the plant journal

SEB Society for Experimental Biology

The Plant Journal (2019) **98,** 745–758 doi: 10.1111/tpj.14270

RESOURCE

Limber pine (*Pinus flexilis* James) genetic map constructed by exome-seq provides insight into the evolution of disease resistance and a genomic resource for genomics-based breeding

Jun-Jun Liu[1,*], Anna W. Schoettle[2], Richard A. Sniezko[3], Fupan Yao[1], Arezoo Zamany[1], Holly Williams[1] and Benjamin Rancourt[1]

[1]*Canadian Forest Service, Natural Resources Canada, Victoria, BC, V8Z 1M5, Canada,*

[2]*USDA Forest Service, Rocky Mountain Research Station, 240 West Prospect Road, Fort Collins, CO, 80526, USA, and*

[3]*USDA Forest Service, Dorena Genetic Resource Center, 34963 Shoreview Road, Cottage Grove, OR, 97424, USA*

Received 23 May 2018; revised 24 December 2018; accepted 28 January 2019; published online 7 February 2019.
*For correspondence (e-mail jun-jun.liu@canada.ca).

SUMMARY

Limber pine (*Pinus flexilis*) is a keystone species of high-elevation forest ecosystems of western North America, but some parts of the geographic range have high infection and mortality from the non-native white pine blister rust caused by *Cronartium ribicola*. Genetic maps can provide essential knowledge for understanding genetic disease resistance as well as local adaptation to changing climates. Exome-seq was performed to construct high-density genetic maps in two seed families. Composite maps positioned 9612 unigenes across 12 linkage groups (LGs). Syntenic analysis of genome structure revealed that the majority of orthologs were positional orthologous genes (POGs) with localization on homologous LGs among conifer species. Gene ontology (GO) enrichment analysis showed relatively fewer constraints for POGs with putative roles in adaptation to environments and relatively more conservation for POGs with roles in basic cell function and maintenance. The mapped genes included 639 nucleotide-binding site leucine-rich repeat genes (*NBS-LRR*s), 290 receptor-like protein kinase genes (*RLK*s), and 1014 genes with potential roles in the defense response and induced systemic resistance to attack by pathogens. Orthologous loci for resistance to rust pathogens were identified and were co-positioned with multiple members of the *R* gene family, revealing the evolutionary pressure acting upon them. This high-density genetic map provides a genomic resource and practical tool for breeding and genetic conservation programs, with applications in genome-wide association studies (GWASs), the characterization of functional genes underlying complex traits, and the sequencing and assembly of the full-length genomes of limber pine and related *Pinus* species.

Keywords: *Cronartium ribicola*, exome-enrichment and sequencing, genomics-based breeding, high-density genetic map, *Pinus flexilis*, single nucleotide polymorphisms (SNPs), white pine blister rust.

INTRODUCTION

Limber pine (*Pinus flexilis*) is a member of the five-needle pine species in the family Pinaceae. It has a wide natural distribution in the mountain regions of western North America, ranging from Mexico to Canada. It is a keystone tree species in forest ecosystems at high elevations, and its seeds provide food for wildlife species, including red squirrels, Clark's nutcrackers, and bears. Unfortunately, limber pine populations have been declining rapidly across portions of their range as a result of white pine blister rust (WPBR), mountain pine beetle infestation, wildfire suppression and succession and climate change (Schoettle and Stritch, 2013), and a continuing decline is expected (Krist *et al.*, 2014). It is listed as an endangered species by the Committee on the Status of Endangered Wildlife in Canada (COSEWIC, 2014). WPBR is caused by the non-native fungal pathogen *Cronartium ribicola* that was inadvertently introduced to western North America around 1910. As a

result of high WPBR-derived mortality in limber pine and other native five-needle pines across their natural distributions, selective breeding and an examination of molecular host–pathogen interactions have been performed, with significant progress in recent years (Sniezko *et al.*, 2014).

Next-generation sequencing (NGS) and NGS-based genotyping platforms enable the detection of a vast number of DNA markers (mainly single-nucleotide polymorphisms, SNPs) in a large number of individuals of non-model species, including conifers, which have very large genomes. Using NGS-based and other high-throughput genotyping platforms, high-density genetic maps have been constructed for several conifer species (Neves *et al.*, 2014; Westbrook *et al.*, 2015; Plomion *et al.*, 2016; Pavy *et al.*, 2017). Exome capture for resequencing (exome-seq) is a suitable method for generating reduced genomic DNA libraries and calling DNA variation in complex conifer genomes (Neves *et al.*, 2013; Suren *et al.*, 2016; Syring *et al.*, 2016). Such tools have enabled the investigation of evolutionary processes of major conifer species through comparative genomic approaches, and have enhanced our understanding of the heritable characteristics of phenotypic traits (Westbrook *et al.*, 2015; Pavy *et al.*, 2017). In limber pine, the lack of such genomic resources (whole-genome linkage group maps, genotype-specific SNP markers, and a reference sequence) has hindered genomic-assisted resistance detection and genetic improvement for conservation and restoration.

Limber pine has exceptional drought and cold tolerance, providing watershed protection in habitats not suitable for other tree species; it can also be very long lived, with a lifespan that can exceed 1000 years. The frequency of major gene resistance (MGR) to WPBR in limber pine was found to be higher in at least some populations than has been documented for other native five-needle pines in North America (Schoettle *et al.*, 2014; Sniezko *et al.*, 2016). The availability of a limber pine genetic map and related genomics resources would greatly facilitate the characterization of gene functions and biological processes underlying phenotypic traits of economic and ecological importance. An understanding of the genetic and molecular basis of these adaptive traits would increase the efficiency of the genetic improvement of limber pine and closely related five-needle pines.

The first comprehensive genomic resource for limber pine, a newly assembled transcriptome, has enabled the *in silico* mining of SNPs in a large number of expressed genes across the genome through comparison of transcriptomes among different genotypes. This work anchored a limber pine *R* gene (*Cr4*) on the *Pinus* consensus linkage group 8 (LG8) by synteny-based comparative mapping (Liu *et al.*, 2016b). Nonetheless, genetic and genomic research on the limber pine has lagged behind the need and potential utility for its breeding and conservation. An increased genomic resource would benefit the characterization and use of genetic diversity in limber pine as well as related five-needle pines. Comparative genetics would benefit limber pine improvement programs by allowing the transfer of information about genes from better characterized species to limber pine, which would then enable an evaluation and utilization of within-species allelic diversity for the development of elite seed families with assembly of the most desirable alleles.

The identification of candidate genes underlying qualitative and quantitative resistance to WPBR, or other pathogens and pests, is of great interest in order to develop durable resistance systems with the integration of a set of *R* genes in breeding programs (Vázquez-Lobo *et al.*, 2017). Genome-wide association studies (GWASs) offer a powerful approach to mine quantitative trait loci (QTLs) and genes contributing to complex traits of interest, but they require reference sequences or genetic maps with high-density DNA markers on individual LGs (Bartholomé *et al.*, 2016). SNP markers are now widely used in LG mapping because of their high abundance throughout all genomes and the relatively low cost of genotyping with high-throughput methods (Ganal *et al.*, 2009). Such SNP-based LG maps increase our knowledge of genome architecture and facilitate the fine genetic dissection of QTLs (Mammadov *et al.*, 2012; Rasheed *et al.*, 2017). In addition, the availability of LG maps facilitates map-based gene cloning and the functional characterization of genes conferring host resistance or other adaptive traits. Ultra-high-density genetic maps provide a genomic tool for comparative map analysis, anchoring genome sequences to chromosomes, and assisting genome assembly, therefore bridging a gap in the current knowledge of conifer genomics.

The present study aimed to construct ultra-high-density genetic maps of limber pine by positioning polymorphic genes through an exome-seq approach. The composite map constructed in two seed families positioned 9612 expressed genes, representing one of the most informative genetic maps with the highest number of mapped genes in conifers. Mapped genes were annotated, and included 639 nucleotide-binding site leucine-rich repeat genes (*NBS-LRR*s), 290 receptor-like protein kinase (*RLK*) genes and 1014 defense-related genes with coordinated expression in response to biotic stresses. The identification of genetic bins across all 12 LGs presents *R* candidates with putative functions in resistance against WPBR in limber pine and other five-needle pines. The limber pine genetic map was compared with the latest maps of other conifers, enhancing current knowledge of the architecture and evolution of conifer genomes, and demonstrating limber pine genomics resources to be transferable and useful for other close species.

RESULTS

Genotyping by exome-seq and SNP analysis

Megagametophyte genomic DNA (gDNA) samples were barcoded in 48-plex for exome enrichment and sequencing. A total of 91 and 99 samples were sequenced on an Illumina HiSeq platform for field-collected seed from open-pollinated families LJ-112 and PHA-106 (Table S1), respectively. After filtering for quality control and demultiplexing, a total of 850.80 million 100 paired-end reads were generated. Clean reads in individual samples ranged from 1.84 million to 6.62 million. Seed family PHA-106 had higher average reads (4.94 million per sample) than seed family LJ-112 (3.98 million per sample) (Table S1). As no genome sequence was available, the reads were mapped against the limber pine reference transcriptome that was used to generate the hybridization probes for gDNA enrichment. The coverage depth of target sequences was assessed across all 190 limber pine haploid megagametophyte samples: 88, 75 and 52% of the target sequences had a minimum 5×, 10×, and 20× coverage depth, respectively (Figure S1).

Limber pine consensus maps

The SNP data sets from exome-seq analysis were subjected to quality control by filtering out SNPs highly distorted from the expected Mendelian segregation ratio of 1 : 1 ($\alpha \leq 0.01$). Because segregation distortion is a common phenomenon in many plant species, we included weakly distorted SNPs (α ranging from 0.01 to 0.05) in our map construction. Following SNP data filtering, a total of 60 591 and 38 328 SNP markers were used for map construction in the seed families PHA-106 and LI-112, respectively. The called SNPs had 27× and 36× average depth in seed family LJ-112 and PHA-106, respectively. A total of 190 samples from the two seed families were used to construct the LG map (Table S1).

The LG maps were first constructed for the two mapping seed families separately using two runs of LEPMAP 2. The first run of mapping analysis using SNPs with less than 10% of missing data positioned 30 811 and 18 887 SNP loci across 12 LGs in seed families PHA-106 and LI-112, respectively. The LG number corresponded to a well-known haploid chromosome number for most *Pinus* species. Mapped SNPs were distributed within 7142 and 4550 unique gene sequences encoding proteins, and an average of 4.23 SNP loci per gene were mapped (Table S1). These mapped genes resided at 1461 and 1149 unique positions across 12 LGs, with total genetic map lengths of 1625.809 cM and 1782.985 cM in seed families PHA-106 and LI-112, respectively (Tables S2 and S3). LG9 was split into two subgroups in family LI-112 because of a large gap. The largest gap distance in LGs ranged from 4.489 cM (LG3 for family PHA-106) to 17.113 cM (LG10 for family LJ-112). The average gap between two adjacent unique positions was much shorter in family PHA-106 than in family LJ-112 (1.11 cM versus 1.55 cM). The gap size distribution showed 74 and 82% of gap distances less than 2 cM in family LJ-112 and PHA-106, respectively (Figure S2). On average about four genes resided at each unique position. To evaluate the consistency and accuracy of the individual LG maps between two seed families, the order and positions of those shared genes were compared along each LG by Pearson correlation analysis. A high coefficient ($R^2 = 0.963 \pm 0.029$) confirmed strong agreement between corresponding LGs constructed independently (Figure S3).

A composite map was generated by merging the LGs of two seed families using LPMERGE. A total of 8780 functionally expressed genes were assigned across 12 LGs and positioned at 1554 unique loci on the composite map, which had an average gap length of 1.14 cM among adjacent unique positions (Table S2). The composite LG lengths varied approximately twofold in a range from 102.00 cM (LG11) to 217.01 cM (LG2), and a total genetic map length of 1775.24 cM was calculated, with an average of about five genes per cM (Figure 1; Tables S2 and S3).

Single-nucleotide polymorphism (SNP) markers with >10% of missing data were added in the second run of mapping analysis, revealing few changes of marker order and positions compared with the first run of mapping analysis (Table S4). Integrating them into the mapping analysis resulted in the assignment of a total of 9612 polymorphic genes across 12 limber pine LGs (Table S3).

Synteny analyses and evolutionary characterization of the limber pine genome

Reciprocal best-hit Blastn analysis (Blastn *E* value < e^{-100}) identified 1771, 1687 and 3507 mapped orthologous gene pairs between the *P. flexilis* (Pifl) map and the previously reported genetic maps of conifer species *Pinus pinaster* (maritime pine; Pipi), *Pinus taeda* (loblolly pine; Pita) and *Picea glauca* (white spruce; Pigl), respectively (De Miguel *et al.*, 2015; Westbrook *et al.*, 2015; Pavy *et al.*, 2017). These paired orthologs were used as bridging sequences for genetic map comparison. Of these limber pine bridging genes, 89.67–92.59% were mapped on the homologous chromosomes (LGs) of maritime pine, loblolly pine or white spruce (Table S4). Intergenome synteny analyses showed good collinearity between maps of limber pine and the other three conifers. The majority of bridging genes showed a near-linear arrangement along all 12 corresponding LGs (Figure 2). Similar to previous reports (Pavy *et al.*, 2017), neither segmental transposition nor other major rearrangements were observed. In Pearson correlation analysis of gene positions along homologous chromosomes, the average R^2 ranged from 0.922 to 0.946 across the 12 LGs in all three paired comparisons (Table S4). Homologous chromosomes showed the lowest conservation level between

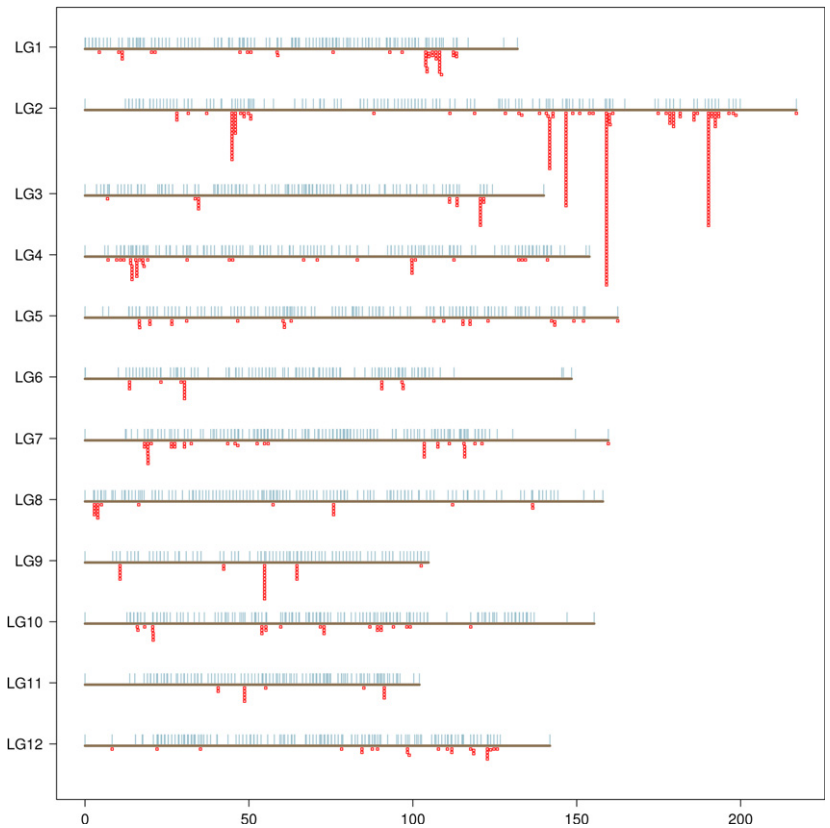

Figure 1. Illustration of the limber pine composite genetic map of linkage groups (LGs). Horizontal gray lines represent all 12 LGs. Black bars indicate the relative gene/marker positions, and red boxes below each LG indicate the positions and clusters of putative *NBS-LRR* genes. The *x*-axis represents LG length in centiMorgans (cM) and the *y*-axis represents LG numbers.

*Pifl*_LG12 and *Pigl*_LG4 (R^2 = 0.858) and between *Pifl*-LG3 and *Pigl*-LG11 (R^2 = 0.866). These R^2 values were significantly lower than the average R^2 values across 12 LGs (*z*-score test $P < 0.05$). The majority of orthologous genes mapped on the homologous LGs retained their relative ancestral genomic positions and were considered as positional orthologous genes (POGs) between compared species (Dewey, 2011). In contrast, orthologous genes that mapped on different LGs were termed as non-positional orthologous genes (nPOGs) and accounted for the remaining 7.41–10.33% of the total bridge genes.

Reciprocal best-hit Blastn analysis showed no significant differences for both the greatest bit scores and the greatest hit lengths between POGs and nPOGs. As compared with all mapped genes, functional analyses showed no significant enrichment of biological process for nPOGs, whereas POGs showed nine under- and 45 over-represented biological processes (Figure S4). The over-represented gene ontology (GO) terms included metabolic process (GO 0008152), biosynthetic process (GO 0009058), and protein localization (GO 0008104), whereas the under-represented GO terms included signaling (GO 0023052), response to stimulus (GO 0050896) and regulation of biological processes (GO 0050789).

Synonymous substitution rates (K_s), non-synonymous substitution rates (K_a) and K_a/K_s ratios were analyzed to infer genome evolution (Figure 3). Mean K_s values were calculated as 0.145 versus 0.185, 0.131 versus 0.223 and 0.205 versus 0.281 for POG and nPOG pairs in the syntenic analysis of *Pifl* with *Pita*, *Pipi* and *Pigl*, respectively. K_s values showed significant differences between POGs and nPOGs for comparisons of *Pifl* with the three other conifer species. As expected, the K_s values were significantly higher for the comparison of *Pifl* versus *Pigl* than for comparisons among pine species, but no significant difference was detected among the pine species [one-way analysis of variance (ANOVA) with Tukey's honestly significant difference (HSD) test, $P < 0.05$; Figure 3a]. Similarly, the K_s distribution of POGs showed a K_s peak, with a higher K_s value for *Pifl* versus *Pigl* (K_s peaked at ~0.175) than for *Pifl* versus *Pita* or *Pipi* (K_s peaked at ~0.075) (Figure 3b), giving average synonymous substitution rates (μs) of 1.15×10^{-9} and 1.25×10^{-9} substitutions per site per year for subgenera *Pinus–Strobus* and *Pinus–Picea* divergence, assuming average divergence times of 65 and 140 Mya, respectively (Willyard *et al.*, 2007; Buschiazzo *et al.*, 2012). When mean K_s values across 12 different LGs were analyzed in a syntenic comparison of *Pifl* versus *Pigl*, *Pigl*-LG3 had the highest K_s mean value (0.2179 ± 0.0192), significantly higher than five other LGs (LG5, LG8–LG11), followed by *Pifl*-LG12 (0.2141 ± 0.0197), with significantly higher K_s values than LG8 and LG10 (Student's *t*-test $P < 0.05$). This larger

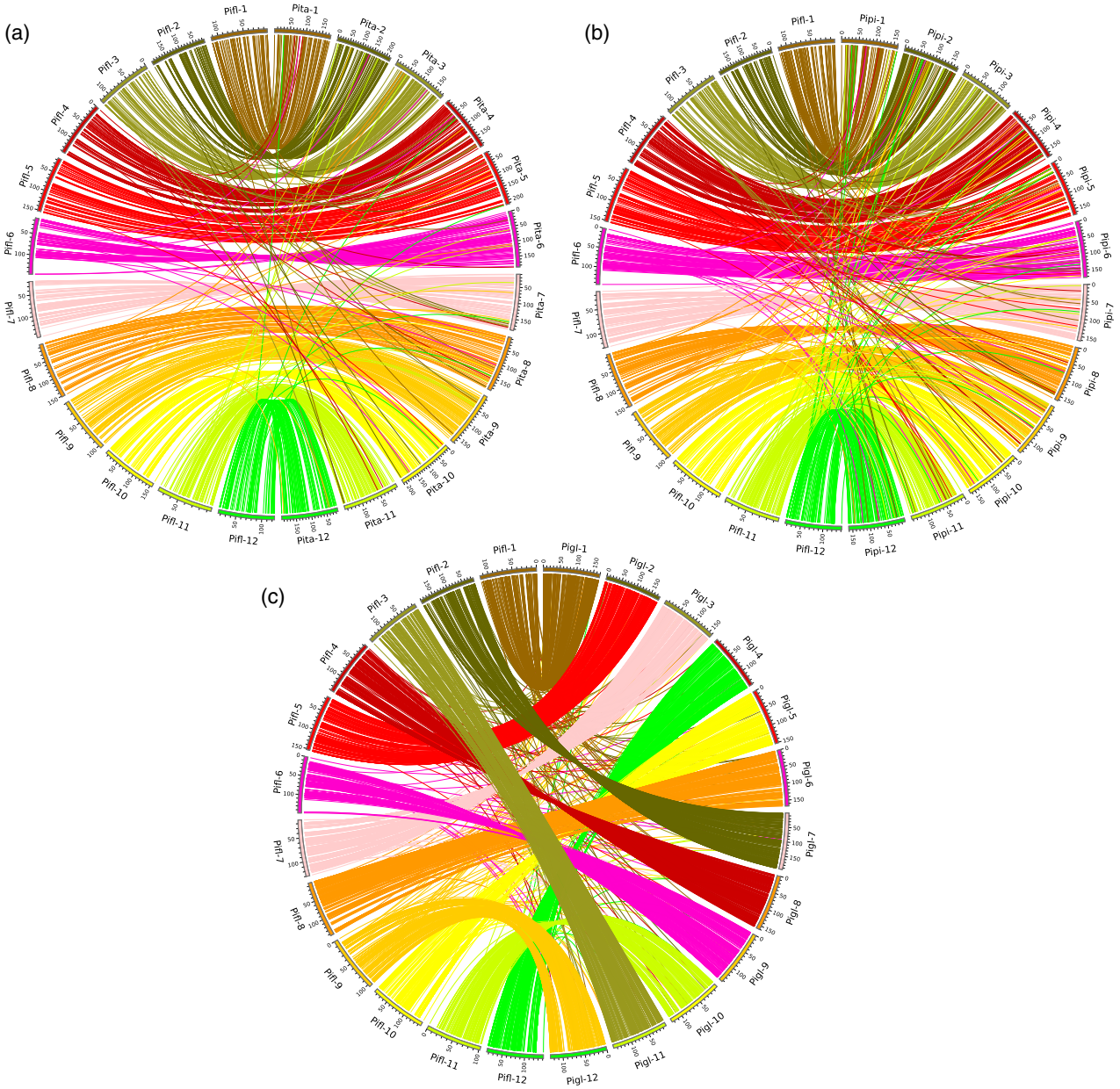

Figure 2. Syntenic relationships of *Pinus flexilis* (Pifl_LG01–LG12) with *Pinus taeda* (Pita_LG01–LG12) (a), *Pinus pinaster* (Pipi_LG01–LG12) (b) and *Picea glauca* (Pigl_LG01-LG12) (c) linkage groups (LGs). Synteny between limber pine and three other conifer species is based on the alignment of bridging genes on the individual genetic maps of each species. Lines with the same color represent bridging genes connected between corresponding LGs.

divergence of *Pifl*-LG3 and -LG12 compared with other LGs provides additional evidence for much less correlation with POG order and position on these two LGs as observed between pine and spruce (Table S4).

The K_a/K_s ratios were significantly lower than one (Fisher's exact test $P < 0.05$) for the majority of ortholog pairs (including POGs and nPOGs): 73.6, 92.3 and 95.8% for *Pifl* versus *Pita*, *Pipi* and *Pigl*, respectively. Purifying selection of these orthologs contributes to the stability of basic biological processes such as the metabolic process and the biosynthetic process, as revealed by GO enrichment analysis. The remaining ortholog pairs (26.4, 7.7 and 4.2%) showed K_a/K_s ratios not significantly different from one, and only one pair (M249871 versus sp_v3.0_unigene17852) had a K_a/K_s ratio significantly larger than one. Moreover, POGs showed significantly higher K_a/K_s ratios than nPOGs in comparisons with *Pifl* versus the other three conifer species (Student's *t*-test, $P < 0.05$; Figure 3c), demonstrating that POGs are subject to stronger diversifying selection and a faster evolutionary rate than the nPOGs.

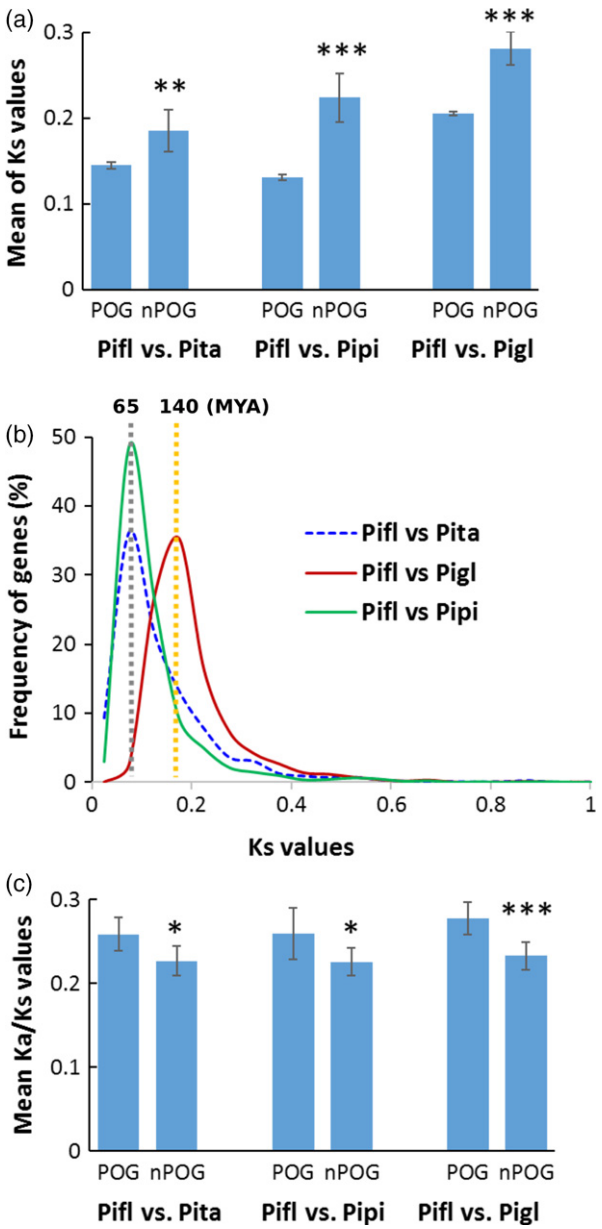

Figure 3. Divergence history of orthologous genes in limber pine and related conifers. Mapped orthologous genes were identified by reciprocal Blastn analysis between *Pinus flexilis* (Pifl) and the other three species: *Pinus pinaster* (Pipi), *Pinus taeda* (Pita) and *Picea glauca* (Pigl). (a) Comparisons of mean synonymous substitution levels (K_s) between positional orthologous genes (POGs) and non-positional orthologous genes (nPOGs). (b) Frequency distribution of K_s values. The *x*- and *y*-axes indicate K_s values and POG frequency, respectively. Three distributions are presented to depict the divergence of limber pines from the other three conifer species. (c) The non-synonymous substitution rate (K_a)/K_s ratios of orthologous genes. Error bars represent sample variance. Differences were analyzed by Student's *t*-test and one-way analysis of variance (ANOVA) with Tukey's honestly significant difference (HSD) test: $*P < 0.05$; $**P < 0.01$; $***P < 0.001$.

Annotation of limber pine mapped genes

A total of 9612 mapped limber pine gene sequences were used for Blastp analysis, and 94.4% of them had significant homology hits (*E*-value cut-off at $1e^{-5}$) against the *Arabidopsis thaliana* (TAIR10) database (Table S3). Similarly, Blastp analysis showed 79.8 and 94.5% of mapped limber pine genes with either identical hits (*E* value $< 1e^{-100}$) or significant homology hits (*E* value $< 1e^{-5}$), respectively, as the putative proteome of sugar pine (*Pinus lambertiana*). In addition, Blastp searches against the NCBI nr database and western white pine (*Pinus monticola*) homologs of *R* gene families (Liu and Ekramoddoullah 2007; Liu *et al.*, 2014) identified 639 and 290 gene sequences encoding putative NBS-LRR and RLK proteins, respectively (Figure 1; Tables S5 and S6). Among them, one NBS-LRR gene (*M287456*) and one RLK gene (*M236700*) were detected close to the *Cr4* locus. Genes of the same family show non-random distribution and many of these genes tend to localize in a tandem pattern along chromosomes. About 42% of the putative NBS-LRR genes were grouped into genetic bins with at least four homologs positioned at identical chromosomal regions (Table S7). A total of 29 genetic bins were detected for the NBS-LRR family across 10 LGs, with the exception being LG5 and LG10. LG2 positioned 266 NBS-LRR genes with the highest bin number at nine, followed by LG4, LG7, LG8 and LG9 with three bins. LG1, LG3 and LG11 were each assigned two NBS-LRR bins, and LG6 and LG12 each harbored only one NBS-LRR bin (Figure 1; Table S5).

Evolutionary relationships among mapped NBS-LRR genes were further examined by phylogenetic analysis based on alignment of the NBS domains with at least 150 amino acids (Figure 4). A total of 176 sequences were included and grouped in two types: NBS-LRR proteins with putative N-terminal coiled-coil (CC) domain (CNLs) or with putative N-terminal Toll interleukin 1 receptor (TIR)-like domains (TNLs). A 1 : 1 ratio of CNL (87 sequences) to TNL (89 sequences) was observed with CNL and TNL separating into eight (CNL_C1-C8) and nine clades (TNL_C1-C9), respectively. Within these phylogenetic clades, about 25% of paralogs were mapped at the same positions of the same LGs, whereas others were mapped at different positions of the same LGs or on different LGs. Most TNLs were distributed on LG2, whereas CNLs were more widely and evenly distributed across all LGs. When K_s values were used to assess duplication events using paralog pairs inside the phylogenetic clades, the mean K_s values were significantly higher among misaligned CNLs on different LGs than among aligned CNLs on the same LGs, but the mean K_s values for the latter were similar to TNLs (either misaligned or aligned) without any significant difference (Figure S5). These relationships among phylogenetic clades, LG locations and K_s values suggest that the limber pine CNLs may have evolved earlier than others, and have undergone diversification through more ancient large-scale gene duplication, whereas TNLs and aligned CNLs have been diversified through more recent local gene duplication.

The NBS-LRR loci for genetic resistance to rust fungi showed syntenic relationships between limber pine and closely related *Pinus* species. *R* candidates of *NBS-LRR* genes were recently identified for sugar pine *Cr1* (Stevens *et al.*, 2016), loblolly pine *Fr1* (Neale *et al.*, 2014) and western white pine *Cr2* (Liu *et al.*, 2017b). Their orthologous loci were placed on limber pine LG2 at 45.87 cM (M503518) and 159.09 cM (M278749), and LG1 at 11.35 cM (M400228), respectively. *Cr1*, *Cr2* and *Fr1* orthologous loci were detected as genetic bins with co-localization of 7, 3 and 52 NBS-LRR sequences (Table S1). Both CNL and TNL sequences were mapped at the *Cr1* orthologous locus. Among them, *Cr1* candidate and limber pine paralogs M362534 showed K_a/K_s ratios of 1.4545 and 1.1540, respectively (Figure S6), suggesting that at least some NBS-LRR genes may be under selection to develop novel resistance

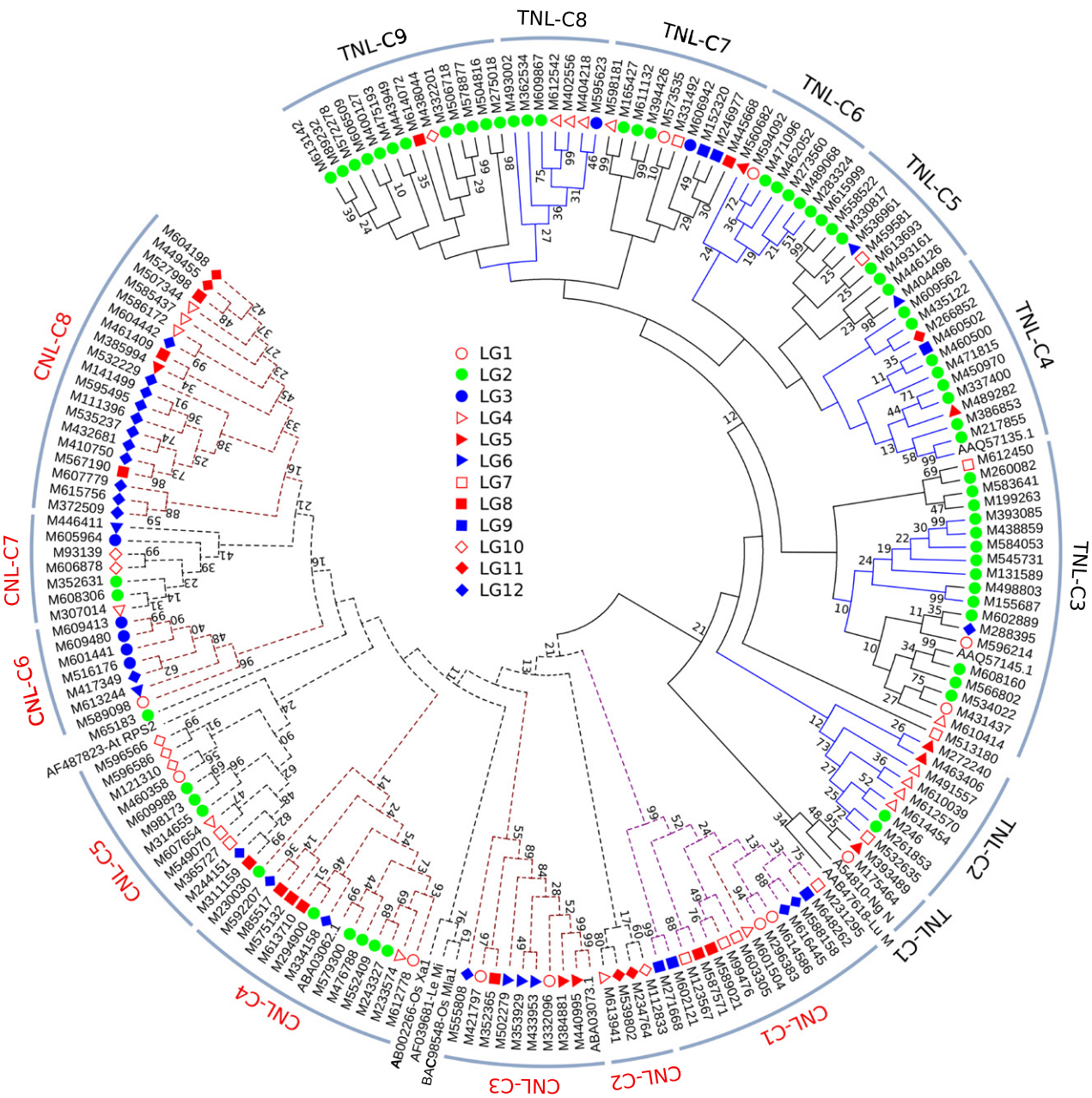

Figure 4. Phylogenetic relationships of mapped *NBS-LRR* genes based on the alignment of putative NBS domains. A total of 176 limber pine intact NBS domains were included in the phylogenetic tree: 87 assigned to eight clades (C1–C8) of the putative *CC-NBS-LRR* (*CNL*) genes and the other 89 sequences assigned in nine clades (C1–C9) of the putative *TIR-NBS-LRR* (*TNL*) genes. Bootstrap values of 1000 replications are shown at branch points. A few angiosperm *R* genes and western white pine *NBS-LRR* genes included in the analysis are indicated by their GenBank accession numbers.

specificity, even though most members of the limber pine NBS-LRR family showed strong evolutionary constraints, with K_a/K_s ratios significantly lower than one.

Seven genetic bins with at least four members were detected for the *RLK* genes (Table S7). In addition to the clustering of paralogs with sequence similarities, we also observed the clustering of defense-related genes with similar expression patterns inside genetic bins across the different LGs. Blast analysis revealed a total of 1013 mapped limber pine genes of different families with identical hits to *Pinus* genes (Blastn with best match at *E* values of $<e^{-100}$). These genes showed defense-responsive expression (Tables S8–S10), including 223 in response to pine weevil attack (Kovalchuk *et al.*, 2015), 253 in response to *C. ribicola* infection (Liu *et al.*, 2013) and 748 in response to methyl jasmonate (MeJA) treatment with potential roles in induced systemic resistance to pathogens and pests (Liu *et al.*, 2017b). A total of four, four and 25 genetic bins with at least four genes were identified for these three groups of defense-related genes under coordinated expression control (Table S7). Notably, 10 genetic bins harbored gene groupings under different categories. For example, the two bins (at 44.84 cM of LG2 and at 54.77 cM of LG9) contained genes of the NBS-LRR family, as well as genes with MeJA-, weevil- and WPBR-responsive expression patterns. The genes with MeJA-, WPBR- and pine weevil-responsive expression patterns were enriched in the bin at 76.23 cM on LG12 (Table S7).

DISCUSSION

High-density LG map as a genomic resource for limber pine and related conifer species

We constructed a high-density genetic map for limber pine, a threatened conifer, using exome-seq. The application of advanced NGS technologies and bioinformatics allows the sequencing of genomes and transcriptomes and high-throughput genotyping of large populations, facilitating the construction of high-density linkage maps. Exome-seq is an NGS-based SNP genotyping technology that reduces genome complexity through the enrichment of gene sequences that can be functionally transcribed. It is a powerful approach for calling short nucleotide variations (SNPs and indels) from protein-coding regions, introns, untranslated and flanking regions, and intergenic spaces, and has been widely applied in SNP discovery, population genetics and the construction of genetic maps in non-model species with complex genomes (Neves *et al.*, 2013; Müller *et al.*, 2015; Lu *et al.*, 2016; Pavy *et al.*, 2016; Suren *et al.*, 2016; Syring *et al.*, 2016).

The accuracy of a genetic map is affected by several factors, including the distribution of targeted genomic regions, the quality of genotypic data, the number of mapped markers and the mapping population size. In this study, a limber pine transcriptome (Liu *et al.*, 2016b) was used for exome capture because a limber pine genome sequence is not yet available. A total of 14 706 unique protein-coding transcripts were selected as exome-enrichment targets. The designed hybridization probes consisted of 87 853 capture targets (14.23 Mb) and covered 93.6% of the total submitted transcripts. The high capture efficiency and sequencing depth in our study are similar to the levels recently reported (Lu *et al.*, 2016; Pavy *et al.*, 2016; Suren *et al.*, 2016). Capturing assays were successfully transferred from one species to a closely related species (Mascher *et al.*, 2013). As our probe was designed to target the highly conserved genes across *Pinus* species, the limber pine capturing assays have great potential to be used for other five-needle species for enriching their exomes.

To collect high-quality genotypic data, we first evaluated SNP data for read depth, the distribution of missing data, the co-segregation of multiple SNPs in the same gene and the co-segregation of different genes. Like other NGS-based platforms, exome-seq reads had missing data for some individuals. SNP loci were removed for mapping analysis if their genotypes were missed in more than 10% of the whole mapping population. Using highly stringent criteria, the bioinformatics pipeline called 26K and 43K SNP loci with the 1 : 1 expected segregation ratio in families LJ-112 and PHA-106, respectively. This confirmed that the SNP calling pipeline was highly reliable for detecting and filtering DNA variations in conifer exome-seq data (Neves *et al.*, 2014).

Exome-seq usually achieves much higher read depths than other NGS technologies, such as genotyping-by-sequencing (Bodi *et al.*, 2013). The SNP loci mapped to limber pine LGs had an average 27× and 36× coverage depth in two mapping families; these high levels ensure a reliable mapping analysis. To improve the accuracy of our high-density LG maps, a single representative SNP was further selected from multiple loci within each polymorphic gene. The accuracy of the limber pine map is satisfactory, as evidenced by several analyses, including map consistency between seed families and between current and previous maps from the same and related species.

The final composite map positioned 9612 genes with high density spanning 12 LGs. A high proportion of DNA variations in the protein-coding regions were nonsynonymous and are likely to cause functional changes in the encoded proteins (Liu *et al.*, 2016a). As SNPs identified by exome-seq are localized within expressed genes, exome-seq-derived maps should be beneficial for functional genomics. As the majority of mapped limber pine sequences showed good alignment with reference sequences in related conifer species (such as sugar pine), the high-density functional gene LG map provides a useful genomic resource of value primarily in limber pine and by extension in related pines. A further exploration of the relationship

between genetic and physical distance may help to anchor genomic scaffolds onto LGs and allow for the improvement of genome assemblies in closely related conifers with complex and huge genomes (Pavy *et al.*, 2017).

Conifer genome evolution

We used orthologous genes as bridges to compare genetic maps between limber pine and three other related conifer species. The gene content and order on limber pine LGs are generally very similar to those on homologous chromosomes of related conifers. The majority of bridging orthologs were mapped on the homologous LGs and they generally retained their ancestral genomic positions as POGs, whereas a minority were mapped on different LGs as nPOGs. GO term-enrichment analyses revealed relatively fewer constraints for POGs with putative roles in adaptation to environment, and relatively more conservation for POGs with roles in basic cell function and maintenance. In comparison with white spruce, however, limber pine LG3 and LG12 presented greater K_s values and lower correlation R^2 values, in contrast with other LGs, which may reflect local small-scale rearrangements among different LGs driven by tandem duplications, a frequent and ongoing event during genome evolution (Delseny, 2004).

The K_s values between orthologs or paralogs can be used as a molecular clock to estimate either divergence time of different species or gene duplication time in one species. Ancient whole-genome duplications (WGDs) are often detected by an examination of K_s distributions within genomes (Tiley *et al.*, 2018). The divergence age between different species, as estimated by orthologous gene pair-based K_s values, has provided evidence for the evolutionary events of WGD in the lineages of several angiosperm species (Ren *et al.*, 2018). At least one ancient WGD event was detected in Pinaceae, occurring between 210 and 342 Mya (Li *et al.*, 2015). Our analysis covering thousands of ortholog pairs between limber pine and related conifer species found mean K_s values that are comparable with results reported previously (Willyard *et al.*, 2007; Buschiazzo *et al.*, 2012; Li *et al.*, 2015). K_s distribution plots demonstrated a big peak at 0.075 for subgenera *Pinus–Strobus* and at 0.175 for subgenera *Pinus–Picea*, providing evidence that WGD did occur during conifer evolutionary history (Li *et al.*, 2015).

A comparison of evolutionary rates revealed nPOGs with K_s values significantly higher than POGs in conifers, indicating the more ancient origins of nPOGs. More recent gene duplications tend to have larger K_a/K_s values because at least part of the duplicated genes would have been subjected to rapid changes after gene duplication (Nembaware *et al.*, 2002; Domazet-Loso and Tautz, 2003). Correspondingly, we found that conifer POGs had less evolutionary constraints than nPOGs, although the vast majority of mapped conifer orthologs have been under purifying selection, with K_a/K_s values significantly below one.

Candidates for MGR and QTLs

Gene ontology (GO) analysis revealed that our genome-wide mapping positioned 639, 290 and 1013 gene sequences encoding putative NBS-LRR, RLK and defense-related proteins of other families, respectively. The *NBS-LRR* genes were localized unevenly across the 12 LGs. LG2 contained the highest proportion (41.6%) of *NBS-LRR* genes, whereas LG11 had only 2.2% of the total. The most predominant R proteins belong to the NBS-LRR and RLK superfamilies. Numerous well-characterized *NBS-LRR* genes confer MGR against biotrophic pathogens through incompatible plant–microbe interactions, as depicted by the 'gene-for-gene' or 'guard' model (Glazebrook, 2005).

Our high-density limber pine LG map allowed the fine genetic dissection of genomic regions of interest, including *Cr4* and orthologous loci (*Cr1* and *Cr2*) for the MGRs to *C. ribicola* in sugar pine and western white pine. Based on their positions on LG2 and LG1, *NBS-LRR* genes were proposed as *R* candidates against *C. ribicola* in these two five-needle pine species (Stevens *et al.*, 2016; Liu *et al.*, 2017a), as well as *Cronartium quercuum* Miyabe ex Shirai f.sp. *fusiforme* (*Cqf*) in loblolly pine (Neale *et al.*, 2014). An *NBS-LRR* gene (*M287456*) considered as a *Cr4* candidate showed only limited similarity to its closest homologs in the sugar pine genome and western white pine transcriptome. In addition to MGR localizations on different LGs, variant orthologous genomic regions provide additional evidence for the hypothesis that *R* loci to WPBR had evolved independently in these five-needle pine species before they were exposed to WPBR (Liu *et al.*, 2016b).

Although partial resistance studies with limber pine are just getting underway (Schoettle *et al.*, 2011), extensive investigations of quantitative resistance to *C. ribicola* have been conducted with some North American white pine species, which is most notable in western white pine, whitebark pine (*Pinus albicaulis*), and southwestern white pine (*Pinus strobiformis*) (Sniezko *et al.*, 2008, 2014). Association studies revealed nucleotide variations in several genes (encoding PR10, chitinase, thaumatin, anti-microbial protein and others) that made significant contributions to phenotypic traits of partial resistance against *C. ribicola* in sugar pine (Vázquez-Lobo *et al.*, 2017), western white pine (Liu *et al.*, 2005, 2011) and whitebark pine (Liu *et al.*, 2016a); however, so far no WPBR-resistant QTLs have been mapped on LGs in five-needle pines. In this study, we detected 38 genetic bins with the co-localization of at least four *NBS-LRR* sequences and eight of them were heterogeneous clusters that co-localized with genetic bins of other defense-related genes. These genes were involved in defense responses or induced systemic resistance to *C. ribicola* or other pathogens and pests in closely related

pine species (Liu *et al.*, 2017b). The Ch4A ortholog was tightly linked to *Cr4*, suggesting that this *R* locus may be associated with a QTL against WPBR (Liu *et al.*, 2005). Interaction of R proteins with pathogenic effectors activates a large set of downstream genes during the host defense response, leading to resistant phenotypes upon infection. The clustering of *NBS-LRR* genes with functionally related genes of other families appears to be consistent with the concept of the genomic context, in which it is assumed that clustered genes have resulted from positive selection to maintain their adjacency for a higher complexity (Chen *et al.*, 2010), which may facilitate coordinated expression within plant resistance signaling networks (Galperin and Koonin, 2000; López-Kleine *et al.*, 2013; Christopoulou *et al.*, 2015).

Genome-wide mapping of a large number of *NBS-LRR* genes offers insight into the evolution of this gene family in conifers as well as the potential involvement of *NBS-LRR* genes in resistance to *C. ribicol*a in different five-needle pine species. A close or co-localization of the *NBS-LRR* genes with known defense-responsive genes presents an additional line of evidence for the role of *NBS-LRR* genes in either MGR or QTLs for disease resistance. Our study provides strong groundwork for further functional studies to validate the roles of *NBS-LRR* genes in MGR and partial resistance. The identification of candidate genes at both *R* loci and QTLs will increase our understanding of the molecular mechanisms for disease resistance.

High-density map for application in limber pine breeding and conservation programs

In order to restore forest ecosystems damaged by WPBR and other environmental stressors across North America, conservation and tree improvement programs have been developed with great progress on five-needle pines, including limber pine (Sniezko *et al.*, 2014). Using conventional methods, however, it is difficult for tree breeders to capture and transfer the genetic variability of the targeted traits, such as quantitative resistance and other related adaptive traits. Although traditional breeding approaches currently are effective in screening five-needle pine resistance to WPBR, the genetic dissection of complex phenotypic traits is still an obstacle. In recent years, genomic resources have been developed in a few five-needle pines, including genome-wide marker discovery (Liu *et al.*, 2014; Syring *et al.*, 2016), high-density genetic maps (Jermstad *et al.*, 2011; Friedline *et al.*, 2015), transcriptome profiles (Lorenz *et al.*, 2012; Liu *et al.*, 2013, 2016b; Gonzalez-Ibeas *et al.*, 2016; Baker *et al.*, 2018) and whole-genome sequences (Stevens *et al.*, 2016). These genomics resources and tools open up a completely new avenue for the capture and use of genome-wide variability in breeding programs; however, the genomic information available so far is still too limited to generate a sufficient foundation of knowledge required for the genetic improvement of five-needle pines (Liu *et al.*, 2016a).

Genetic maps with a high density of markers throughout the genome have great potential to detect markers tightly linked to the traits of interest in sufficiently large mapping populations. Association studies are an effective genomic approach to identify DNA variants associated with phenotypes, which in turn may elucidate gene function and regulation as well as the allelic architectures of complex traits. Limber pine *Cr4* was first discovered in the southern Rocky Mountains of the USA (Schoettle *et al.*, 2014). Recently, limber pine MGR was found in Alberta, Canada, more than 1100 km north of the previously documented occurrence of *Cr4* (Sniezko *et al.*, 2016). For breeding and conservation programs, it is very important to know whether the MGRs found within the species among distant geographical sites are the same locus. Limber pine is closely related to southwestern white pine, and hybridization has taken place in at least some areas where these species overlap (Menon *et al.*, 2018). MGR has been previously documented in southwestern white pine in the US portion of the species range (Kinloch and Dupper, 2002; Sniezko *et al.*, 2008). The availability of the limber pine genetic map will enable a comparison of limber pine MGRs that have originated in different regions, and between limber pine and southwestern white pine MGRs.

Our previous work focused on limber pine transcriptome profiling, where only 25 genes were mapped over a genetic distance of about 50 cM on LG8 by genotyping 324 *in-silico* SNP loci using the Sequenom MassARRAY iPLEX platform (Liu *et al.*, 2016b). This study further refined the *Cr4* locus, which resulted in the localization of 50 genes within 5 cM of *Cr4* when SNPs with <10% of missing data were mapped (Table S3). These *Cr4*-linked polymorphic genes and *Cr4* functional candidates have high potential application in marker-assisted selection of resistance to WPBR. A TaqMan tool was developed using SNPs within a *Cr2*-linked *NBS-LRR* gene, a practical MAS tool applicable for the selection of MGR in five-needle pine breeding programs (Liu *et al.*, 2017b). The selection of favorable alleles or allelic combinations is a lengthy process in the breeding of outcrossing conifer species. To select genotypes related to a few genes with large phenotypic effects (such as MGR), the application of such MAS tools has the potential to shorten the length of each breeding cycle by allowing the precise prediction of phenotypes.

In order to select and predict complex traits with minor gene effects, genomic selection is an attractive approach for breeders. The genetic characterization of genes or QTLs contributing to various complex traits of interest is an essential prerequisite for genomics-based breeding: for example, quantitative resistance to WPBR, to mountain pine beetle, to cold, and to other related biotic and abiotic stressors are highly desirable traits in limber pines and

related five-needle pines. Our limber pine genetic map offers a valuable resource for future GWAS, QTL mapping and genomic selection of these complex traits. In particular, genetic bins consisting of *R* candidates and related genes will facilitate the identification of novel QTLs for WPBR and other pests or pathogens, help refine QTL positions and accelerate gene cloning in five-needle pines. Novel SNP markers, genes and QTLs will improve breeding programs for durable disease resistance and enhanced adaptability to environmental changes in five-needle pines.

EXPERIMENTAL PROCEDURES

Plant materials, resistance phenotyping and genomic DNA extraction

Two seed families LJ-112 and PHA-106 were used for exome-seq-based genetic mapping. These two populations were previously used in a study on the genetic mapping of the *Cr4* locus for resistance against *C. ribicola* (Liu *et al.*, 2016b). Seeds were collected from parental trees with heterozygous genotype *Cr4/cr4* from northern Colorado (40.79/−106.49, elevation 2527 m a.s.l.) and southern Wyoming (41.27/−105.43, elevation 2665 m), respectively (Schoettle *et al.*, 2014). Megagametophyte tissues were harvested for DNA extraction. Genomic DNA was extracted with a DNeasy Plant Mini kit (QIAGEN, https://www.qiagen.com). Phenotypes were determined for each seedling for resistance to *C. ribicola* as described previously (Schoettle *et al.*, 2014; Sniezko *et al.*, 2016).

Exome enrichment and sequencing

Based on a limber pine needle transcriptome reported previously (Liu *et al.*, 2016b), 14 706 unigenes in a total transcript length of 21.2 Mb were selected and submitted to Roche NimbleGen Inc. for the design of hybridization probes (Roche, https://design.nimblegen.com/nimbledesign). The majority of these targets were genes highly conserved in *Pinus* (BLASTn *E* value < $10e^{-100}$). In addition, about 1100 and 400 genes encoding putative NBS-LRR and RLK proteins (Liu *et al.*, 2016b) were included. After screening, the designed probes covered a total of 87 853 captured targets (14.23 Mb) with an estimated coverage of 93.6% of all submitted sequences. Of the 14 706 unigenes, only 182 genes were not targeted. KAPA Library Preparation Kits (Illumina Platforms, https://emea.illumina.com) were used to construct a genomic DNA library for each DNA sample. The Roche NimbleGen SeqCap EZ system was used for hybridization and target enrichment. Following quality assessment, the captured multiplex DNA samples were sequenced at 100-bp pair ends (PEs) using an Illumina HiSeq 2000 platform at McGill University and Génome Québec Innovation Centre (http://gqinnovationcenter.com). Genomic DNA seq reads were downloaded in FASTQ files per sample after quality filtering and demultiplexing for further bioinformatics analyses.

Exome-seq read alignment and SNP analysis

Single-nucleotide polymorphisms (SNPs) were mined in each sample using the reference transcriptome with 14 706 unigene sequences in read mapping. This reference file was formatted using PICARD-TOOLS 2.3.0, rebuilt using BOWTIE2 2.2.9 (Langmead and Salzberg, 2012). An FAI file was generated for read-mapping using SAMTOOLS 1.3.1 (Li *et al.*, 2009). The paired clean reads of each sample in the input data FASTA files were aligned with the generated reference using BOWTIE2 2.2.9 with the arguments 'local' and 'very-sensitive-local'. The generated SAM files were converted to BAM files, and then sorted and indexed using SAMTOOLS 1.3.1. The BAM files were then analyzed for sequence variant detection and genotype calling using FREEBAYES 1.0.2-16-gd466dde using default parameters with haploid mode by setting ploidy = 1, outputting VCF files (Garrison and Marth, 2012). Finally, SNP data were processed in the VCF files using VCFTOOLS 0.1.12b (Danecek *et al.*, 2011). Short indel, MNV and presence/absence variants (PAVs) were not included in this study. All extracted SNP data in tab-format files were merged and analyzed using in-house R scripts.

Genetic map construction

Haploid megagametophyte samples of the same seed family were maternally inherited from one mother tree and were used to calculate recombination rates between DNA markers during meiosis. SNPs with segregation ratios distorted from the expected 1 : 1 (χ^2 test, $P < 0.01$) were removed before inputting data for genetic map construction. Over 80% of all the SNPs had missing data levels of less than 10%, and these were used for the first run of the mapping analysis. SNPs with >10% of missing data were added in the second run of the mapping analysis, making little impact on SNP order and position when compared with the first run ($R^2 = 0.9883 \pm 0.0104$ for PHA-106, $R^2 = 0.9950 \pm 0.0067$ for LJ-112).

Genetic maps were constructed for each seed family using the software suite LEPMAP2 (Rastas *et al.*, 2015). The SEPARATE CHROMOSOMES module was used to assign SNP markers into linkage groups (LGs) at lodLimit = 10, and other remaining SNP markers were added by the JOINSINGLES module to existing LGs at lodLimit = 6. Finally, the ORDERMARKERS module was used to calculate the relative positions of SNP loci within each LG by maximizing the likelihood of the data given the order using input parameters alpha = 0.1, polishWindow = 100, filterWindow = 10, sexAveraged = 1, chromosome = X (for each chromosome X). Based on the maps constructed in the first run of LEPMAP2, marker assignment to LGs and marker positions on each LG were checked for potential genotyping errors. Genes and their SNP data were removed from the final map construction if multiple SNP markers in the same genes were assigned to different LGs in the same mapping population. LEPMAP2 was run for a second time using SNP data with one SNP marker per gene. For each polymorphic gene, representative SNP markers were selected with the lowest missing data in the mapping population, the lowest error estimate and the closest position to the median position if multiple SNP markers of the same genes were mapped on the same LG in the first run of map construction. Markers and genes with conflicting LG assignments between families were removed from seed family LJ-112 because it had a smaller mapping population size, fewer polymorphic genes and lower SNP coverage depth, compared with family PHA-106.

The genetic maps from the two seed families were integrated using LPMERGE with 10 trial runs using a maximum interval size between bins ranging from 1 to 10 (Endelman and Plomion, 2014). Input maps were weighted based on the sizes of individual mapping populations and percentages of missing data. For each LG, the best consensus map was chosen according to the software developers' recommendation by minimizing the average root mean-squared error (RMSE) and achieving a total map length comparable with the mean of the LG maps. Each merged LG was compared with its individual input maps. If terminal markers distorted the merged map from the individual contributing maps, they were deleted from input data for the second run of LPMERGE

for the removal of singleton markers at the end of a merged LG [International Cassava Genetic Map Consortium (ICGMC), 2015].

Synteny analysis with *Pinus* and *Picea* species

A total of 3856 and 5194 genes were mapped on *Pinus* consensus maps based on data from *P. taeda* and *Pinus elliottii* (Westbrook *et al.*, 2015), or from *P. pinaster* and *P. taeda* (De Miguel *et al.*, 2015; INRA_Pinus_composite_2016, released 24 March, 2016 at http://w3.pierroton.inra.fr/cgi-bin/cmap_pinus), and 8793 genes were mapped on a white spruce (*Picea glauca*) composite map (Pavy *et al.*, 2017). These data were downloaded to explore intergenome syntenic relationships. Reciprocal best-hit Blastn analysis was performed to identify orthologous gene pairs between mapped limber pine protein-coding sequences and the mapped genes of the above three species at an *E*-value threshold of $1e^{-100}$. The paired orthologous genes were used as bridging genes for syntenic analysis and genome evolutionary analysis. The resulting synteny in paired intergenome comparison was visualized using CIRCOS (Krzywinski *et al.*, 2009). Mapped genes with matching chromosomes were further subjected to Pearson correlation analysis to compare gene positions between two species.

Paired orthologs were considered as POGs for two species if their relative position on homologous LGs were preserved, whereas orthologs with localizations across heterologous LGs were considered to be the outside synteny group as nPOGs. Genome evolution was analyzed by the measurement of synonymous substitution levels (K_s) and nonsynonymous nucleotide substitution rates (K_a) between paired orthologous sequences using PARAAT (Zhang *et al.*, 2012) and the KAKS calculator, with model averaging (Zhang *et al.*, 2006). A K_s cutoff of 2 was set for distribution analysis because of K_s saturation and stochastic effects (Vanneste *et al.*, 2013).

Annotation of limber pine mapped genes

Gene ontology (GO) analysis was performed using BLAST2GO (Conesa *et al.*, 2005). The databases used in Blast analysis included the NCBI nr database, Arabidopsis genome, sugar pine putative proteome (Stevens *et al.*, 2016), *NBS-LRR* genes, *RLK* genes and WPBR-responsive genes in western white pine (Liu and Ekramoddoullah, 2007; Liu *et al.*, 2013), MeJA-responsive genes in whitebark pine (Liu *et al.*, 2017b), and pine weevil-responsive genes in Scots pine (*Pinus sylvestris*) (Kovalchuk *et al.*, 2015). GO term enrichment analysis was performed for gene groups and the significance of each comparison was evaluated by Fisher's exact test using a false discovery rate (FDR) threshold of 0.01.

To reveal relationships between the linkage location of *NBS-LRR* genes and their phylogenetic clustering, full-type *NBS-LRR* genes were determined using criteria based on well-defined NBS motifs (Seo *et al.*, 2016). Amino acid sequences of NBS domains with at least 150 amino acid residues were aligned and used for phylogenetic analysis. A phylogenetic tree was constructed using a neighbor-joining method with 1000 bootstrap retests. Limber pine paralogs localized on either the same LGs or across different LGs were used to assess local gene duplication and whole genome duplication, respectively. For paired paralogs from gene duplication inside phylogenetic clades, K_s, K_a and K_a/K_s ratios were calculated using the KAKS calculator, as outlined above. For *NBS-LRR* sequences mapped at the same position, nucleotide sequences were aligned for the construction of phylogenetic trees to detect branches of the NBS-LRR family trees under evolutionary pressure. Ancestral sequences were reconstructed to calculate K_a/K_s ratios for phylogenetic tree branches using a covarion-based approach (Siltberg and Liberles, 2002).

ACKNOWLEDGEMENTS

We would like to thank Gary Zhang and Emily Ayala at the Canadian Forest Service (CFS) for bioinformatic analyses, colleagues at the Dorena Genetic Resource Center (DGRC) for sample collection and phenotype assessment, and Dr Andrew J. Eckert at Virginia Commonwealth University and two anonymous reviewers for their constructive comments that greatly contributed to improving the article. This research was supported in part by the CFS-GRDI and the USDA-Forest Service Special Technology Development Program (award STDP-R2-2014-01).

CONFLICT OF INTEREST

The authors declare no known conflicts of interest.

AUTHOR CONTRIBUTION

J-JL, AWS and RAS designed the study and performed the interpretation. AZ and HW performed the experiments. FY and BR performed bioinformatics analyses of exome-seq data. J-JL drafted the article. All authors reviewed and approved the final version for publication.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Cumulative distribution of the coverage depth of sequenced targets in two mapping seed families. Each line represents one megagametophyte sample. Coverage depth showed a similar pattern across all 190 samples.

Figure S2. Distribution of gap distances (cM) between two adjacent unique positions on the limber pine genetic maps.

Figure S3. Concordance of gene order and positions between two limber pine seed families. The *x*-axis shows the genetic distance in centimorgan (cM) of the linkage group (LG) map from seed family PHA-106, and the *y*-axis is the genetic distance (cM) of the linkage group (LG) map from seed family LJ-112.

Figure S4. Enrichment analysis of gene ontology (GO) terms for positional orthologous genes (POGs). All mapped genes were used as reference and significance for the GO term difference was evaluated by Fisher's exact test with a false-discovery rate (FDR) threshold at $P < 0.01$.

Figure S5. Comparisons of mean K_s values between *NBS-LRR* paralogous groups. *NBS-LRR* paralogs within each phylogenetic clade as shown in Figure 4 were paired for the K_s calculation. Paralogs mapped onto the homologous LGs are referred to as aligned, whereas paralogs mapped onto different LGs were referred to as misaligned. Error bars represent sample variance. Differences were analyzed Student's *t*-test and one-way ANOVA with Tukey's HSD test: $*P < 0.05$; $**P < 0.01$; $***P < 0.001$.

Figure S6. The K_a/K_s values for each node in the phylogenetic tree of the limber pine orthologous locus for resistance to *Cronartium ribicola*. The analysis was performed using on-line server of K_a/K_s calculation tool (http://services.cbu.uib.no/tools/kaks). Seven limber pine *NBS-LRR* sequences and the *Cr1* candidate (PILAAhq_017371-R) were included. K_a/K_s values were larger than one for limber pine gene M362534 and the *Cr1* candidate, indicating that they were subject to positive selection.

Table S1. Summary of exome-seq and SNP analyses for genetic map construction.

Table S2. Statistics of linkage groups (LGs) constructed using LepMAP2.

Table S3. Annotation of mapped limber pine genes by Blastp search against Arabidopsis genome.
Table S4. Pearson correlation analysis of orders and positions of bridging genes along homologous chromosomes (linkage groups) among conifer species.
Table S5. Limber pine mapped gene sequences encoding putative NBS-LRR proteins as annotated by Blastp search against GenBank nr database.
Table S6. Limber pine mapped gene sequences encoding putative RLK proteins as annotated by Blastp search against GenBank nr database.
Table S7. Genetic bins of gene groups based on sequence similarity and expression patterns.
Table S8. Limber pine mapped gene sequences with potential roles in induced resistance as predicted by Blastn search against whitebark pine MJ-responsive transcripts.
Table S9. Limber pine mapped gene sequences with potential response to insect attack as predicted by Blastn search against Scots pine genes.
Table S10. Limber pine mapped gene sequences with potential response to WPBR infection as predicted by Blastn search against western white pine genes.

REFERENCES

Baker, E.A.G., Wegrzyn, J.L., Sezen, U.U. *et al.* (2018) Comparative transcriptomics among four white pine species. *G3*, **8**, 1461–1474. https://doi.org/10.1534/g3.118.200257.

Bartholomé, J., Bink, M.C., van Heerwaarden, J., Chancerel, E., Boury, C., Lesur, I., Isik, F., Bouffier, L. and Plomion, C. (2016) Linkage and association mapping for two major traits used in the maritime pine breeding program: height growth and stem straightness. *PLoS ONE*, **11**(11), e0165323.

Bodi, K., Perera, A.G., Adams, P.S. *et al.* (2013) Comparison of commercially available target enrichment methods for next-generation sequencing. *J. Biomol. Tech.* **24**, 73–86.

Buschiazzo, E., Ritland, C., Bohlmann, J. and Ritland, K. (2012) Slow but not low: genomic comparisons reveal slower evolutionary rate and higher dN/dS in conifers compared to angiosperms. *BMC Evol. Biol.* **12**, 8.

Chen, Q., Han, Z., Jiang, H., Tian, D. and Yang, S. (2010) Strong positive selection drives rapid diversification of R-genes in Arabidopsis relatives. *J. Mol. Evol.* **70**, 137–148.

Christopoulou, M., McHale, L.K., Kozik, A., Wo, S.R.-C., Wroblewski, T. and Michelmore, R.W. (2015) Dissection of two complex clusters of resistance genes in lettuce (*Lactuca sativa*). *Mol. Plant Microbe Interact.* **28**, 751–765.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M. and Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**, 3674–3676.

COSEWIC. (2014) COSEWIC assessment and status report on the limber pine *Pinus flexilis* in Canada. Committee on the Status of Endangered Wildlife in Canada. Ottawa. ix + 49 pp. (www.registrelepsararegistry.gc.ca/default_e.cfm).

Danecek, P., Auton, A., Abecasis, G. *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

De Miguel, M., Bartholomé, J., Ehrenmann, F. *et al.* (2015) Evidence of intense chromosomal shuffling during conifer evolution. *Genome Biol. Evol.* **7**, 2799–2809.

Delseny, M. (2004) Re-evaluating the relevance of ancestral shared synteny as a tool for crop improvement. *Curr. Opin. Plant Biol.* **7**, 126–131.

Dewey, C.N. (2011) Positional orthology: putting genomic evolutionary relationships into context. *Brief. Bioinform.* **12**, 401–412.

Domazet-Loso, T. and Tautz, D. (2003) An evolutionary analysis of orphan genes in Drosophila. *Genome Res.* **13**, 2213–2219.

Endelman, J.B. and Plomion, C. (2014) LPmerge: an R package for merging genetic maps by linear programming. Bioinformatics, https://doi.org/10.1093/bioinformatics/btu091.

Friedline, C.J., Lind, B.M., Hobson, E.M., Harwood, D.E., Mix, A.D., Maloney, P.E. and Eckert, A.J. (2015) The genetic architecture of local adaptation I: the genomic landscape of foxtail pine (*Pinus balfouriana* Grev. & Balf.) as revealed from a high-density linkage map. *Tree Genet. Genomes*, **11**, 49.

Galperin, M.Y. and Koonin, E.V. (2000) Who's your neighbor? New computational approaches for functional genomics. *Nat. Biotechnol.* **18**, 609–613.

Ganal, M.W., Altmann, T. and Röder, M.S. (2009) SNP identification in crop plants. *Curr. Opin. Plant Biol.* **2**, 211–217.

Garrison, E. and Marth, G. (2012) Haplotype-based variant detection from short-read sequencing. *ArXiv e-Prints*, **1207**, 3907.

Glazebrook, J. (2005) Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu. Rev. Phytopathol.* **43**, 205–227.

Gonzalez-Ibeas, D., Martinez-Garcia, P.J., Famula, R.A., Delfino-Mix, A., Stevens, K.A., Loopstra, C.A., Langley, C.H., Neale, D.B. and Wegrzyn, J.L. (2016) Assessing the gene content of the megagenome: sugar Pine (*Pinus lambertiana*). *G3*, **6**, 3787–3802.

International Cassava Genetic Map Consortium (ICGMC). (2015) High-resolution linkage map and chromosome-scale genome assembly for cassava (*Manihot esculenta* Crantz) from 10 Populations. *G3*, **5**, 133–144.

Jermstad, K.D., Eckert, A.J., Wegrzyn, J.L., Delfino-Mix, A., Davis, D.A., Burton, D.C. and Neale, D.B. (2011) Comparative mapping in Pinus: sugar pine (*Pinus lambertiana* Dougl.) and loblolly pine (*Pinus taeda* L.). *Tree Genet. Genomes*, **7**, 457–468.

Kinloch, B.B. and Dupper, G.E. (2002) Genetic specificity in the white pine-blister rust pathosystem. *Phytopathology*, **92**, 278–280.

Kovalchuk, A., Raffaello, T., Jaber, E., Keriö, S., Ghimire, R., Lorenz, W., Dean, J.F.D., Holopainen, J.K. and Asiegbu, F.O. (2015) Activation of defence pathways in Scots pine bark after feeding by pine weevil (*Hylobius abietis*). *BMC Genom.* **16**, 352.

Krist, F.J., Ellenwood, J.R., Woods, M.E., McMahan, A.J., Cowardin, J.P., Ryerson, D.E., Sapio, F.J., Zweifler, M.O. and Romero, S.A. (2014) 2013-2027 National insect and disease forest risk assessment. U.S. Department of Agriculture, Forest Service, Forest Health Technology Enterprise Team.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J. and Marra, M.A. (2009) Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645.

Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) The sequence alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Li, Z., Baniaga, A., Sessa, E., Scascitelli, M., Graham, S., Rieseberg, L. and Barker, M. (2015) Early genome duplications in conifers and other seed plants. *Science Advances*, **1**, e1501084.

Liu, J.-J. and Ekramoddoullah, A. (2007) The CC-NBS-LRR subfamily in western white pine (*Pinus monticola* D. Don.): targeted identification, gene expression and genetic linkage with disease resistance against *Cronartium ribicola*. *Phytopathology*, **97**, 728–736.

Liu, J.-J., Ekramoddoullah, A. and Zamani, A. (2005) A class IV chitinase is up-regulated upon fungal infection and abiotic stresses and associated with slow-canker-growth resistance to *Cronartium ribicola* in western white pine (*Pinus monticola*, Dougl. Ex D. Don). *Phytopathology*, **95**, 284–291.

Liu, J.-J., Sniezko, R.A. and Ekramoddoullah, A.K. (2011) Association of a novel *Pinus monticola* chitinase gene (*PmCh4B*) with quantitative resistance to *Cronartium ribicola*. *Phytopathology*, **101**, 904–911.

Liu, J.-J., Sturrock, R.N. and Benton, R. (2013) Transcriptome analysis of *Pinus monticola* primary needles by RNA-seq provides novel insight into host resistance to *Cronartium ribicola*. *BMC Genom.* **14**, 884.

Liu, J.-J., Sniezko, R.A., Sturrock, R.N. and Chen, H. (2014) Western white pine SNP discovery and high-throughput genotyping for breeding and conservation applications. *BMC Plant Biol.* **14**, 1586.

Liu, J.-J., Sniezko, R., Murray, M., Wang, N., Chen, H., Zamany, A., Sturrock, R.N., Savin, D. and Kegley, A. (2016a) Genetic diversity and population structure of whitebark pine (*Pinus albicaulis* Engelm.) in western North America. *PLoS ONE*, **11**, e0167986.

Liu, J.-J., Schoettle, A.W., Sniezko, R.A. *et al.* (2016b) Genetic mapping of *Pinus flexilis* major gene (*Cr4*) for resistance to white pine blister rust using transcriptome-based SNP genotyping. *BMC Genom.* **17**, 753.

Liu, J.-J., Sniezko, R.A., Zamany, A., Williams, H., Wang, N., Kegley, A., Savin, D.P., Chen, H. and Sturrock, R.N. (2017a) Saturated genic SNP mapping identified functional candidates and selection tools for the *Pinus monticola Cr2* locus controlling resistance to white pine blister rust. *Plant Biotechnol. J.* **15**, 1149–1162.

Liu, J.-J., Williams, H., Li, X.R., Schoettle, A.W., Sniezko, R.A., Murray, M., Zamany, A., Roke, G. and Chen, H. (2017b) Profiling methyl jasmonate-responsive transcriptome for understanding induced systemic resistance in whitebark pine (*Pinus albicaulis*). *Plant Mol. Biol.* **95**, 359–374.

López-Kleine, L., Pinzón, A., Chaves, D., Restrepo, S. and Riaño-Pachón, D.M. (2013) Chromosome 10 in the tomato plant carries clusters of genes responsible for field resistance/defence to *Phytophthora infestans*. *Genomics*, **101**, 249–255.

Lorenz, W.W., Ayyampalayam, S., Bordeaux, J.M., Howe, G.T., Jermstad, K.D., Neale, D.B., Rogers, D.L. and Dean, J.F. (2012) Conifer DBMagic: a database housing multiple de novo transcriptome assemblies for 12 diverse conifer species. *Tree Genet. Genomes*, **8**, 1477–1485.

Lu, M., Krutovsky, K.V., Nelson, C.D., Koralewski, T.E., Byram, T.D. and Loopstra, C.A. (2016) Exome genotyping, linkage disequilibrium and population structure in loblolly pine (*Pinus taeda* L.). *BMC Genom.* **17**, 730.

Mammadov, J., Aggarwal, R., Buyyarapu, R. and Kumpatla, S. (2012) SNP markers and their impact on plant breeding. *Int. J. Plant Genomics*, **2012**, 728398.

Mascher, M., Richmond, T.A., Gerhardt, D.J. *et al.* (2013) Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. *Plant J.* **76**, 494–505.

Menon, M., Bagley, J.C., Friedline, C.J. *et al.* (2018) The role of hybridization during ecological divergence of southwestern white pine (*Pinus strobiformis*) and limber pine (*P. flexilis*). *Mol. Ecol.* **27**, 1245–1260.

Müller, T., Freund, F., Wildhagen, H. and Schmid, K.J. (2015) Targeted re-sequencing of five Douglas-fir provenances reveals population structure and putative target genes of positive selection. *Tree Genet. Genomes*, **11**, 1–17.

Neale, D.B., Wegrzyn, J.L., Stevens, K.A. *et al.* (2014) Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biol.* **15**, R59.

Nembaware, V., Crum, K., Kelso, J. and Seoighe, C. (2002) Impact of the presence of paralogs on sequence divergence in a set of mouse-human orthologs. *Genome Res.* **12**, 1370–1376.

Neves, L.G., Davis, J.M., Barbazuk, W.B. and Kirst, M. (2013) Whole-exome targeted sequencing of the uncharacterized pine genome. *Plant J.* **75**, 146–156.

Neves, L.G., Davis, J.M., Barbazuk, W.B. and Kirst, M. (2014) A high-density gene map of loblolly pine (*Pinus taeda* L.) based on exome sequence capture genotyping. *G3*, **4**, 29–37.

Pavy, N., Gagnon, F., Deschênes, A., Boyle, B., Beaulieu, J. and Bousquet, J. (2016) Development of highly reliable in silico SNP resource and genotyping assay from exome capture and sequencing: an example from black spruce (*Picea mariana*). *Mol. Ecol. Resour.* **16**, 588–598.

Pavy, N., Lamothe, M., Pelgas, B., Gagnon, F., Birol, I., Bohlmann, J., Mackay, J., Isabel, N. and Bousquet, J. (2017) A high-resolution reference genetic map positioning 8.8 K genes for the conifer white spruce: structural genomics implications and correspondence with physical distance. *Plant J.* **90**, 189–203.

Plomion, C., Bartholomé, J., Lesur, I. *et al.* (2016) High-density SNP assay development for genetic analysis in maritime pine (*Pinus pinaster*). *Mol. Ecol. Resour.* **16**, 574–587.

Rasheed, A., Hao, Y., Xia, X., Khan, A., Xu, Y., Varshney, R.K. and He, Z. (2017) Crop breeding chips and genotyping platforms: progress, challenges and perspectives. *Mol. Plant*, **10**, 1047–1064.

Rastas, P., Calboli, F.C., Guo, B., Shikano, T. and Merilä, J. (2015) Construction of ultra-dense linkage maps with Lep-MAP2: stickleback F2 recombinant crosses as an example. *Genome Biol. Evol.* **8**, 78–93.

Ren, R., Wang, H., Guo, C., Zhang, N., Zeng, L., Chen, Y., Ma, H. and Qi, J. (2018) Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. *Mol. Plant*, **11**, 414–428.

Schoettle, A. and Stritch, L. (2013) *Pinus flexilis*. The IUCN Red List of Threatened Species, e.T42363A2975338.

Schoettle, A.W., Sniezko, R.A., Kegley, A. and Burns, K.S. (2011) Preliminary overview of the first extensive rust resistance screening tests of *Pinus flexilis* and *Pinus aristata*. In: *The Future of High-Elevation, Five-Needle White Pines in Western North America: Proceedings of the High Five Symposium, 28–30 June 2010, Missoula, MT*. Proceedings RMRS-P-63 (Keane, R.E., Tomback, D.F., Murray, M.P. and Smith, C.M., eds). Fort Collins, CO: USDA Forest Service, pp 265–269.

Schoettle, A.W., Sniezko, R.A., Kegley, A. and Burns, K.S. (2014) White pine blister rust resistance in limber pine: evidence for a major gene. *Phytopathology*, **104**, 163–173.

Seo, E., Kim, S., Yeom, S.-I. and Choi, D. (2016) Genome-wide comparative analyses reveal the dynamic evolution of nucleotide-binding leucine-rich repeat gene family among Solanaceae plants. *Front. Plant Sci.* **7**, 1205.

Siltberg, J. and Liberles, D.A. (2002) A simple covarion-based approach to analyse nucleotide substitution rates. *J. Evol. Biol.* **15**, 588–594.

Sniezko, R.A., Kegley, A.J. and Danchok, R. (2008) White pine blister rust resistance in North American, Asian and European species - results from artificial inoculation trials in Oregon. *Ann. For. Res.* **51**, 53–66.

Sniezko, R., Smith, J., Liu, J.-J. and Hamelin, R. (2014) Genetic resistance to fusiform rust in southern pines and white pine blister rust in white pines—A contrasting tale of two rust pathosystems—Current status and future prospects. *Forests*, **5**, 2050–2083.

Sniezko, R.A., Danchok, R., Savin, D.P., Liu, J.-J. and Kegley, A. (2016) Genetic resistance to white pine blister rust in limber pine (*Pinus flexilis*): major gene resistance in a northern population. *Can. J. For. Res.* **46**, 1173–1178.

Stevens, K.A., Wegrzyn, J.L., Zimin, A. *et al.* (2016) Sequence of the sugar pine megagenome. *Genetics*, **204**, 1613–1626.

Suren, H., Hodgins, K.A., Yeaman, S., Nurkowski, K.A., Smets, P., Rieseberg, L.H., Aitken, S.N. and Holliday, J.A. (2016) Exome capture from the spruce and pine giga-genomes. *Mol. Ecol. Resour.* **16**, 1136–1146.

Syring, J.V., Tennessen, J.A., Jennings, T.N., Wegrzyn, J., Scelfo-Dalbey, C. and Cronn, R. (2016) Targeted capture sequencing in whitebark pine reveals range-wide demographic and adaptive patterns despite challenges of a large, repetitive genome. *Front. Plant Sci.* **7**, 484.

Tiley, G.P., Barker, M.S. and Burleigh, J.G. (2018) Assessing the performance of Ks plots for detecting ancient whole genome duplications. *Genome Biol. Evol.* **10**, 2882–2898.

Vanneste, K., Van de Peer, Y. and Maere, S. (2013) Inference of genome duplications from age distributions revisited. *Mol. Biol. Evol.* **30**, 177–190.

Vázquez-Lobo, A., De La Torre, A.R., Martínez-García, P.J. *et al.* (2017) Finding loci associated to partial resistance to white pine blister rust in sugar pine (*Pinus lambertiana* Dougl.). *Tree Genet. Genomes*, **13**, 108.

Westbrook, J.W., Chhatre, V.E., Wu, L.-S. *et al.* (2015) A consensus genetic map for *Pinus taeda* and *Pinus elliottii* and extent of linkage disequilibrium in two genotype-phenotype discovery populations of *Pinus taeda*. *G3*, **25**, 1685–1694.

Willyard, A., Syring, J., Gernandt, D., Liston, A. and Cronn, R. (2007) Molecular evolutionary rates indicate a recent and rapid diversification for modern pine lineages. *Mol. Biol. Evol.* **23**, 1–12.

Zhang, Z., Li, J., Zhao, X.Q., Wang, J., Wong, G.K. and Yu, J. (2006) KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics*, **4**, 259–263.

Zhang, Z., Xiao, J., Wu, J., Zhang, H., Liu, G., Wang, X. and Dai, L. (2012) ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments. *Biochem. Biophys. Res. Commun.* **419**, 779–781.