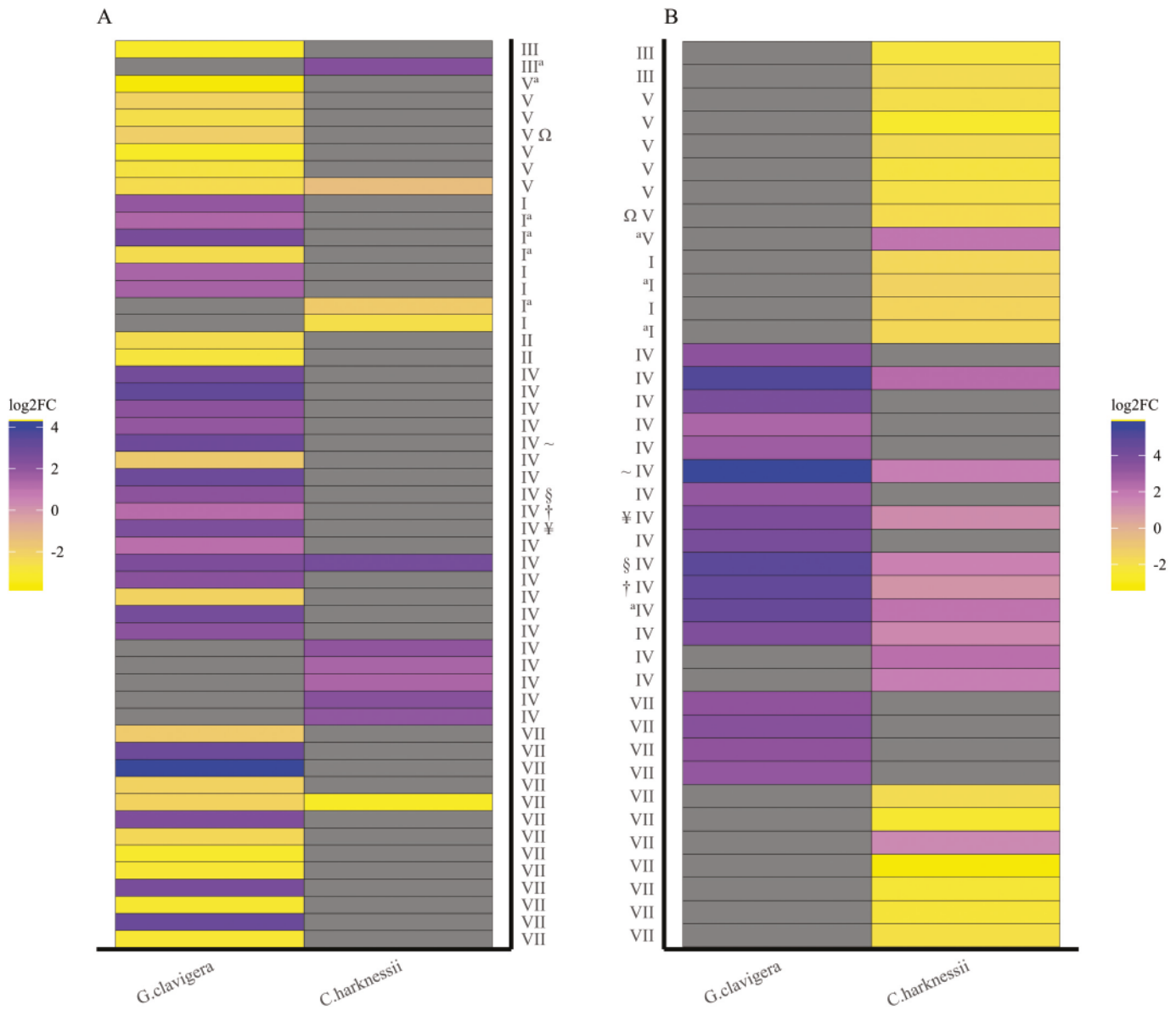
and F_{ST} values for genes involved in pathogen response. Lodgepole and jack pine are thus ideally suited for comparative genomic analyses designed to uncover genotype–phenotype connections for adaptive traits such as host responses to pathogens.

Generating transcriptomes for comparative genomic studies

Vital to any comparative genomics study is the availability of high-quality genomic-scale sequence resources for each species that have been developed in such a way that comparison between them is not subjected to systematic bias. Our analyses of Embryophyta COGs (via BUSCO), annotation, and comparison to existing transcriptomes indicate that lodgepole pine and jack pine master transcriptomes are well assembled with good coverage of coding regions. The degree of over-assembly remaining in our data are likely due to in silico pooling of numerous individual transcriptomes and splice variants. The five Pinaceae species in our RBH comparison ranged in size from 105 454

Fig. 7. Null distribution of pairwise F_{ST} values between lodgepole and jack pines. Replicates were generated from 100 randomized subsets of 20 loci. Dashed line indicates the calculated F_{ST} value between identified chitinase single nucleotide polymorphism loci in lodgepole pine and jack pine.
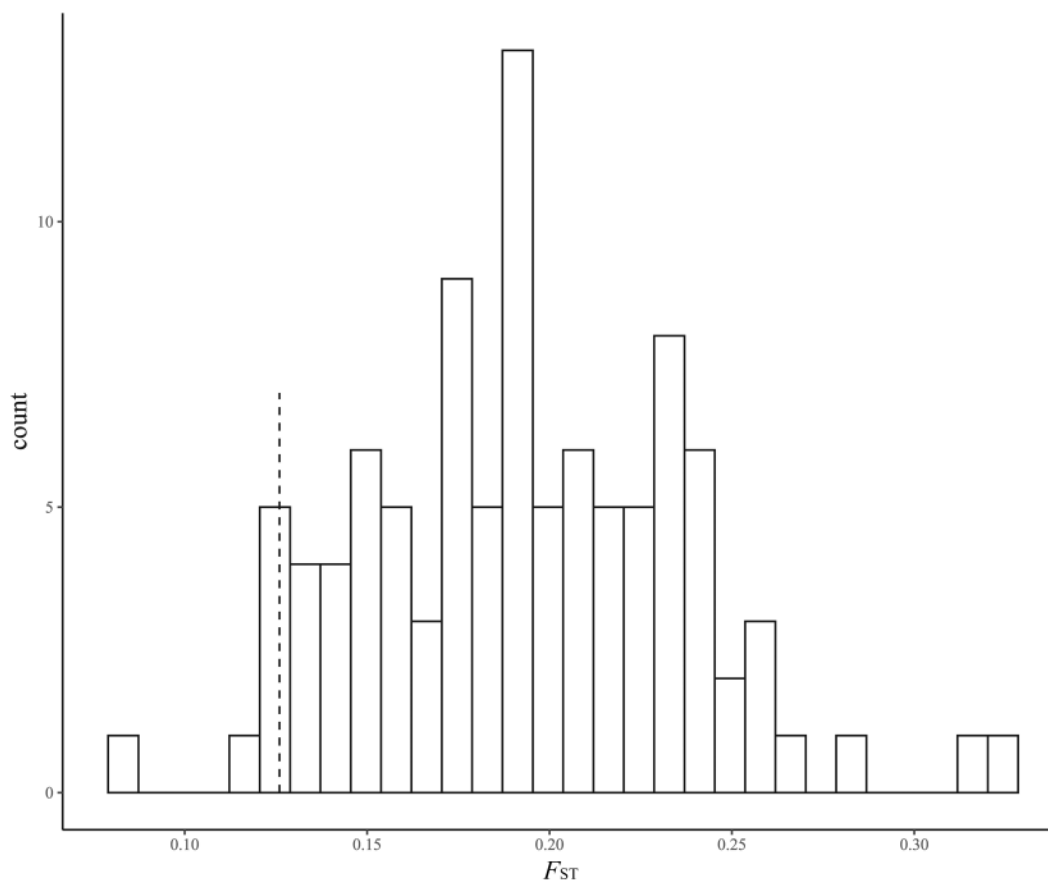
contigs in Mexican weeping pine (Visser et al. 2015) to 455 504 in white spruce (Warren et al. 2015). Our newly generated transcriptome assemblies reported a similar number to the genes, between 32 000 and 50 000, predicted within conifer transcriptomes (Visser et al. 2015).

The four lodgepole and jack pine master transcriptomes reported here represent an advancement from previously available lodgepole and jack transcriptome resources, which were assembled from Sanger, Roche 454 Titanium, Illumina GA, and HiSeq2000 data (Hall et al. 2013). Firstly, our master transcriptomes included more treatments and tree ages. The variety of treatments, together with increased depth of sequencing afforded by using newer technology, enabled us to develop a more comprehensive catalogue of stem-expressed genes. Secondly, the transcriptomes included biological replicates for each of the treatments, enabling RNA-Seq gene expression analyses. Thirdly, our master transcriptomes were developed from 44 genetically unique lodgepole pine individuals and 36 genetically unique jack pine individuals. In contrast, the lodgepole and jack pine transcriptomes published by Hall et al. (2013) were generated from 1–4 individuals depending on sequencing method and species. The large number of individuals that we

included in our transcriptomes made assembly of the master transcriptomes more challenging and has very likely contributed to an inflated number of contigs due to the splitting of alleles representing the same locus. However, the sequence variants embodied in the transcriptomes that arise from including multiple genetically unique individuals in the assembly represent an important resource for future SNP and indel discovery. Additionally, replicates and depth of sequencing allowed us to identify additional low-expression transcripts. The addition of high-quality resources for future SNP discovery and expression analyses is especially important for jack pine, which has fewer available genome level resources from multiple individuals and experimental treatments. However, both these transcriptomes and those reported by Hall et al. (2013) focused solely on stem tissues. Consequently, genes uniquely expressed in other tissues are not represented in either transcriptome resource.

Of the species included in our direct comparisons, Mexican weeping pine is the most closely related species to lodgepole and jack pines (Gernandt et al. 2005). The Mexican weeping pine transcriptome resource was generated from six-month-old seedlings inoculated with *Fusarium circinatum* Nirenberg & O'Donnell and

harvested at 1 dpi. We identified thousands of orthologous contigs between lodgepole/jack pine and Mexican weeping pine. The number of orthologous contigs found between Mexican weeping pine and lodgepole and jack pines was lower than that found between lodgepole/jack and both Chinese red pine (Fan et al. 2014) and whitebark pine (Baker et al. 2018). Chinese red pine had the largest number of putative orthologs to lodgepole/jack pines. The Chinese red pine transcriptome data was generated from three different phosphorous treatments and seedlings were not exposed to any pests/pathogens before harvesting at 22, 34, 58, and 70 days post emergence. It is surprising that this set of transcripts had the highest number of putative orthologs considering the seedlings were never pathogen challenged. It is also surprising because Mexican weeping pine is the more complete (highest number of single copy COGs) and potentially better assembled (higher N50 score) transcriptome in comparison to Chinese red pine. The Mexican weeping pine assembly has the fewest overall contigs so the lower number of identified orthologs could be due to a bias because, like lodgepole and jack pines, the Chinese red pine transcriptome is over-assembled creating a false inflation of RBH.

Lodgepole and jack pine chitinase gene families

A total of 134 lodgepole and 94 jack pine chitinase contigs were identified in the master transcriptome through sequence similarity searches, phylogenetic analyses, and motif evaluation. Similar to what has been described in other studies of conifer chitinases, we identified chitinases from classes I–V and VII in both lodgepole pine and jack pine master transcriptomes. We also found that classes I, II, IV, and VII had signal peptide domains, CBD, and varying presence of loop domains and carboxy-terminal extension. Classes III and V had the GH-18 catalytic domain and, in some contigs, the peptide signaling domain was identified. Through phylogenetic analysis we identified the same phylogenetic groups as presented in Kolosova et al. (2014) and González et al. (2015), although in our analyses chitinases identified as class III were not monophyletic. The lack of recovered monophyly could be due to sampling and not because class III chitinases in these species are not monophyletic.

Norway and white spruce both have expanded chitinase gene families in comparison to *Arabidopsis* (González et al. 2015). Likewise, the copy number evidence presented herein supports the expansion of the chitinase gene family in lodgepole and jack pines as well. The expansion appears to be focused within classes IV and VII. The exact nature of the expansion cannot be explicitly described with these data, however, because master and control transcriptomes analyzed comprise several individuals. This could lead to the possibility that copy numbers are inflated due to allelic diversity as well as physical numbers of genes. The expansion of class IV may prove interesting since there is direct evidence that novel class

IV chitinase gene(s) are correlated with resistance to *Cronartium ribicola* J.C. Fisch in western white pine (*Pinus monticola*; Liu et al. 2011).

Although we achieved deep coverage in our transcriptome sequencing, we did not find evidence of class VI chitinases within lodgepole and jack pine. This is perhaps unsurprising as there is no evidence of class VI chitinases in the few other conifer transcriptomes published to date (summarized in Islam et al. 2010). Chitinase class VI was described originally in sugar beet (*Chia6*) as having a proline-rich hinge and having only half of the CBD (Berglund et al. 1995; Grover 2012). The dissimilarity of the non-CBD sequence precludes the placement of the novel chitinases into previously described classes, and the instances of detected class VI chitinases remain small. The other chitinase variant described as belonging to class VI, *Chic1* (Melchers et al. 1994), has a full CBD (Zhao and Chye 1999). This class could be a catch all for the few newly described chitinase-like genes, specific to angiosperms, resultant from expansion of the family. Alternatively, class VI could be an artificial class that is describing singular evolutionary events within angiosperms. Our study supports earlier reports that there is no evidence of class VI chitinases in conifers.

We determined that GH-18 and GH-19 chitinases required separate phylogenetic analyses due to the lack of homology between GH-18 and GH-19 sequences. By definition, multisequence alignments are hypotheses of homology among sequences (Phillips et al. 2000) and the goal of phylogenetic analysis is to determine the evolutionary relationships among homologous markers. The absence of homologous sites for alignment between GH-18 and GH-19 chitinases, which also have no common catalytic binding domains and only the short signal peptide motif to base homology on, calls into question whether or not the two families should be combined for phylogenetic analysis. Moving forward, identification of class using phylogenetic methods should be limited to clustering algorithms to determine the family. To infer evolutionary relationships and for further phylogenetic inference, GH-18 and GH-19 should be analyzed separately. It would be best to make determinations about evolutionary history of duplications, losses, and other evolutionary processes through ancestral character state reconstruction or comparative phylogenetic methods.

Grosmannia clavigera- and *Cronartium harknessii*-induced chitinase profiles in lodgepole and jack pine: core similarities and differences

In angiosperms, chitinases belonging to every class have been associated with defense responses to filamentous pathogenic fungi (van Loon et al. 2006). This is reflected in their designation as PR proteins, falling into groups PR-3 (classes I–VII), PR-4 (classes I–II), PR-8 (class III), and PR-11 (class I; van Loon et al. 2006). Our RNA-Seq analyses indicate that while chitinases belonging to classes I–V and VII were responsive to fungal pathogens

in lodgepole and (or) jack pine, the majority of the pathogen-responsive chitinases belonged to classes I, IV, and VII. As indicated in the previous section, chitinases belonging to classes I and IV are predicted to have chitinolytic activity, with the key difference being that class I chitinases are often targeted to the vacuole, while class IV chitinases are targeted to the secretory pathway and reside in the apoplastic space (Neuhaus et al. 1991; Dombrowski et al. 1993). In lodgepole pine, upregulation of class I chitinases in response to *G. clavigera* is consistent with previous reports of *G. clavigera*-upregulated class I chitinases in lodgepole pine by Kolosova et al. (2014). The greatest proportion of upregulated chitinases in lodgepole and jack pine in response to both *G. clavigera* and *Cronartium harknessii* were class IV chitinases. Upregulation of class IV chitinases is consistent with numerous studies that have implicated class IV chitinases in defense responses across conifer species (Hietala et al. 2004; Morse et al. 2004; Liu et al. 2005, 2011, 2017; Jøhnk et al. 2005; Myburg et al. 2006; Islam et al. 2010; Kolosova et al. 2014).

Chitinolytic chitinases, like class I and IV chitinases, play key roles in pathogen-associated molecular pattern (PAMP)-triggered immunity (PTI). Chitooligosaccharides (chitin fragments) derived from plant chitinase activity on fungal cell wall chitin serve as effective PAMP (Mengiste 2012; Buendia et al. 2018). PTI invoked by chitooligosaccharides is often referred to as chitin-triggered immunity (CTI) (Espinoza et al. 2017; Volk et al. 2019). Given that many fungal pathogens initiate host invasion via the apoplastic space (Rovenich et al. 2014), constitutively expressed apoplastic class IV chitinases likely generate the initial wave of chitooligosaccharide PAMP, which are detected by pattern recognition receptors at the plasma membrane surface belonging to the LysM receptor-like kinase family (Buendia et al. 2018). The CTI response includes upregulation of additional chitinases, including class I and class IV chitinases that induce lysis of pathogen hyphae (Rovenich et al. 2014). The prevalence of conifer chitinases, particularly class IV chitinases, expressed in secondary stem tissues under constitutive or pathogen-induced conditions reported in this study and others (reviewed in Kolosova and Bohlmann 2012) raises the possibility that CTI is important for conifer defense against woody stem fungal pathogens such as *G. clavigera* and *Cronartium harknessii*. It is plausible that CTI assumes greater importance in conifer woody stem PTI than PTI of herbaceous plants, given that other common PAMP such as cutin monomers are absent from this host environment (Mengiste 2012; Rovenich et al. 2014).

In this study, we examined the responses of chitinases to a necrotroph — *G. clavigera* — and a biotroph — *Cronartium harknessii*. Biotrophs not only actively suppress cell death, they seek to evade host detection through mechanisms that shelter or mask PAMP such as chitooligosaccharides (Ökmen and Doehlemann 2014; Toruño et al. 2016). Based

on this paradigm, the necrotrophic *G. clavigera* would be predicted to elicit a stronger PTI in lodgepole and jack pine than the biotrophic *Cronartium harknessii*. Host responses to necrotrophs are often mediated through JA signaling networks, while responses to biotrophs are often mediated through SA signaling networks (Thomma et al. 1998; Glazebrook 2005; Arango-Velez et al. 2016). There is evidence to suggest that, in conifers, members of the chitinase family are uniquely elicited in response to application of either JA and (or) SA — both important defense-signaling hormones involved in tree host defense response. Exogenous application of JA in slash pine induced expression of class I and IV chitinases, while application of SA induced expression of the same class I and class IV chitinases as well as a class II chitinase (Davis et al. 2002).

The dissimilar response of induction across the chitinases family in response to *G. clavigera* compared to *Cronartium harknessii* observed in both jack and lodgepole pine may reflect activation through distinct defensive signaling pathways. Moreover, the comparatively greater number of *G. clavigera*-upregulated chitinases in both lodgepole and jack pine relative to *Cronartium harknessii*-upregulated chitinases, especially those belonging to class IV, could play an important role in releasing elicitors during necrotrophic pathogen challenge thereby triggering a PTI/CTI response. PTI is quantitative rather than all-or-none, meaning that the efficacy of PTI (and conceivably CTI) in a given pathosystem has a genotype-dependent potential to span the full range of resistance capacity from low to high (Mengiste 2012). It is possible that differential PTI, including CTI, contributes to the documented differences in lodgepole and jack pine susceptibility to *Cronartium harknessii* and *G. clavigera* (Yang et al. 1999; Hall et al. 2013; Clark et al. 2014; Arango-Velez et al. 2014, 2016; Lusebrink et al. 2016; Erbilgin et al. 2017). Jack pine, which is considered to be more resistant to *Cronartium harknessii* than lodgepole pine (Yang et al. 1999), showed a greater number of upregulated class IV chitinases in response to *Cronartium harknessii* inoculation. Interestingly, class I chitinases were downregulated in both lodgepole and jack pine in response to *Cronartium harknessii*, with more class I chitinases being downregulated in the more resistant jack pine than the more susceptible lodgepole pine. This could be a function of the biotroph's suppression of PTI. In the case of *G. clavigera*, the pathogen more effectively colonizes its co-evolved host lodgepole pine in the first few days following inoculation than its new host jack pine, although ultimately both hosts are able to constrain the *G. clavigera* infection (McAllister et al. 2018). In this study, lodgepole pine showed upregulation of both class IV and class I chitinases, while jack pine showed upregulation of class IV chitinases but not class I chitinases. Class I chitinases, which are often localized to the vacuole, become important to the induced defense response if plant cell integrity becomes compromised. A possible explanation for these

different induction patterns is that the co-evolved lodgepole pine can mount a more specialized defense response than jack pine. An alternative but not mutually exclusive explanation is that *G. clavigera* is more effective at cell invasion of its co-evolved host than its new host, i.e., exhibits greater necrotrophic facility, leading to a greater upregulation of defenses designed to protect cellular contents. Ongoing studies are testing these hypotheses.

Relative to the chitinolytic chitinases belonging to classes I and IV, less is known about chitinases belonging to classes V (GH-19) and VII (GH-18), which have only been reported in conifers in the last decade (Islam et al. 2011). Relative to the class I and IV chitinases, the class V and VII chitinases exhibited more complex expression profiles. For example, most of the DE class V chitinases were downregulated in lodgepole and jack pine in response to pathogen challenge, whereas *G. clavigera* resulted in downregulation of more class V chitinases in lodgepole pine and *Cronartium harknessii* resulted in downregulation of more class V chitinases in jack pine. Different class VII chitinases were both up and downregulated by pathogen challenge in the two host species. We found that lodgepole and jack pine class V chitinases have opposite responses to pathogen challenge. In lodgepole pine, several class V chitinases were downregulated when challenged with *G. clavigera*, while no class V chitinases were DE in jack pine. Class V and VII sequences lack chitin binding domains, and thus are not predicted to exhibit chitinolytic activity. Consistent with this prediction, Kolosova et al. (2014) reported a lack of detectable chitinolytic activity for several class V chitinases characterized from interior spruce and lodgepole pine. On the other hand, a class V chitinase gene was characterized as having transglycosylation activity and no direct antifungal activity in the gymnosperm *Cycas revoluta* Thunb. (Taira et al. 2009; Islam et al. 2011). Given that classes V and VII are less characterized than other chitinase classes, these new insights in class V and VII chitinases in responses to pathogen attack may highlight new avenues of investigations regarding the function of the proteins encoded by these genes and their role in tree responses to pathogen attack.

Single nucleotide polymorphism genotyping and population genetics

In a study of western white pine resistance to *Cronartium ribicola*, the causative agent of white pine blister rust, Liu et al. (2011) showed that a class IV chitinase gene exhibited high haplotype diversity, significant genetic differentiation between seedlots with varying degrees of *Cronartium ribicola* resistance, and significant association with the resistance phenotype. Islam et al. (2011) reported within class amino acid sequence similarity from 90% in class I to 37%–98% within classes II and IV chitinases, and Liu et al. (2014) report genetic distances of up to 0.5 among western white pine class IV genes. This pattern of high variability is seen outside the conifers in other gymnosperms (Franceschi et al. 2005; Islam et al. 2011; Liu et al. 2011, 2014), suggesting that this variation is necessary to combat the large number of pathogen pressures hosts face. We investigated genetic differentiation of chitinases using provenance (wild-collected) seed sources from sites across Canada strategically chosen to represent different genetic populations that had previously been reported by Godbout et al. (2005, 2008). Within species, we identified low levels of differentiation between stands of lodgepole pine, but we did not detect population level differentiation in jack pine or the putative chitinase SNP dataset. Similar to previous studies (Parchman et al. 2011; Cullingham et al. 2013a), the variation found in lodgepole pine was evenly distributed across the landscape. Our selection analysis on chitinase contigs, a conservative estimate on consensus sequences, also indicated that there was strong purifying selection acting upon these chitinase genes. Thus, our findings did not support the hypothesis that chitinases would exhibit signatures of diversifying selection.

Conclusions

Our study combined analyses from three molecular datasets and several inquiry methods including comparative transcriptomics, differential gene expression, population genetics, and phylogenetics to compare the genetics underlying defense-related responses in lodgepole pine and jack pine. Focusing on chitinases, we identified putative contigs then confirmed them as genes within the chitinase gene family in both species, and we used these data as the input to determine changes in expression levels in response to pathogen exposure. In addition, we used SNP data to determine the landscape level variation of lodgepole pine and jack pine chitinase genes but were unable to identify signatures of selection for individual chitinase SNP. To investigate the chitinases as a family of genes, we analyzed the chitinase SNP as a group and compared the F_{ST} values of these loci between species to a null distribution. This analysis indicates that the chitinase gene family is constrained and may be under purifying selection. However, the diversity seen in the assembly of the transcriptome implies that overdominance may be a factor and could be contributing to excessive heterozygosity, which could mean increased resistance. Sequencing of genes rather than SNP will be necessary to fully understand the landscape level diversity of chitinases in lodgepole and jack pine.

The large genome sizes of the Pinaceae pose unique challenges in the age of genomics and next-generation sequencing. Data reduction methods such as transcriptome sequencing and genome-wide SNP datasets circumvent the still-intractable problem of creating well-assembled genome sequences for these species, and as such will continue to serve as valuable resources for conifer genomic studies into the foreseeable future. This is particularly true for these non-model forest tree species for which the research community is relatively small, as

is the case for both lodgepole and jack pine. Analyses of the chitinase gene families provided a stringent measure by which to assess the utility and information depth of the transcriptome data. The transcriptome and SNP data sets reported here represent important resources, in large part because of the pressures that North American coniferous forests are facing from MPB and other emerging and existing pathogen systems. Continuing to generate and exploit genome scale data across a diverse sampling of species is key to better understanding conifer response to pathogens in the context of the evolutionary potential of these species on the landscape, with the goal of linking genotype to phenotype for adaptive traits such as disease resistance.

Acknowledgements

References

Adams, A.S., Aylward, F.O., Adams, S.M., Erbilgin, N., Aukema, B.H., Currie, C.R., et al. 2013. Mountain pine beetles colonizing historical and naïve host trees are associated with a bacterial community highly enriched in genes contributing to terpene metabolism. Appl. Environ. Microbiol. 79(11): 3468–3475. doi:10.1128/AEM.00068-13. PMID:23542624.

Allen, R.L., Bittner-Eddy, P.D., Grenville-Briggs, L.T., Meitz, J.C., Rehmany, A.P., Rose, L.E., and Beynon, J.L. 2004. Host-parasite coevolutionary conflict between *Arabidopsis* and downy mildew. Science, 306(5703): 1957–1960. doi:10.1126/science.1104022. PMID:15591208.

Altenhoff, A.M., and Dessimoz, C. 2009. Phylogenetic and functional assessment of orthologs inference projects and methods. PLoS Comput. Biol. 5(1): e1000262. doi:10.1371/journal.pcbi.1000262. PMID:19148271.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. J. Mol. Biol. 215(3): 403–410. doi:10.1016/S0022-2836(05)80360-2. PMID:2231712.

Andrews, S., Lindenbaum, P., Howard, B., and Ewels, P. 2018. FastQC: a quality control tool for high throughput sequence data. Java, Babraham Bioinformatics. Available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Arango-Velez, A., Gonzalez, L.M.G., Meents, M.J., El Kayal, W., Cooke, B.J., Linsky, J., et al. 2014. Influence of water deficit on the molecular responses of *Pinus contorta* x *Pinus banksiana* mature trees to infection by the mountain pine beetle fungal associate, *Grosmannia clavigera*. Tree Physiol. 34(11): 1220–1239. doi:10.1093/treephys/tpt101. PMID:24319029.

Arango-Velez, A., El Kayal, W., 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(4): 726–744. doi:10.1111/pce.12615. PMID:26205849.

Avice, J.-C., Dily, F.L., Goulas, E., Noquet, C., Meuriot, F., Volenec, J.J., et al. 2003. Vegetative storage proteins in overwintering storage organs of forage legumes: roles and regulation. Can. J. Bot. 81(12): 1198–1212. doi:10.1139/b03-122.

Baker, E.A.G., Wegrzyn, J.L., Sezen, U.U., Falk, T., Maloney, P.E., Vogler, D.R., et al. 2018. Comparative transcriptomics among four white pine species. G3. 8(5): 1461–1474. doi:10.1534/g3.118.200257. PMID:29559535.

Baptiste, A. 2017. gridExtra: Miscellaneous Functions for "Grid" Graphics. R. Available at https://CRAN.R-project.org/package=gridExtra.

Berglund, L., Brunstedt, J., Nielsen, K.K., Chen, Z., Mikkelsen, J.D., and Marcker, K.A. 1995. A proline-rich chitinase from *Beta vulgaris*. Plant Mol. Biol. 27(1): 211–216. doi:10.1007/BF00019193. PMID:7865792.

Bolger, A.M., Lohse, M., and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15): 2114–2120. doi:10.1093/bioinformatics/btu170. PMID:24695404.

Buendia, L., Girardin, A., Wang, T., Cottret, L., and Lefebvre, B. 2018. LysM receptor-like kinase and LysM receptor-like protein families: an update on phylogeny and functional characterization. Front. Plant Sci. 9: 1531. doi:10.3389/fpls.2018.01531. PMID:30405668.

Burns, I., James, P.M.A., Coltman, D.W., and Cullingham, C.I. 2019. Spatial and genetic structure of the lodgepole × jack pine hybrid zone. Can. J. For. Res. 49(7): 844–853. doi:10.1139/cjfr-2018-0428.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. 2009. BLAST+: architecture and applications. BMC Bioinformatics, 10(1): 421. doi:10.1186/1471-2105-10-421. PMID:20003500.

Chang, S., Puryear, J., and Cairney, J. 1993. A simple and efficient method for isolating RNA from pine trees. Plant Mol. Biol. Rep. 11(2): 113–116. doi:10.1007/BF02670468.

Chen, H., and Boutros, P.C. 2011. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics, 12(1): 35. doi:10.1186/1471-2105-12-35. PMID:21269502.

Clark, E.L., Pitt, C., Carroll, A.L., Lindgren, B.S., and Huber, D.P.W. 2014. Comparison of lodgepole and jack pine resin chemistry: implications for range expansion by the mountain pine beetle, Dendroctonus ponderosae (Coleoptera: Curculionidae). PeerJ, 2: e240. doi:10.7717/peerj.240. PMID:24688833.

Cock, P.J.A., Chilton, J.M., Grüning, B., Johnson, J.E., and Soranzo, N. 2015. NCBI BLAST+ integrated into Galaxy. Gigascience, 4(1): 39. doi:10.1186/s13742-015-0080-7. PMID:26336600.

Collinge, D.B., Kragh, K.M., Mikkelsen, J.D., Nielsen, K.K., Rasmussen, U., and Vad, K. 1993. Plant chitinases. Plant J. 3(1): 31–40. doi:10.1046/j.1365-313X.1993.t01-1-00999.x. PMID:8401605.

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. Manage. 396: 11–25. doi:10.1016/j.foreco.2017.04.008.

Cui, L., Kerr Wall, P., Leebens-Mack, J.H., Lindsay, B.G., Soltis, D.E., Doyle, J.J., et al. 2006. Widespread genome duplications throughout the history of flowering plants. Genome Res. 16(6): 738–749. doi:10.1101/gr.4825606. PMID:16702410.

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 the boreal forest. Mol. Ecol. 20(10): 2157–2171. doi:10.1111/j.1365-294X.2011.05086.x. PMID:21457381.

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 × jack pine hybrid zone: mosaic structure and differential introgression. Evol. Appl. 5(8): 879–891. doi:10.1111/j.1752-4571.2012.00266.x. PMID:23346232.

Cullingham, C.I., Cooke, J.E.K., and Coltman, D.W. 2013a. Effects of introgression on the genetic population structure of two ecologically and economically important conifer species: lodgepole pine (Pinus contorta var. latifolia) and jack pine (Pinus banksiana). Genome, 56(10): 577–585. doi:10.1139/gen-2013-0071. PMID:24237338.

Cullingham, C.I., Cooke, J.E.K., Dang, S., and Coltman, D.W. 2013b. A species-diagnostic SNP panel for discriminating lodgepole pine, jack pine, and their interspecific hybrids. Tree Genet. Genomes. 9(4): 1119–1127. doi:10.1007/s11295-013-0608-x.

Davies, G., and Henrissat, B. 1995. Structures and mechanisms of glycosyl hydrolases. Structure, 3(9): 853–859. doi:10.1016/S0969-2126(01)00220-9. PMID:8535779.

Davis, J.M., Wu, H., Cooke, J.E.K., Reed, J.M., Luce, K.S., and Michler, C.H. 2002. Pathogen challenge, salicylic acid, and jasmonic acid regulate expression of chitinase gene homologs in pine. Mol. Plant Microbe. Interact. 15(4): 380–387. doi:10.1094/MPMI.2002.15.4.380. PMID:12026177.

Dombrowski, J.E., Schroeder, M.R., Bednarek, S.Y., and Raikhel, N.V. 1993. Determination of the functional elements within the vacuolar targeting signal of barley lectin. Plant Cell, 5(5): 587–596. doi:10.1105/tpc.5.5.587. PMID:8518558.

Duplessis, S., Cuomo, C.A., Lin, Y.-C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., et al. 2011. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. Proc. Natl. Acad. Sci. U.S.A. 108(22): 9166–9171. doi:10.1073/pnas.1019315108. PMID:21536894.

Erbilgin, N., Cale, J.A., Lusebrink, I., Najar, A., Klutsch, J.G., Sherwood, P., et al. 2017. Water-deficit and fungal infection can differentially affect the production of different classes of defense compounds in two host pines of mountain pine beetle. Tree Physiol. 37(3): 338–350. doi:10.1093/treephys/tpw105. PMID:27881799.

Espinoza, C., Liang, Y., and Stacey, G. 2017. Chitin receptor CERK1 links salt stress and chitin-triggered innate immunity in Arabidopsis. Plant J. 89(5): 984–995. doi:10.1111/tpj.13437. PMID:27888535.

Fan, F., Cui, B., Zhang, T., Qiao, G., Ding, G., and Wen, X. 2014. The temporal transcriptomic response of Pinus massoniana seedlings to phosphorus deficiency. PLoS One, 9(8): e105068. doi:10.1371/journal.pone.0105068. PMID:25165828.

Franceschi, V.R., Krokene, P., Christiansen, E., and Krekling, T. 2005. Anatomical and chemical defenses of conifer bark against bark beetles and other pests. New Phytol. 167(2): 353–376. doi:10.1111/j.1469-8137.2005.01436.x. PMID:15998390.

Gauthier, S., Bernier, P., Burton, P.J., Edwards, J., Isaac, K., Isabel, N., et al. 2014. Climate change vulnerability and adaptation in the managed Canadian boreal forest. Environ. Rev. 22(3): 256–285. doi:10.1139/er-2013-0064.

Gernandt, D.S., López, G.G., García, S.O., and Liston, A. 2005. Phylogeny and classification of Pinus. Taxon. 54(1): 29–42. doi:10.2307/25065300.

Glazebrook, J. 2005. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. Annu. Rev. Phytopathol. 43(1): 205–227. doi:10.1146/annurev.phyto.43.040204.135923. PMID:16078883.

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. doi:10.1111/j.1365-294X.2005.02674.x. PMID:16156818.

Godbout, J., Fazekas, A., Newton, C.H., Yeh, F.C., and Bousquet, J. 2008. Glacial vicariance in the Pacific Northwest: evidence from a lodgepole pine mitochondrial DNA minisatellite for multiple genetically distinct and widely separated refugia. Mol. Ecol. 17(10): 2463–2475. doi:10.1111/j.1365-294X.2008.03761.x. PMID:18430147.

González, L.M.G., El Kayal, W., Morris, J.S., and Cooke, J.E.K. 2015. Diverse chitinases are invoked during the activity-dormancy transition in spruce. Tree Genet. Genomes. 11(3): 41. doi:10.1007/s11295-015-0871-0.

Goudet, J., and Jombart, T. 2015. hierfstat: estimation and tests of hierarchical F-statistics. R. Available at https://CRAN.R-project.org/package=hierfstat.

Griffith, M., Antikainen, M., Hon, W.-C., Pihakaski-Maunsbach, K., Yu, X.-M., Chun, J.U., and Yang, D.S.C. 1997. Antifreeze proteins in winter rye. Physiol. Plant. 100(2): 327–332. doi:10.1111/j.1399-3054.1997.tb04790.x.

Groome, M.C., Axler, S.R., and Gifford, D.J. 1991. Hydrolysis of lipid and protein reserves in loblolly pine seeds in relation to protein electrophoretic patterns following imbibition. Physiol. Plant. 83(1): 99–106. doi:10.1111/j.1399-3054.1991.tb01287.x.

Grover, A. 2012. Plant chitinases: genetic diversity and physiological roles. Crit. Rev. Plant Sci. 31(1): 57–73. doi:10.1080/07352689.2011.616043.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8(8): 1494–1512. doi:10.1038/nprot.2013.084. PMID:23845962.

Hall, D.E., Yuen, M.M.S., Jancsik, S., Quesada, A., Dullat, H.K., Li, M., et al. 2013. Transcriptome resources and functional characterization of monoterpene synthases for two host species of the mountain pine beetle, lodgepole pine (Pinus contorta) and jack pine (Pinus banksiana). BMC Plant Biol. 13(1): 80. doi:10.1186/1471-2229-13-80. PMID:23679205.

Heineman, J.L., Sachs, D.L., Mather, W.J., and Simard, S.W. 2010. Investigating the influence of climate, site, location, and treatment factors on damage to young lodgepole pine

in southern British Columbia. Can. J. For. Res. **40**(6): 1109–1127. doi:10.1139/X10-055.

Hietala, A.M., Kvaalen, H., Schmidt, A., Jøhnk, N., Solheim, H., and Fossdal, C.G. 2004. Temporal and spatial profiles of chitinase expression by Norway spruce in response to bark colonization by *Heterobasidion annosum*. Am. Environ. Microbiol. **70**(7): 3948–3953. doi:10.1128/AEM.70.7.3948-3953.2004. PMID:15240268.

Hocking, E. 1971. Preparation and use of a nutrient solution for culturing seedlings of lodgepole pine and white spruce, with selected bibliography. Canadian Forest Service, Department of the Environment, Edmonton, AB, Canada.

Hodgins, K.A., Yeaman, S., Nurkowski, K.A., Rieseberg, L.H., and Aitken, S.N. 2016. Expression divergence is correlated with sequence evolution but not positive selection in conifers. Mol. Biol. Evol. **33**(6): 1502–1516. doi:10.1093/molbev/msw032. PMID:26873578.

Hossain, M.A., Noh, H.-N., Kim, K.-I., Koh, E.-J., Wi, S.-G., Bae, H.-J., et al. 2010. Mutation of the chitinase-like protein-encoding AtCTL2 gene enhances lignin accumulation in dark-grown *Arabidopsis* seedlings. J. Plant Physiol. **167**(8): 650–658. doi:10.1016/j.jplph.2009.12.001. PMID:20056293.

Huang, X. 1999. CAP3: a DNA sequence assembly program. Genome Res. **9**(9): 868–877. doi:10.1101/gr.9.9.868. PMID:10508846.

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., et al. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. **44**(D1): D286–D293. doi:10.1093/nar/gkv1248. PMID:26582926.

Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., von Mering, C., and Bork, P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol. Biol. Evol. **34**(8): 2115–2122. doi:10.1093/molbev/msx148. PMID:28460117.

Islam, M.A., Sturrock, R.N., Williams, H.L., and Ekramoddoullah, A.K.M. 2010. Identification, characterization, and expression analyses of class II and IV chitinase genes from Douglas-fir seedlings infected by *Phellinus sulphurascens*. Phytopathology, **100**(4): 356–366. doi:10.1094/PHYTO-100-4-0356. PMID:20205539.

Islam, M.A., Sturrock, R.N., and Ekramoddoullah, A.K.M. 2011. Conifer chitinases. *In* Journal of Plant Science and Biotechnology. Plant science and biotechnology in North America: Focus on Canada II. *Edited by* K. Al-Taweel and W.G.D. Fernando. pp. 22–36. The Americas Journal of Plant Science and Biotechnology **5**(1): 22–36.

Jøhnk, N., Hietala, A.M., Fossdal, C.G., Collinge, D.B., and Newman, M.-A. 2005. Defense-related genes expressed in Norway spruce roots after infection with the root rot pathogen *Ceratobasidium bicorne* (anamorph: *Rhizoctonia* sp). Tree Physiol. **25**(12): 1533–1543. doi:10.1093/treephys/25.12.1533. PMID:16137939.

Jombart, T., and Ahmed, I. 2011. *adegenet 1.3-1*: new tools for the analysis of genome-wide SNP data. Bioinformatics, **27**(21): 3070–3071. doi:10.1093/bioinformatics/btr521. PMID:21926124.

Kamvar, Z.N., Tabima, J.F., and Grünwald, N.J. 2014. *Poppr*: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ. **2**: e281. doi:10.7717/peerj.281. PMID:24688859.

Kamvar, Z.N., Brooks, J.C., and Grünwald, N.J. 2015. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front. Genet. **6**: 208. doi:10.3389/fgene.2015.00208. PMID:26113860.

Kanehisa, M., Sato, Y., and Morishima, K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J. Mol. Biol. **428**(4): 726–731. doi:10.1016/j.jmb.2015.11.006. PMID:26585406.

Katoh, K. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. **30**(14): 3059–3066. doi:10.1093/nar/gkf436. PMID:12136088.

Kelley, S.T., and Farrell, B.D. 1998. Is specialization a dead end? The phylogeny of host use in *Dendroctonus* bark beetles (Scolytidae). Evolution, **52**(6): 1731–1743. doi:10.1111/j.1558-5646.1998.tb02253.x. PMID:28565330.

Kolosova, N., and Bohlmann, J. 2012. Conifer defense against insects and fungal pathogens. *In* Growth and Defence in Plants. *Edited by* R. Matyssek, H. Schnyder, W. Oßwald, D. Ernst, J.C. Munch, and H. Pretzsch. Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 85–109. doi:10.1007/978-3-642-30645-7_4.

Kolosova, N., Breuil, C., and Bohlmann, J. 2014. Cloning and characterization of chitinases from interior spruce and lodgepole pine. Phytochemistry, **101**: 32–39. doi:10.1016/j.phytochem.2014.02.006. PMID:24564978.

Kozik, A. 2010. Atgc-tools. Python. Available at https://code.google.com/archive/p/atgc-tools/downloads.

Laluk, K., and Mengiste, T. 2010. Necrotroph attacks on plants: wanton destruction or covert extortion? Arabidopsis Book, **8**: e0136. doi:10.1199/tab.0136. PMID:22303261.

Lambrechts, L. 2010. Dissecting the genetic architecture of host–pathogen specificity. PLoS Pathog. **6**(8): e1001019. doi:10.1371/journal.ppat.1001019. PMID:20700450.

Langmead, B., and Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods, **9**(4): 357–359. doi:10.1038/nmeth.1923. PMID:22388286.

Laurie, C.C., Doheny, K.F., Mirel, D.B., Pugh, E.W., Bierut, L.J., Bhangale, T., et al. 2010. Quality control and quality assurance in genotypic data for genome-wide association studies. Genet. Epidemiol. **34**(6): 591–602. doi:10.1002/gepi.20516. PMID:20718045.

Lewontin, R.C., and Krakauer, J. 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics. **74**(1): 175–195. doi:10.1093/genetics/74.1.175. PMID:4711903.

Li, W., and Godzik, A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics, **22**(13): 1658–1659. doi:10.1093/bioinformatics/btl158. PMID:16731699.

Liao, Y., Smyth, G.K., and Shi, W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics, **30**(7): 923–930. doi:10.1093/bioinformatics/btt656. PMID:24227677.

Lieutier, F., Yart, A., and Salle, A. 2009. Stimulation of tree defenses by Ophiostomatoid fungi can explain attack success of bark beetles on conifers. Ann. For. Sci. **66**(8): 801–801. doi:10.1051/forest/2009066.

Little, S.A., Boyes, I.G., Donaleshen, K., von Aderkas, P., and Ehlting, J. 2016. A transcriptomic resource for Douglas-fir seed development and analysis of transcription during late megagametophyte development. Plant Reprod. **29**(4): 273–286. doi:10.1007/s00497-016-0291-9. PMID:27699505.

Liu, J.-J., Ekramoddoullah, A.K.M., and Zamani, A. 2005. A class IV chitinase is up-regulated by fungal infection and abiotic stresses and associated with slow-canker-growth resistance to *Cronartium ribicola* in western white pine (*Pinus monticola*). Phytopathology, **95**(3): 284–291. doi:10.1094/PHYTO-95-0284. PMID:18943122.

Liu, J.-J., Sniezko, R.A., and Ekramoddoullah, A.K.M. 2011. Association of a novel *Pinus monticola* chitinase gene (*PmCh4B*) with quantitative resistance to *Cronartium ribicola*. Phytopathology, **101**(8): 904–911. doi:10.1094/PHYTO-10-10-0282. PMID:21469933.

Liu, J.-J., Chan, D., Sturrock, R., and Sniezko, R.A. 2014. Genetic variation and population differentiation of the

endochitinase gene family in Pinus monticola. Plant Syst. Evol. **300**(6): 1313–1322. doi:10.1007/s00606-013-0963-y.

Liu, J.-J., Williams, H., Li, X.R., Schoettle, A.W., Sniezko, R.A., Murray, M., et al. 2017. Profiling methyl jasmonate-responsive transcriptome for understanding induced systemic resistance in whitebark pine (*Pinus albicaulis*). Plant Mol. Biol. **95**(4–5): 359–374. doi:10.1007/s11103-017-0655-z. PMID:28861810.

Lohse, M.A.R.C., Nagel, A.X.E.L., Herter, T., May, P., Schroda, M., Zrenner, R.I.T.A., et al. 2014. Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data: Mercator: sequence functional annotation server. Plant Cell Environ. **37**(5): 1250–1258. doi:10.1111/pce.12231. PMID:24237261.

Lovett, G.M., Canham, C.D., Arthur, M.A., Weathers, K.C., and Fitzhugh, R.D. 2006. Forest ecosystem responses to exotic pests and pathogens in eastern North America. BioScience. **56**(5): 395–405. doi:10.1641/0006-3568(2006)056[0395:FERTEP]2.0.CO;2.

Lusebrink, I., Evenden, M.L., Blanchet, F.G., Cooke, J.E.K., and Erbilgin, N. 2011. Effect of water stress and fungal inoculation on monoterpene emission from an historical and a new pine host of the mountain pine beetle. J. Chem. Ecol. **37**(9): 1013–1026. doi:10.1007/s10886-011-0008-3. PMID:21874397.

Lusebrink, I., Erbilgin, N., and Evenden, M.L. 2016. The effect of water limitation on volatile emission, tree defense response, and brood success of *Dendroctonus ponderosae* in two pine hosts, lodgepole, and jack pine. Front. Ecol. Evol. **4**: 2. doi:10.3389/fevo.2016.00002.

MacLachlan, I.R., Yeaman, S., and Aitken, S.N. 2018. Growth gains from selective breeding in a spruce hybrid zone do not compromise local adaptation to climate. Evol. Appl. **11**(2): 166–181. doi:10.1111/eva.12525. PMID:29387153.

Mahony, C.R., MacLachlan, I.R., Lind, B.M., Yoder, J.B., Wang, T., and Aitken, S.N. 2020. Evaluating genomic data for management of local adaptation in a changing climate: a lodgepole pine case study. Evol. Appl. **13**(1): 116–131. doi:10.1111/eva.12871. PMID:31892947.

McAllister, C.H., Fortier, C.E., St Onge, K.R., Sacchi, B.M., Nawrot, M.J., Locke, T., and Cooke, J.E.K. 2018. A novel application of RNase H2-dependent quantitative PCR for detection and quantification of *Grosmannia clavigera*, a mountain pine beetle fungal symbiont, in environmental samples. Tree Physiol. **38**(3): 485–501. doi:10.1093/treephys/tpx147. PMID:29329457.

McCarthy, D.J., Chen, Y., and Smyth, G.K. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. **40**(10): 4288–4297. doi:10.1093/nar/gks042. PMID:22287627.

Melchers, L.S., Groot, M.A., van der Knaap, J.A., Ponstein, A.S., Sela-Buurlage, M.B., Bol, J.F., et al. 1994. A new class of tobacco chitinases homologous to bacterial exo-chitinases displays antifungal activity. Plant J. **5**(4): 469–480. doi:10.1046/j.1365-313X.1994.5040469.x. PMID:8012401.

Mengiste, T. 2012. Plant immunity to necrotrophs. Annu. Rev. Phytopathol. **50**(1): 267–294. doi:10.1146/annurev-phyto-081211-172955. PMID:22726121.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler A., et al. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. **37**(5): 1530–1534. doi:101093/molbev/msaa015.

Morse, A.M., Nelson, C.D., Covert, S.F., Holliday, A.G., Smith, K.E., and Davis, J.M. 2004. Pine genes regulated by the necrotrophic pathogen *Fusarium circinatum*. Theor. Appl. Genet. **109**(5): 922–932. doi:10.1007/s00122-004-1719-4. PMID:15221139.

Myburg, H., Morse, A.M., Amerson, H.V., Kubisiak, T.L., Huber, D., Osborne, J.A., et al. 2006. Differential gene expression in loblolly pine (*Pinus taeda* L.) challenged with the fusiform rust fungus, Cronartium quercuum f.sp. fusiforme. Physiol. Mol. Plant Pathol. **68**(1–3): 79–91. doi:10.1016/j.pmpp.2006.07.002.

Myrholm, C.L., and Hiratsuka, Y. 1993. A new method for inoculating jack pine seedlings with the western gall rust fungus *Endocronartium harknessii*. Can. J. Plant Pathol. **15**(1): 29–33. doi:10.1080/07060669309500846.

Neale, D.B., Martínez-García, P.J., De La Torre, A.R., Montanari, S., and Wei, X.-X. 2017. Novel insights into tree biology and genome evolution as revealed through genomics. Annu. Rev. Plant Biol. **68**(1): 457–483. doi:10.1146/annurev-arplant-042916-041049. PMID:28226237.

Neuhaus, J.M. 1999. Plant chitinases (PR-3, PR-4, PR-8, PR-11). *In* Pathogenesis-Related Proteins in Plants. *Edited by* S.K. Datta and S. Muthukrishnan. CRC Press. Boca Raton, FL. pp. 77–105. doi:10.1201/9781420049299.

Neuhaus, J.M., Sticher, L., Meins, F., and Boller, T. 1991. A short C-terminal sequence is necessary and sufficient for the targeting of chitinases to the plant vacuole. Proc. Natl. Acad. Sci. U.S.A. **88**(22): 10362–10366. doi:10.1073/pnas.88.22.10362. PMID:1946457.

Neuhaus, J.M., Fritig, B., Linthorst, H.J.M., Meins, F., Mikkelsen, J.D., and Ryals, J. 1996. A revised nomenclature for chitinase genes. Plant Mol. Biol. Rep. **14**(2): 102–104. doi:10.1007/BF02684897.

Noël, A., Levasseur, C., Le, V.Q., and Séguin, A. 2005. Enhanced resistance to fungal pathogens in forest trees by genetic transformation of black spruce and hybrid poplar with a *Trichoderma harzianum* endochitinase gene. Physiol. Mol. Plant Pathol. **67**(2): 92–99. doi:10.1016/j.pmpp.2005.09.010.

Nsolomo, V.R., and Woodward, S. 1994. Glucanohydrolase enzyme activity in embryos of Scots, Corsican and lodgepole pines infected in vitro with *Heterobasidion annosum*. Forest Pathol. **24**(3): 144–153. doi:10.1111/j.1439-0329.1994.tb00980.x.

Ökmen, B., and Doehlemann, G. 2014. Inside plant: biotrophic strategies to modulate host immunity and metabolism. Curr. Opin. Plant Biol. **20**: 19–25. doi:10.1016/j.pbi.2014.03.011. PMID:24780462.

Parchman, T.L., Geist, K.S., Grahnen, J.A., Benkman, C.W., and Buerkle, C.A. 2010. Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. BMC Genomics. **11**(1): 180. doi:10.1186/1471-2164-11-180. PMID:20233449.

Parchman, T.L., Benkman, C.W., Jenkins, B., and Buerkle, C.A. 2011. Low levels of population genetic structure in *Pinus contorta* (Pinaceae) across a geographic mosaic of co-evolution. Am. J. Bot. **98**(4): 669–679. doi:10.3732/ajb.1000378. PMID:21613166.

Pavy, N., Boyle, B., Nelson, C., Paule, C., Giguère, I., Caron, S., et al. 2008. Identification of conserved core xylem gene sets: conifer cDNA microarray development, transcript profiling and computational analyses. New Phytol. **180**(4): 766–786. doi:10.1111/j.1469-8137.2008.02615.x. PMID:18811621.

Phillips, A., Janies, D., and Wheeler, W. 2000. Multiple sequence alignment in phylogenetic analysis. Mol. Phylogenet. Evol. **16**(3): 317–330. doi:10.1006/mpev.2000.0785. PMID:10991785.

Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics. **155**(2): 945–459. PMID:10835412.

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

Raffa, K.F., and Berryman, A.A. 1982. Physiological differences between lodgepole pines resistant and susceptible to

the mountain pine beetle and associated microorganisms. Environ. Entomol. **11**(2): 486–492. doi:10.1093/ee/11.2.486.

Rigault, P., Boyle, B., Lepage, P., Cooke, J.E.K., Bousquet, J., and MacKay, J.J. 2011. A white spruce gene catalog for conifer genome analyses. Plant Physiol. **157**(1): 14–28. doi:10.1104/pp.111.179663. PMID:21730200.

Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S.D., et al. 2010. *De novo* assembly and analysis of RNA-seq data. Nat. Methods, **7**(11): 909–912. doi:10.1038/nmeth.1517. PMID:20935650.

Robinson, M.D., and Oshlack, A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. **11**(3): R25. doi:10.1186/gb-2010-11-3-r25. PMID:20196867.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, **26**(1): 139–140. doi:10.1093/bioinformatics/btp616. PMID:19910308.

Roe, A.D., Rice, A.V., Bromilow, S.E., Cooke, J.E.K., and Sperling, F.A.H. 2010. Multilocus species identification and fungal DNA barcoding: insights from blue stain fungal symbionts of the mountain pine beetle. Mol. Ecol. Resour. **10**(6): 946–959. doi:10.1111/j.1755-0998.2010.02844.x. PMID:21565104.

Roe, A.D., James, P.M.A., Rice, A.V., Cooke, J.E.K., and Sperling, F.A.H. 2011. Spatial community structure of mountain pine beetle fungal symbionts across a latitudinal gradient. Microb. Ecol. **62**(2): 347–360. doi:10.1007/s00248-011-9841-8. PMID:21468661.

Rovenich, H., Boshoven, J.C., and Thomma, B.P. 2014. Filamentous pathogen effector functions: of pathogens, hosts and microbiomes. Curr. Opin. Plant Biol. **20**: 96–103. doi:10.1016/j.pbi.2014.05.001. PMID:24879450.

RStudio Team. 2017. RStudio: Integrated Development for R. R, RStudio, Inc., Boston, MA. Available at http://www.rstudio.com/.

RStudio Team. 2018. RStudio: Integrated Development for R. R, RStudio, Inc., Boston, MA. Available at http://www.rstudio.com/.

Safranyik, L., and Carroll, A.L. 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 B. Wilson. Canadian Forest Service, Pacific Forestry Centre, Victoria, BC. pp. 3–66. Available at http://www.for.gov.bc.ca/hfd/library/documents/bib96122.pdf.

Safranyik, L., Carroll, A.L., Régnière, J., Langor, D.W., Riel, W.G., Shore, T.L., et al. 2010. Potential for range expansion of mountain pine beetle into the boreal forest of North America. Can. Entomol. **142**(5): 415–442. doi:10.4039/n08-CPA01.

Sánchez-Rodríguez, C., Bauer, S., Hématy, K., Saxe, F., Ibáñez, A.B., Vodermaier, V., et al. 2012. CHITINASE-LIKE1/POM-POM1 and its homolog CTL2 are glucan-interacting proteins important for cellulose biosynthesis in *Arabidopsis*. Plant Cell, **24**(2): 589–607. doi:10.1105/tpc.111.094672. PMID:22327741.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics, **31**(19): 3210–3212. doi:10.1093/bioinformatics/btv351. PMID:26059717.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J.M., and Birol, I. 2009. ABySS: A parallel assembler for short read sequence data. Genome Res. **19**(6): 1117–1123. doi:10.1101/gr.089532.108. PMID:19251739.

Six, D.L. 2012. Ecological and evolutionary determinants of bark beetle—fungus symbioses. Insects. **3**(1): 339–366. doi:10.3390/insects3010339. PMID:26467964.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. **30**(9): 1312–1313. doi:10.1093/bioinformatics/btu033. PMID:24451623.

Sundell, D., Mannapperuma, C., Netotea, S., Delhomme, N., Lin, Y.-C., Sjödin, A., et al. 2015. The plant genome integrative explorer resource: PlantGenIE.org. New Phytol. **208**(4): 1149–1156. doi:10.1111/nph.13557. PMID:26192091.

Suren, H., Hodgins, K.A., Yeaman, S., Nurkowski, K.A., Smets, P., Rieseberg, L.H., et al. 2016. Exome capture from the spruce and pine giga-genomes. Mol. Ecol. Resour. **16**(5): 1136–1146. doi:10.1111/1755-0998.12570. PMID:27428061.

Taira, T., Hayashi, H., Tajiri, Y., Onaga, S., Uechi, G., Iwasaki, H., et al. 2009. A plant class V chitinase from a cycad (*Cycas revoluta*): Biochemical characterization, cDNA isolation, and posttranslational modification. Glycobiology, **19**(12): 1452–1461. doi:10.1093/glycob/cwp119. PMID:19696236.

Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., et al. 2004. mapman: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J. **37**(6): 914–939. doi:10.1111/j.1365-313X.2004.02016.x. PMID:14996223.

Thomma, B.P.H.J., Eggermont, K., Penninckx, I.A.M.A., Mauch-Mani, B., Vogelsang, R., Cammue, B.P.A., and Broekaert, W.F. 1998. Separate jasmonate-dependent and salicylate-dependent defense-response pathways in *Arabidopsis* are essential for resistance to distinct microbial pathogens. Proc. Natl. Acad. Sci. U.S.A. **95**(25): 15107–15111. doi:10.1073/pnas.95.25.15107. PMID:9844023.

Tiffin, P., and Moeller, D. 2006. Molecular evolution of plant immune system genes. Trends Genet. **22**(12): 662–670. doi:10.1016/j.tig.2006.09.011. PMID:17011664.

Toruño, T.Y., Stergiopoulos, I., and Coaker, G. 2016. Plant-pathogen effectors: cellular probes interfering with plant defenses in spatial and temporal manners. Annu. Rev. Phytopathol. **54**(1): 419–441. doi:10.1146/annurev-phyto-080615-100204. PMID:27359369.

van Aalten, D.M.F., Komander, D., Synstad, B., Gaseidnes, S., Peter, M.G., and Eijsink, V.G.H. 2001. Structural insights into the catalytic mechanism of a family 18 exo-chitinase. Proc. Natl. Acad. Sci. U.S.A. **98**(16): 8979–8984. doi:10.1073/pnas.151103798. PMID:11481469.

Van Bel, M., Proost, S., Van Neste, C., Deforce, D., Van de Peer, Y., and Vandepoele, K. 2013. TRAPID: an efficient online tool for the functional and comparative analysis of *de novo* RNA-Seq transcriptomes. Genome Biol. **14**(12): R134. doi:10.1186/gb-2013-14-12-r134. PMID:24330842.

van der Kamp, B.J., and Spence, M. 1987. Stem diseases of lodgepole pine in the British Columbia interior following juvenile spacing. For. Chron. **63**(5): 334–339. doi:10.5558/tfc63334-5.

van Loon, L.C., Rep, M., and Pieterse, C.M.J. 2006. Significance of inducible defense-related proteins in infected plants. Annu. Rev. Phytopathol. **44**(1): 135–162. doi:10.1146/annurev.phyto.44.070505.143425. PMID:16602946.

Veluthakkal, R., and Dasgupta, M.G. 2012. Isolation and characterization of pathogen defence-related class I chitinase from the actinorhizal tree *Casuarina equisetifolia*. For. Pathol. **42**(6): 467–480. doi:10.1111/j.1439-0329.2012.00781.x.

Visser, E.A., Wegrzyn, J.L., Steenkmap, E.T., Myburg, A.A., and Naidoo, S. 2015. Combined *de novo* and genome guided assembly and annotation of the *Pinus patula* juvenile shoot transcriptome. BMC Genomics, **16**(1): 1057. doi:10.1186/s12864-015-2277-7. PMID:26652261.

Volk, H., Marton, K., Flajšman, M., Radišek, S., Tian, H., Hein, I., et al. 2019. Chitin-binding protein of *Verticillium nonalfalfae* disguises fungus from plant chitinases and suppresses chitin-triggered host immunity. Mol. Plant Microbe. Interact. **32**(10): 1378–1390. doi:10.1094/MPMI-03-19-0079-R. PMID:31063047.

Warren, R.L., Keeling, C.I., Yuen, M.M.S., Raymond, A., Taylor, G.A., Vandervalk, B.P., et al. 2015. Improved white spruce (*Picea glauca*) genome assemblies and annotation of large gene families of conifer terpenoid and phenolic defense metabolism. Plant J. **83**(2): 189–212. doi:10.1111/tpj.12886. PMID:26017574.

Wegrzyn, J.L., Lee, J.M., Tearse, B.R., and Neale, D.B. 2008. TreeGenes: a forest tree genome database. Int. J. Plant Genomics. **2008**: 412875. doi:10.1155/2008/412875. PMID:18725987.

Wickham, H. 2016. ggplot2: elegant graphics for data analysis. Springer Verlag New York. Available at https://ggplot2.tidyverse.org.

Woods, A.J., Nussbaum, A., and Golding, B. 2000. Predicted impacts of hard pine stem rusts on lodgepole pine dominated stands in central British Columbia. Can. J. For. Res. **30**(3): 476–481. doi:10.1139/x99-236.

Wu, H.X., Ying, C.C., and Muir, J.A. 1996. Effect of geographic variation and jack pine introgression on disease and insect resistance in lodgepole pine. Can. J. For. Res. **26**(5): 711–726. doi:10.1139/x26-081.

Yang, R.-C., Ye, Z., and Hiratsuka, Y. 1999. Susceptibility of *Pinus contorta – Pinus banksiana* complex to *Endocronartium harknessii*: host–pathogen interactions. Can. J. Bot. **77**(7): 1035–1043. doi:10.1139/b99-095.

Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. **24**(8): 1586–1591. doi:10.1093/molbev/msm088.

Yeaman, S., and Jarvis, A. 2006. Regional heterogeneity and gene flow maintain variance in a quantitative trait within populations of lodgepole pine. Proc. Biol. Sci. **273**(1594): 1587–1593. doi:10.1098/rspb.2006.3498. PMID:16769628.

Yeaman, S., Hodgins, K.A., Lotterhos, K.E., Suren, H., Nadeau, S., Degner, J.C., et al. 2016. Convergent local adaptation to climate in distantly related conifers. Science, **353**(6306): 1431–1433. doi:10.1126/science.aaf7812. PMID:27708038.

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. PMID:21887479.

Zhao, K.-J., and Chye, M.-L. 1999. Methyl jasmonate induces expression of a novel *Brassica juncea* chitinase with two chitin-binding domains. Plant Mol. Biol. **40**(6): 1009–1018. doi:10.1023/A:1006266407368. PMID:10527425.

Zhou, X., Lindsay, H., and Robinson, M.D. 2014. Robustly detecting differential expression in RNA sequencing data using observation weights. Nucleic Acids Res. **42**(11): e91. doi:10.1093/nar/gku310. PMID:24753412.

Zimin, A., Stevens, K.A., Crepeau, M.W., Holtz-Morris, A., Koriabine, M., Marçais, G., et al. 2014. Sequencing and assembly of the 22-Gb loblolly pine genome. Genetics, **196**(3): 875–890. doi:10.1534/genetics.113.159715. PMID:24653210.