

The CC-NBS-LRR Subfamily in *Pinus monticola*: Targeted Identification, Gene Expression, and Genetic Linkage with Resistance to *Cronartium ribicola*

Jun-Jun Liu and Abul K. M. Ekramoddoullah

Pacific Forestry Centre, Canadian Forest Service, Natural Resources Canada, 506 West Burnside Road, Victoria, BC, V8Z 1M5.
Accepted for publication 26 December 2006.

ABSTRACT

Liu, J.-J., and Ekramoddoullah, A. K. M. 2007. The CC-NBS-LRR subfamily in *Pinus monticola*: Targeted identification, gene expression, and genetic linkage with resistance to *Cronartium ribicola*. *Phytopathology* 97:728-736.

To investigate disease resistance gene analogs (RGAs) encoding coiled-coil–nucleotide-binding site–leucine-rich repeats (CC-NBS-LRR) proteins in western white pine, degenerate primers targeting the conserved motifs in the NBS domain were designed to amplify RGAs from genomic DNA and cDNA. Sixty-one distinct RGAs were identified with identities to well-known R proteins of the CC-NBS-LRR subfamily. These RGAs exhibited variation of putative amino acid sequences from 13% to 98%, representing a complex CC-NBS-LRR subfamily. A phylo-

genetic tree constructed from the amino acid sequence alignment revealed that these 61 RGAs were grouped with other CC-NBS-LRR members from angiosperms, and could be further divided into six classes with an identity threshold of 68%. To map RGAs, RGA polymorphisms and a modified amplified fragment length polymorphism (AFLP) method with incorporated sequences from the NBS domain were used to reveal NBS or NBS-AFLP markers. RGA polymorphism study revealed that three of the identified RGAs were not linked to the *Cr2* gene imparting resistance to white pine blister rust. However, the AFLP strategy, using bulk segregant analysis (BSA) and haploid segregation analysis, identified 11 NBS-AFLP markers localized in the *Cr2* linkage, the closest two to the gene being 0.41 cM and 1.22 cM away at either side. Eight of these markers showed significant amino acid sequence homologies with RGAs.

Since its introduction to North America about 100 years ago, white pine blister rust (WPBR), caused by *Cronartium ribicola* J.C. Fisher ex Rabh., has devastated populations of western white pine (*Pinus monticola* D. Don.) and other five-needle pines in many regions of western North America. Breeding for resistance to WPBR and maintenance of a variety of genetic resources have been high priorities in tree improvement programs for five-needle pines in British Columbia (BC), Canada (17). A single dominant gene for resistance (R) to WPBR, *Cr2*, was discovered in western white pine (12). This R gene provides complete resistance to most races of *C. ribicola* by a typical hypersensitive reaction (12). In addition, partial resistances to WPBR, such as slow canker growth resistance (SCG), and different levels of disease tolerance have also been identified and utilized for western white pine breeding (17). Although R loci have been genetically described and mapped in a few conifer species (15), to the best of our knowledge, no R gene has yet been cloned and characterized at the molecular level in any gymnosperm.

Based on the gene-for-gene model proposed by Flor (5), it is believed that race-specific resistance is activated once a host R gene product recognizes a pathogen avirulence gene product (4). The interaction of these two complementary genes is the initial and crucial step that triggers a battery of plant defense mechanisms. Since the first R gene was characterized in plants (19), more than 40 R genes have been cloned in angiosperms (18). The majority of identified R genes encode proteins that belong to a superfamily with a putative amino-terminal signaling domain, a nu-

cleotide-binding site (NBS), and a series of leucine-rich repeats (LRR). This family of NBS-LRR proteins is further classified into two major subfamilies based on structural characteristics of their putative amino-terminal signaling domain (1). One subfamily, TIR-NBS-LRR, contains an amino-terminal signaling domain that shares homology with *Drosophila* Toll and mammalian interleukin-1 receptor (TIR), and includes R proteins encoded by flax *L*, *M*, and tobacco *N*. The other subfamily, CC-NBS-LRR, does not have a TIR domain, but instead has a putative coiled-coil structure (CC) in the amino-terminal signaling domain, as seen in R proteins encoded by *Arabidopsis* *ADR1*, *RPS2*, and *RPM1* (4).

Both TIR-NBS-LRR and CC-NBS-LRR subfamilies have been found in dicots. The *Arabidopsis* genome contains 149 potentially expressed NBS-LRR genes; of these, 55 are the CC-type and at least 11 are functional R genes (21). However, the NBS-LRR genes identified in monocots are exclusively of the CC-type. The rice genome contains 535 NBS-coding sequences, including 480 CC-NBS-LRR genes (25).

Conifer species contain large genomes and have long generation times. Western white pine DNA content reaches 26,656 Mbp per haploid genome (27.20 pg/1C) (2), with approximately 10 Mb per centimorgan (cM) (15). It is also recalcitrant for genetic transformation. These factors make positional cloning or transposon-tagging cloning of a conifer R gene difficult at the present time. A candidate-gene-based approach may provide a more feasible and cost-effective strategy for R gene identification in conifer species. The NBS domains of R proteins contain a few highly conserved motifs that are useful to design degenerate primers to amplify resistance gene analogs (RGAs) (24). This polymerase chain reaction (PCR)-based RGA cloning approach has been successfully applied in many angiosperm species. A number of RGA loci have been mapped in the vicinity of specific R loci, and PCR- or restriction fragment length polymorphism (RFLP)-based RGA markers provide easier genetic analysis, gene characterization, and use of R genes in breeding (3). Recently, RGA sequences

Corresponding author: A. K. M. Ekramoddoullah;
E-mail address: aekramoddoull@nrcan.gc.ca

doi:10.1094/PHYTO-97-6-0728

Authors A. K. M. Ekramoddoullah and J.-J. Liu (© 2007, Her Majesty the Queen in right of Canada, Natural Resources Canada, Canadian Forest Service).

have been reported in gymnosperms, including *Pinus taeda* (L.) (20), *Cryptomeria japonica* (22), *Picea abies* (L.) (10), *Pinus lambertiana*, and *P. monticola* (9). Our previous investigation showed that there is a highly diverse group of TIR-NBS-LRR genes in western white pine (14).

In this study, a targeted PCR approach was used to clone RGA sequences of the CC-NBS-LRR subfamily in western white pine. To map RGAs, the genetic linkage of RGAs with the R gene (*Cr2*) for WPBR was explored.

MATERIALS AND METHODS

Plant materials. *Pinus monticola* (D. Don.) seed lots #3277, #3278, and #3566 were used as they are known to possess the *Cr2* gene. Seed lots #3277 and #3278 were obtained from Dorena Genetic Resource Center of the United States Department of Agriculture (USDA) Forest Service (Cottage Grove, OR) (8). They were open-pollinated seeds from two mother resistant trees with USDA registration #119-15045-845 and #119-15045-845 X 15045-841, respectively (12). Seed lot #3566 was open-pollinated seeds from one mother tree growing at Port Alberni, BC, Canada, originally from a bulked seed orchard collection from Dorena Genetic Resource Center. This mother tree was previously identified as heterozygous resistant (*Cr2/cr2*) (15).

PCR cloning of RGA sequences from western white pine genomic DNA. Genomic DNA was extracted from needles of resistant seedlings from seed lot #3277 using a Qiagen DNeasy kit (Qiagen Inc., Mississauga, ON). A targeted PCR approach was used to clone the CC-NBS-LRR sequences from genomic DNA. Design of five degenerate primers for the targeted PCR cloning was based on conserved motifs in the NBS domain of the CC-NBS-LRR family (Table 1). Two forward primers, GKT-d and DVW-d, targeted the conserved P-loop/kin-1a motif, MGGLGKTT, and kin-2 motif, (V/L)(I/V)LDDVW, respectively, whereas three reverse primers, CFL-d, HD1-d, and HD2-d, targeted the conserved RNBS-D-nonTIR motif, CFLDLG(A/S)F, and HD-motif, QHDVMR, or IHDELRLN, respectively.

PCR amplification of RGA sequences was performed in a volume of 25 µl with 30 ng of genomic DNA with 2× Taq Master Mix (Qiagen). The PCR conditions included initial denaturation for 3 min at 94°C, followed by 35 cycles of denaturing for 30 s at 94°C, primer annealing for 1 min at 45°C, and primer extension for 2 min at 72°C, with final incubation for 10 min at 72°C. Following separation by agarose gel electrophoresis and purification with a gel excision kit (Qiagen), PCR fragments were cloned into the pGEM-T easy vector (Promega, Madison, WI).

Detection of expressed RGAs using reverse-transcriptase (RT)-PCR. Total RNA was isolated from needles, stems, and roots of 2-year-old seedlings of seed lot #3278 using a procedure described previously by Liu et al. (16). Total RNA was treated with RQ 1 RNase-Free DNase (Promega) and purified again using a Plant RNeasy Extraction Kit (Qiagen). cDNA was synthesized with the pooled total RNA of needles, stem, and roots using a SMART cDNA library construction kit (Clontech Laboratories, Palo Alto, CA). To detect expressed RGAs using RT-PCR, the primers and PCR conditions were the same as those for the genomic DNA PCR process except that 2 µl of cDNA was used as the PCR template.

DNA sequence analysis. Plasmid DNA with PCR inserts from genomic DNA or cDNA was purified using a QIAprep Spin Miniprep kit (Qiagen). DNA sequences of RGAs were determined on both strands on an ABI310 DNA sequencer (Applied Biosystems, Foster City, CA) using a Thermo-cycle sequence kit (Amersham, Baie d'Urfe, Quebec) with T7, SP6 primer, and other internal primers as needed, according to manufacturer instructions. RGA sequence data were compiled and analyzed using BLASTX and BLASTP (National Center for Biotechnology Information [NCBI], Bethesda, MD), DNASTAR software (DNASTAR, Inc., Madison,

WI), and ExPASy Proteomics tools (Swiss Institute of Bioinformatics, Geneva). Nucleotide or protein sequences were aligned online with the Clustal W network service at the European Bioinformatics Institute (EBI, Cambridge, UK). Based on the alignment of nucleotide or deduced amino acid sequences, phylogenetic trees were constructed using a Neighbor-Joining (NJ) method in the MEGA3 software package (13). The reliability of the tree was established by conducting 1,000 Neighbor-Joining bootstrap sampling steps. Western white pine RGA sequences have been deposited in the GenBank database under accession numbers DQ144912 to DQ144971.

Identification of NBS-AFLP markers. A population of 128 individuals from the heterozygous tree (*Cr2/cr2*) #3566, described in a previous *Cr2* mapping study with random amplified polymorphic DNA (RAPD) markers (15), was used to map RGAs and NBS-AFLP markers to the *Cr2* linkage. To explore a possible linkage of RGAs with *Cr2*, PCR with RGA gene-specific primers and a modified AFLP technique (6) were used to reveal RGA and RGA-AFLP polymorphisms. For the development of PCR-based RGA markers, 38 pairs of RGA-specific primers were designed based on the divergent regions. Eighteen of these pairs were used for a previous RGA-expression study (14). Twenty other primer combinations came from 14 forward and 17 reverse primers that were designed to target the second intron in the western white pine CC-type NBS sequences (Table 1). PCR conditions were optimized to detect DNA polymorphisms in NBS sequences.

For RGA mapping with a modified AFLP method (6), genomic DNA of individual megagametophytes from the seeds of female tree #3566 was digested with restriction enzymes *EcoRI* and *MseI*, and ligated with adaptors as described by Vos et al. (23). Pre-amplification was conducted using *MseI*-CC and *EcoRI*-AC primers. Each pre-amplification reaction mixture consisted of 5 µl of diluted restriction/ligation DNA mixture as template, 0.4 µM of E-AC and M-CC primers, and 12.5 µl of 2× Taq Master Mix (Qiagen) in a total volume of 25 µl. For the second round of selective amplification, each of 17 NBS primers, including 5 degenerate primers for RGA cloning and 12 primers designed according to western white pine RGA sequences (Table 1), was coupled with one of 16 *MseI*-CC + 2 and 16 *EcoRI*-AC + 2 selective primers, respectively. In total, 544 primer combinations were screened in bulked segregation analysis (BSA). Each selective amplification reaction contained 1 µl of 50× diluted pre-amplification mixture as template, 0.4 µM of RGA primer, 0.4 µM of E-AC+2 or M-CC+2 primers, and 12.5 µl 2× Taq Master Mix (Qiagen) in a total volume of 25 µl.

To select NBS/AFLP primer combinations that could reveal NBS-AFLP polymorphic markers in the mapping population, bulked segregant analysis (BSA) and haploid segregation analysis were performed as described previously (15).

Selective amplification products were separated in a 7 M urea-6% polyacrylamide gel, and the amplicon patterns were visualized by silver staining. NBS-AFLP markers were scored as present (+) or absent (-). The NBS/AFLP data were treated as dominant markers and the polymorphic bands were mapped on the *Cr2* genetic linkage as described previously (15). The polymorphic NBS-AFLP markers were mapped relative to the existing RAPD markers using JoinMap version 3.0 software (J. W. Van Ooijen and R. E. Voorrips. 2001. Plant Res. Int., Wageningen, Netherlands) with the Kosambi mapping function provided by the software.

NBS-AFLP fragments of interest were excised from the gels, and incubated in an elution buffer (0.5 M NH₄Ac, 10 mM MgAc₂, 1 mM EDTA, 0.1% SDS) for 30 min at 50°C, then purified with a QIAquick gel extraction kit (Qiagen). Those fragments were re-amplified and cloned into the pGEM-T easy vector (Promega). The insert sizes of the recombinant plasmids were checked with PCR using the same primers, and verified by comparison with the original AFLP reactions on denatured sequencing gels.

The nucleotide sequence of each marker was translated into an amino acid sequence according to NBS primer direction, and searched for homologous NBS-LRR proteins by the BLASTP program of “search for short, nearly exact matches” (NCBI).

RESULTS AND DISCUSSION

PCR isolation of NBS genomic sequences. When primer GKT-d was combined with a mixture of primers HD1-d and HD2-d, PCR amplified five genomic fragments with estimated sizes of 1.5, 1.1, 0.9, 0.5, and 0.4 kb, whereas primer pair DVW-d/CFL-d amplified four genomic fragments with estimated sizes of 1.1, 0.85, 0.65, and 0.45 kb. Amplicons of 0.65 kb or longer were purified from agarose gels and cloned into the pGEM-T easy vector. The recombinant plasmids were grouped by restriction enzyme analysis. In total, 160 genomic clones were selected for sequence analysis. A BLASTX search showed that the 0.65- and 0.85-kb fragments from primer pair DVW-d/CFL-d represented various RGA sequences because all of the sequenced clones from these two fragments had, at the amino acid level, significant similarities with R genes or other RGAs of CC-NB-LRR subfamily. However, only a small portion of the 0.9-, 1.1-, and

1.5-kb genomic fragments from primers GKT-d/(HD1-d/HD2-d) showed homologies with NBS sequences, indicating that they were mixed with nonspecific PCR products. In case of possible PCR errors in RGA amplification, only RGA clones with nucleotide sequence identities of 98% or less were considered as distinct RGA members for further analysis. Forty-two distinct RGA genomic sequences were identified.

RT-PCR identification of expressed RGAs. When RT-PCR was used to amplify NBS sequences expressed in western white pine, one main fragment of approximately 0.7 or 0.9 kb was detected with primer pairs GKT-d/CFL-d and GKD-d/HD1-d. Sequence analysis of 86 cDNA clones revealed that all were expressed RGA sequences. Nineteen different RGA sequences with an identity threshold of 98% were identified at the mRNA level. The frequencies of the distinct RGA clones in all cDNA sequences ranged from 1.2 to 11.6%, indicating different expression levels among western white pine RGAs.

Nucleotide sequence analysis of western white pine RGAs. The nucleotide sequences of the 42 genomic RGAs were more divergent (generally 16 to 98% identities among themselves) than the 19 cDNA RGAs (70 to 98% identities among themselves). None of cDNA RGA sequences were identical to any genomic

TABLE 1. Oligonucleotide sequences used for the polymerase chain reaction cloning and genetic mapping of western white pine nucleotide-binding site–leucine-rich repeats (NBS-LRR) genes

Degenerate primers for RGA cloning, expression analysis, and detection of nucleotide-binding site–amplified fragment length polymorphism (NBS-AFLP) markers		
Primer name	Amino acid motif	Primer sequence (5' to 3') ^a
GKT-d (For.)	MGGLGKTT	ATGGGNGGNCNTGGNAARACNAC
DVW-d (For.)	(V/L)(I/V)LDDVW	KTRRTATTGGAYGAYGTMGTGG
CFL-d (Rev.)	CFLDLG(A/S)F	AAAWGMNCCCARRTCYAAARAARCA
HD1-d (Rev.)	QHDVMR	CYCGCATKACATCATGYTG
HD2-d (Rev.)	IHDELNRN	ATTNCGNAGYTCRTCTRGDGT

Western white pine resistance gene analog (RGA)-specific primers for NBS-AFLP markers detection

Primer name	Amino acid motif	Primer sequence (5' to 3') ^a
GKT2 (For.)	MPGVGKT	ATGCCTGGTGTGGAAAGAC
GKT3 (For.)	LGGVGKTT	TTAGGTGGTGTGGGAAAGACNAC
GKT4 (For.)	MGGLGKTT	ATGGGAGGTCTCGGTAAAACT AC
DVD1 (For.)	LLVLDDVD	TTGCTAGTTCTAGATGATGTAGA
DVD2 (For.)	VILDDVD	GTCATTTTAGATGATGTAGA
CPL1 (Rev.)	GLPLALK	CTTTAACGCGAGTGGTAAGCC
CPL2 (Rev.)	GLPLALK	CTTCAGAGCCAGTGGCAGACC
CFL1 (Rev.)	CFLDLGAF	AAAAGCCCCAAATCCAAAAAGCA
CFL2 (Rev.)	CFLDLGSF	AAATGACCCCAGGTCTAAGAAGCA
MHD5 (For.)	MHDHLQNM	ATGCACGATCACCTGCAAAAYATG
HD2 (Rev.)	IHDELNRN	ATTGCGCAGTTCGTCTGTGGAT
MG3 (Rev.)	DMGRDL	AAGATCTCGTCCCATATC

Western white pine RGA-specific primers to detect RGA polymorphism in the second NBS intron

Forward primer sequence (5' to 3')		Reverse primer sequence (5' to 3')	
F1	CAACCAGCTAGTCAAGCAG	R15	ACCTTTACATTCCGCTTGAC
F2	CAACCAACTAGTCAAGCAG	R16	ACCTTTACATTTCGCTTGAC
F3	CAACCAGCTAATCAAGTAG	R17	ACCTTTACGTTCCGCTTGAC
F4	CAAGCAGCTAGTCAAGCAG	R18	ACCTTTACATTCCACTTGAC
F5	CGAGCAACTAGTCAAGCAG	R19	ACCTCTACATTCTGCTTGAC
F6	TGAAACTCTAGTCAAGCAG	R20	GCCTTTACATTTCGCTTGAAC
F7	TGAAACTCTAGTATGGCAG	R21	GCCTTTACATTCTGCTTGAC
F8	TGAACATTCAGTCAAGCAG	R22	GCCTTTACATTCTACTTGAC
F9	GGAGAACTTAGTCAAGCAG	R23	AACTTTACATTGTGCTTGAC
F10	GGAGCAGCTAGTCAAGCAG	R24	GCCTTTACATTGCGCTTGAC
F11	GGAGTAGCTAGTCAAGCAG	R25	GCCTTTATATTCCACTTTCAC
F12	TGAGGATCTTGTTAAATG	R26	GCCTTTACATTCCACTTTCAC
F13	CGAAGTTGTTGTTAAACAG	R27	GCCCTTACACTCTGCTTCAAC
F14	TAAGGATCTTGTTCAACAG	R28	ACCCTTACAGACTGCTTCAAC
		R29	GCCACACGCTGCTGCAAC
		R30	GCCACAGGCTGCTGCAAC
		R31	ACCACCACACGCTTCTGCAAC

^a Degenerate IUB code : R = A+G, Y = C+T, M = A+C, K = G+T, W = A+T, D = G+A+T, and N = A+T+C+G.

RGA sequences. The highest identity between them was only 98%. One genomic RGA member, GKD11-7, shared exceptionally low identities of less than 10% with any of the other RGAs. Among these RGAs, 13 cDNA and 28 genomic RGAs encoded the NBS domain with a continuous open reading frame (ORF) after putative posttranscriptional splicing, whereas the other RGAs contained premature stop codons, reading frame shift mutations, or both, suggesting that 32% of the total RGA genes may be pseudogenes in the western white pine genome.

The 42 genomic RGAs were divided into two types according to gene structure. Thirty-eight of them, derived from the 1.5-kb fragment with primers GKT-d/(HD1-d/HD2-d), as well as from the 0.65- and 0.85-kb fragments with primers DVW-d/CFL-d, contained putative introns. The other four RGAs (GKD9-7, -13, -26, and KGD11-27), derived from the 0.9- and 1.1-kb fragments

with primers GKT-d/(HD1-d/HD2-d), showed no putative splicing site when compared with cDNA sequences. Alignment analysis of genomic sequences of western white pine RGAs with an *Arabidopsis* CC-NBS-LRR-type R gene (*ADRI*) revealed that the positions of three introns in NBS sequences are conserved between gymnosperms and angiosperms. Most 5'-exon/intron and 3'-intron/exon boundaries conformed to the well-known GT/AG donor/acceptor site rule; however, a few rare donor/acceptor sites were also identified in western white pine NBS sequences with putative ORFs, including GC/AG, AT/AG, GT/GG, and GT/AC. The introns ranged from 161 to 468 bp in length, and were AT-rich, an essential characteristic for pre-mRNA processing. The length of the second intron divided these 38 genomic RGAs into two types. One type had a long intron (399 to 468 bp); the other had a short intron (182 to 247 bp).

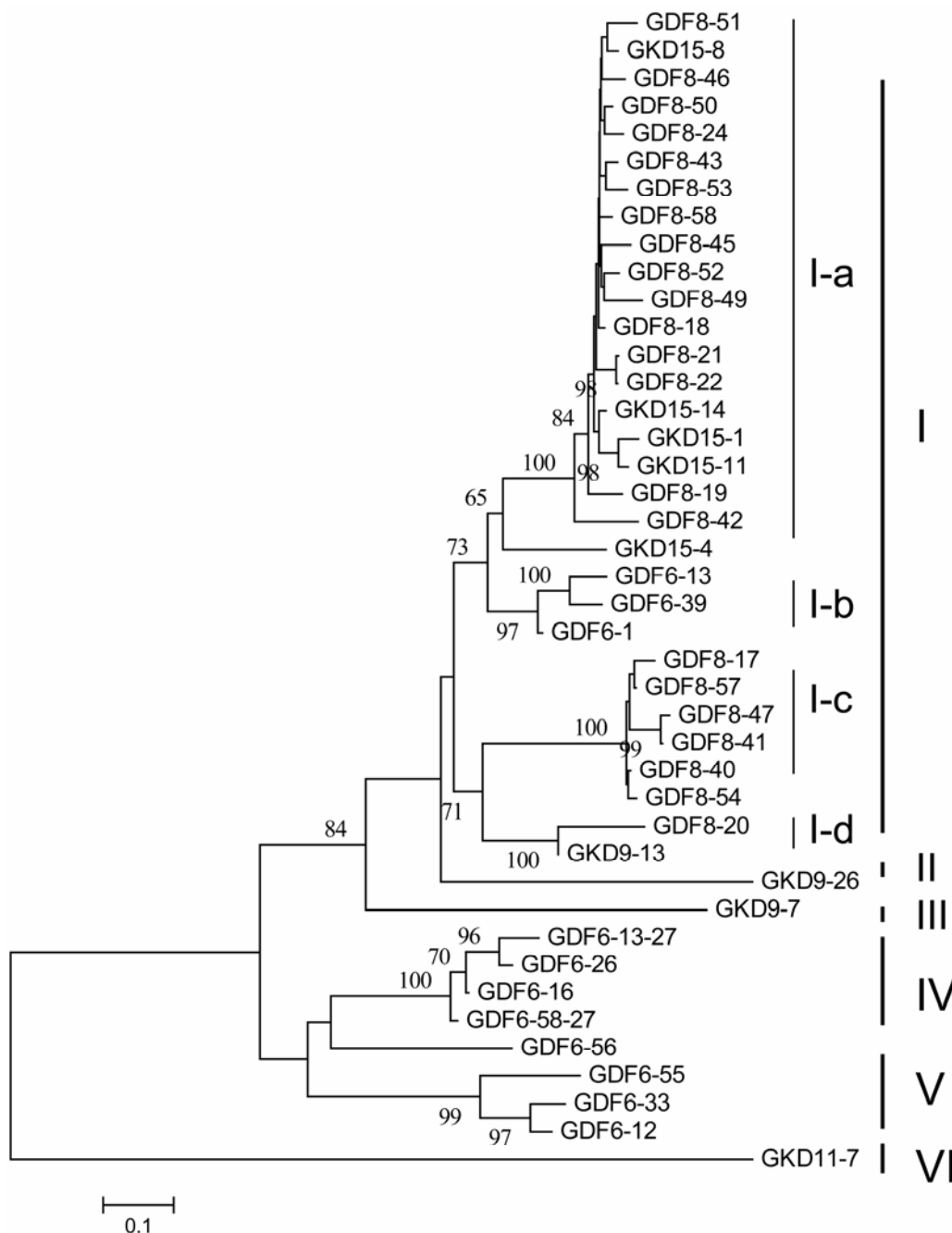


Fig. 1. Phylogenetic tree constructed from the nucleotide sequence alignment analysis of the western white pine resistance gene analogs (RGAs) identified in this study. The tree was generated based on Clustal W (Eur. Bioinformatics Inst., Cambridge, UK) alignment analysis of 42 genomic sequences using a Neighbor-Joining method. A bootstrap of 1,000 replications was employed to evaluate the reliability of the tree branching. The numbers I to VI indicate classification of the RGAs into classes with a 75% identity threshold. The scale at the bottom left indicates genetic distance proportional to the nucleotide substitutions per site.

The phylogenetic tree from the nucleotide sequence alignment of 42 genomic clones is composed of six monophyletic clusters (I to VI), with general identity below 75% between classes (Fig. 1). Class I is the most complex, and is subdivided into four sub-classes (I-a to I-d). Sub-classes I-a, I-c, and I-d consist of RGA members with long second introns, whereas subclass I-b, as well as classes IV and V, contain RGAs that have only short second introns. The intron-free RGA members constitute the single-member classes II (GKD9-26), III (GKD9-7), and VI (GKD11-27), with the exception of GKD9-13 in subclass I-d.

Comparative analyses of putative RGA proteins with R proteins. Nucleotide sequences of western white pine RGAs were translated into amino acid sequences to search for characteristic motifs in NBS domains. The western white pine NBS sequences were identified as RGAs belonging to the CC-NBS-LRR subfamily by the presence of characteristic motifs of this subfamily. One of the western white pine RGAs (CKD24-1) shared highest identity of 98% with a putative CC-NBS-LRR protein recently reported in sugar pine (9). Multiple alignments of deduced amino acid sequences (Fig. 2) revealed that the western white pine RGAs contained the conserved motifs that are common in the NBS region of angiosperm NBS-LRR proteins (20). Among these conserved motifs, the P-loop/kinase-1 and GLPL motifs are present in all NBS-LRR proteins, whereas kinase-2 and RNBS-D are very similar to those motifs found exclusively in the CC-NBS-LRR subfamily (20). A tryptophan (W) residue at the last position in the kinase-2 motif of western white pine NBS sequences, DD(I/V)W, strongly indicates that these sequences probably belong to the CC-NBS-LRR subfamily, as most angiosperm CC-NBS-LRR proteins have a tryptophan (W), whereas most TIR-NBS-LRR proteins have an aspartic acid (D) at this position (20).

At the deduced amino acid level, western white pine CC-NBS-LRR-type RGAs shared from 13 to 98% identity among themselves, and from 5 to 28% identity with the western white pine TIR-NBS-LRR reported previously. In a consensus phylogenetic tree constructed from 90 aligned amino acid sequences, western white pine RGA sequences were clearly grouped into two monophyletic branches corresponding to the TIR-NBS-LRR and CC-NBS-LRR subfamilies (Fig. 3). The western white pine RGAs described previously (14) were clustered with Nt-N, Lu-L, and Lu-M, and belonged to the TIR-NBS-LRR subfamily, whereas the RGAs identified in present study were grouped into the CC-NBS-LRR subfamily (Fig. 3). The subtree of the CC-NBS-LRR subfamily included AtADR1, AtPRS5, HvMla10, OsMla1, OsRPR1, and ZmRP3 from the CC-NBS-LRR subfamily of angiosperms, and is very similar to the phylogenetic tree constructed from the alignment of nucleotide sequences in which western white pine RGA proteins are further subgrouped into six classes (I to VI) that may have evolved from different ancestor genes. The amino acid alignment analysis showed that amino acid identity between classes is significantly lower (13 to 68%) than that within the same class with multiple members (68 to 98%). Classes I to IV are grouped with AtADR1 (AJ581996) and its homologues PtAAM28916, AtNP196092, and AtNP195056. In these classes western white pine RGAs show from 30 to 44% amino acid identity to AtADR1. The single-member class VI (GKD11-27) is clearly separated from all other western white pine RGAs, and is grouped with CC-NBS-LRR-type R protein AtRPS5 (AF074916) and others; its identities to other western white pine RGAs range only from 13 to 27%.

Interestingly, the amino acid sequence identities among these RGAs do not always reflect their source organisms. Most western white pine CC-NBS-LRR-type RGAs showed greater similarity to *Arabidopsis* R protein AtADR1 than to western white pine RGA GKD11-7. For example, the NBS motif RNBS-B is highly conserved between AtADR1 and GKD9-26, but divergent in GKD11-7. A similar result was observed for the western white pine TIR-NBS-LRR-type RGA GI5033, which has higher identity to angiosperm R proteins (Nt-N, Lu-L, and Lu-M) than to other

conifer RGAs despite the evolutionary distance between gymnosperms and angiosperms (14). The sequence alignments and subsequent phylogenetic analyses indicate that some western white pine RGA clusters are ancient and may have arisen from gene duplication events prior to the differentiation of the last common ancestor of gymnosperms and angiosperms about 325 million years ago.

The presence of both TIR-type and CC-type NBS-LRR genes in western white pine indicates that the organization and evolution of the NBS-LRR superfamily in conifers may be different from that in monocot species and may be more similar to that in dicot species. Monocots, such as rice, contain only CC-type NBS-LRR genes (25). High levels of molecular divergence have been detected in both TIR-NBS-LRR and CC-NBS-LRR subfamilies in western white pine. The western white pine TIR-NBS-LRR subfamily contained at least 67 members that could be classified into 14 classes with amino acid sequence similarity less than 75% (14). This number of RGA classes may reflect a difference in the genomic abundance between these two NBS-LRR subfamilies in western white pine. The reported ratio of TIR-type and CC-type NBS-LRR genes is about 2:1 (94 versus 55) in the whole genome of *Arabidopsis* (21). The number of western white pine RGA classes (14 TIR-type and 6 CC-type) is higher than the 11 reported for the angiosperm soybean (24) and the 10 reported for the angiosperm cotton (7). Based on current data, it is too early to say whether this is due to a richer abundance of NBS-LRR genes in the *P. monticola* genome or due to limited PCR primers used for RGA cloning in angiosperm species.

All RGAs of CC-NBS-LRR subfamily characterized at the mRNA level are grouped in the subclasses I-a and I-d within class I, indicating that the mRNA expression of the other five classes is below the level we could detect. Some expressed RGAs from seed lot #3278 showed highest identity of 98% to genomic clones from seed lot #3277, implicating that those RGA clones might be alleles of the same RGA because of their different genetic backgrounds. Note that the absence of cDNA clones from subclass I-b and I-c, and classes II, IV, V, and VI does not rule out a potential function for the RGAs in these groups, because the groups do contain RGA member(s) with a continuous ORF. Class III has only one RGA, GKD9-7, which is presumed to be a pseudogene due to the presence of both a premature stop codon and an ORF shift. Continued expansion of our knowledge about the distribution, organization, and functional genomics of the NBS-LRR superfamily in a conifer species would give us a better understanding of forest pathosystems and the evolution of tree defense systems.

Identification of RGA-AFLP markers linked with disease resistance loci (*Cr2*). Three pairs of primers specific for RGA members GM15, G9034, and GDF8-20 revealed DNA polymorphisms in the megagametophyte population of tree #3566. For the other 35 pairs of RGA-specific primers, only fragments with expected sizes, some accompanied by weak nonspecific products, were amplified in all individuals (data not shown). After genotyping the entire population, RGA members GM15, G9034, and GDF8-20 were mapped to genetic linkages other than the *Cr2* linkage.

For the modified AFLP strategy, about 30% of 544 primer combinations revealed differences between resistance and susceptible bulks in BSA. The NBS-AFLP marker was confirmed in the individuals within bulks. The linkage of each DNA marker to *Cr2* was tested with the entire population of 128 megagametophyte individuals. A representative example of a *Cr2*-linked NBS-AFLP fragment, NBS/M19-231, is shown in Figure 4. A total of 11 new NBS-AFLP markers, plus the 5 RAPD loci identified previously (15), now constitute a genetic map of the *Cr2* region (Fig. 5). In this *Cr2* linkage, two NBS-AFLP markers (NBS/E14-161r and NBS/E109-335r) are linked to *Cr2* in repulsion and the other nine are linked in coupling. Markers NBS/E199-90 and NBS/E33-665 were mapped closest to either side of *Cr2* at genetic distances of 0.41 and 1.22 cM, respectively.

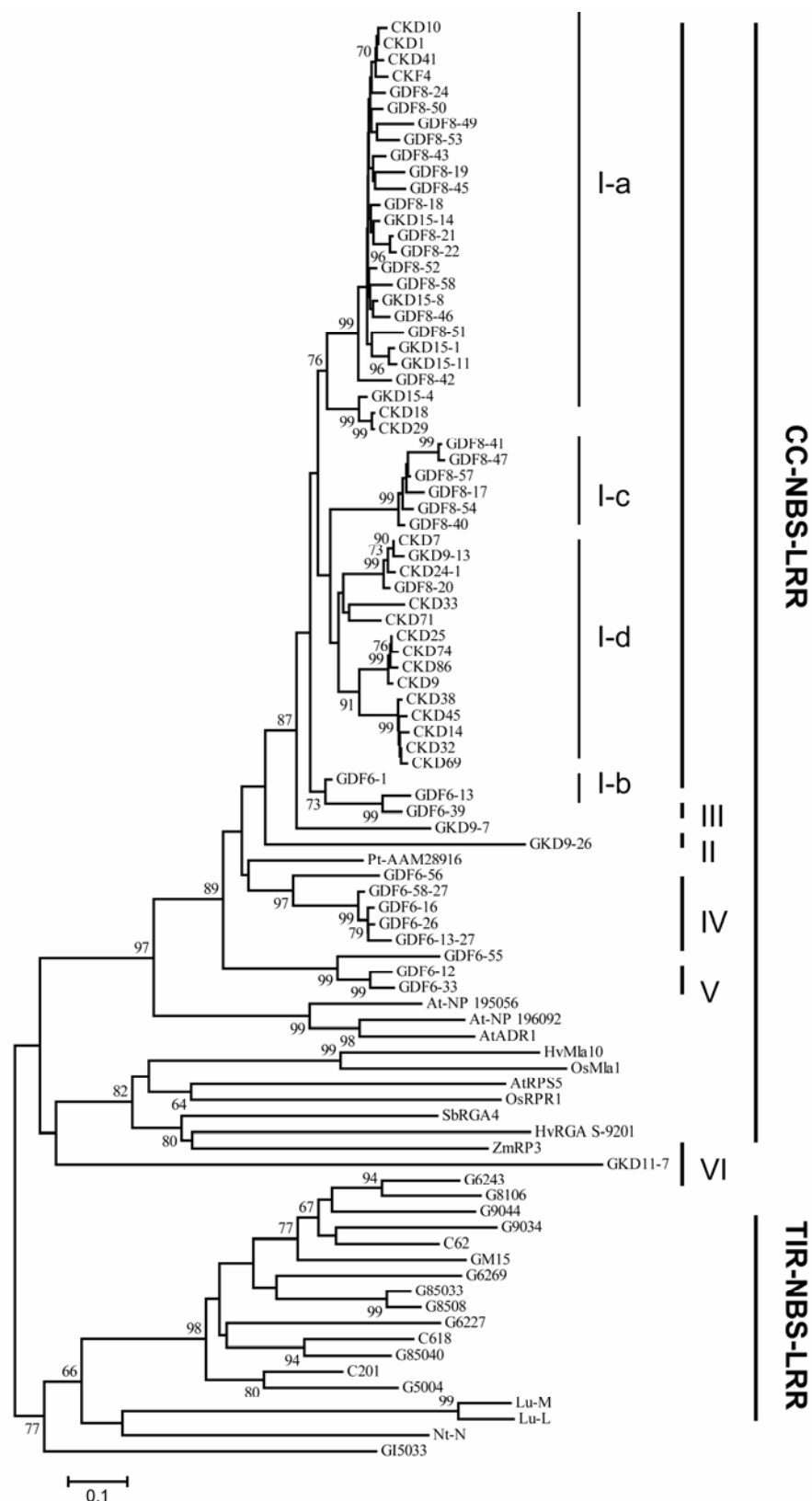


Fig. 3. Phylogenetic tree of amino acid sequences of western white pine resistance gene analogs (RGAs) and selected R proteins from angiosperms. Western white pine RGA sequences were aligned to additional nucleotide-binding site (NBS)-leucine-rich repeats (LRR) proteins, including *Nicotiana glutinosa* Nt-N (A54810), *Linum usitatissimum* Lu-L5 (AF093645) and Lu-M (AAB47618), *Arabidopsis* AtADR1 (AJ581996) and AtRPS5 (AF074916), *Oryza sativa* OsRPR1 (BAC55682) and Os Mla1 (BAC98548), *Hordeum vulgare* HvMla10 (AY266445) and Hvrga (S-9201, CAD45029), *Zea mays* ZmRP3 (AAN23084), *Solanum bulbocastanum* SbRGA4 (AAR29072), and *Pinus taeda* Pt-AAM28916. Western white pine RGAs in the CC-NBS-LRR branch are subgrouped into six classes with amino acid sequence identities less than 68% between members of different clusters. The TIR-NBS-LRR branch includes representative members of western white pine RGA classes with amino acid identities less than 75% between them (14). The scale at the bottom left indicates genetic distance proportional to the amino acid substitutions per site.

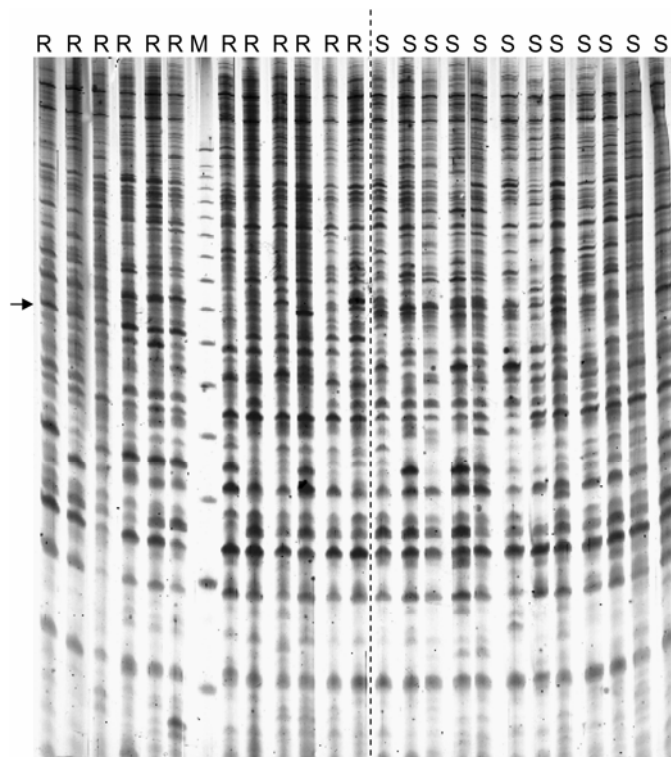


Fig. 4. Band pattern of polymerase chain reaction (PCR) products amplified from primer pair GKT4/Mccgc. The arrow indicates the 231-bp size of the nucleotide-binding site-amplified fragment length polymorphism (NBS-AFLP) marker NBS-M19-231. Samples of 12 resistant individuals (R) and 12 susceptible individuals (S) are shown here. A molecular marker of 25-bp DNA ladder (M) from Invitrogen Canada Inc. (Burlington, ON) was used to estimate AFLP amplicon sizes.

After this step, the 11 NBS-AFLP markers were cloned and sequenced. Sequence analysis demonstrated that 10 markers contained NBS-specific and AFLP-selective primers at their 5'- and 3'- ends. Thus, all but one of the markers contained amino acid sequences of the conserved NBS motifs corresponding to NBS primers. That one marker, NBS/E51-750, was a nonspecific AFLP fragment amplified from one single AFLP selective primer, Eacag (Table 2). Nucleotide sequence alignment analysis of 10 NBS-AFLP markers, 61 CC-type, and 67 TIR-type western white pine RGAs showed 97% identity between markers NBS/E41-121 and NBS/E217-118, 53% identity between these two markers and cluster IV members of the CC-NBS-LRR subfamily as well as the expressed TIR-type RGA C601, 44% identity between marker NBS/E14-161 and TIR-type RGA G8029, and 42% identity between marker NBS/E199-90 and TIR-type RGA G8106. The BLASTP program of "search for short, nearly exact matches" found that eight markers encoded additional amino acid sequences extended from NBS primers that were significantly homologous with NBS-LRR sequences, whereas the other two markers, NBS/M19-231 and NBS/E199-90, contained stop codons immediately following the NBS primer sequence, and showed no homology with any R or RGA proteins (Table 2).

The low level of NBS-AFLP marker homologies with R genes or with RGAs from angiosperms may be caused by a few factors. Firstly, most NBS-AFLP markers are short, with limited ORFs. The marker ORFs are difficult to determine because of the presence of introns in most NBS sequences and the higher ratios of RGA pseudogenes in western white pine (32% of CC-NBS-LRR and 47% of TIR-NBS-LRR) than in angiosperms, with 25% in cotton (7) and 10% in *Arabidopsis* (21). Long-distance genomic DNA walking is necessary to definitively identify western white pine NBS-AFLP markers.

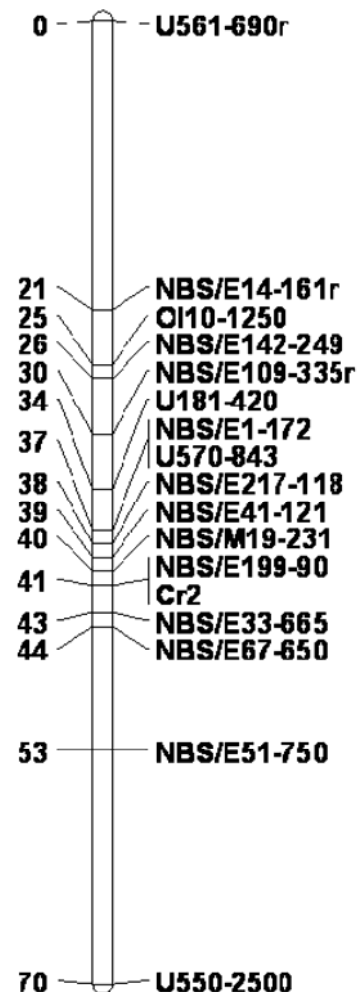


Fig. 5. Genetic map of the *Cr2* linkage. Five random amplified polymorphic DNA (RAPD) loci were described and mapped previously in Liu et al. (15). Genetic distance is indicated on the left in Kosambi cM.

The NBS-AFLP marker closest to *Cr2* is NBS/E199-90, localized 0.41 cM from *Cr2*; this suggests that it is several Mb from *Cr2* on a physical map. To facilitate physical mapping of the area, the linked markers should be located no further than 0.01 cM away from *Cr2*. To achieve this resolution, larger populations with thousands of individuals from different genetic backgrounds will need to be employed for a fine and comparative RGA mapping of the *Cr2* region. The present study failed to observe perfect co-segregation between a NBS-AFLP locus and *Cr2*. More NBS primers or primers targeting the variable LRR domain will need to be screened to identify RGA localized at the *Cr2* locus. It is also possible that the *Cr2* gene may not be a member of the NBS-LRR family. In future studies, further genomic searching and mapping of other R families in western white pine, including *Cf*, *Pto*, *Xa21*, and *RPW8* analogs, may help to answer this question. Several RGA loci have been mapped near or at the same position of known resistance genes (11). Therefore, the putative RGAs tightly linked to *Cr2* may be taken as entry points for further characterization of the genomic organization of conifer R genes. This is especially useful for R gene identification in conifer species with large genomes, such as western white pine.

ACKNOWLEDGMENTS

This research was supported by the Forest Genomics funding from Natural Resources Canada, Canadian Forest Service. We thank D. Patel for technical assistance on AFLP gel electrophoresis, and R. Sturrock, P. Dale, and H. Williams for thoughtful comments on the manuscript.

TABLE 2. Characteristics of the nucleotide-binding site–amplified fragment length polymorphism (NBS-AFLP) markers in the genetic linkage of the western white pine *Cr2* locus using a modified AFLP procedure

Marker name ^a	Primer pairs	Putative short amino acid sequence	Homologous resistance gene analogs (RGAs) [Species] (GenBank accession number)	Homology score (bits)
NBS/M19-231	GKT4/Mccgc	MGGLGKTT* ^b	NS ^c	---
NBS/E1-172	GKT2/Eacaa	MPGVGKTGGNLCYD	R-like protein [<i>Populus tremula</i>] (AAZ30322.1)	28.6
NBS/E14-161	GKT2/Eactc	MPGVGKTTLSLGL	R protein PLTR [<i>Arachis hypogaea</i>] (AAH81254.1)	30.3
NBS/E33-665	DVD1/Eacaa	VVVDDVDGVADLYGVLEALA	NBS-LRR [<i>Malus x domestica</i>] (AAM77249.1)	29.9
NBS/E41-121	DVD1/Eacga	LVVVDDVDNCKDRE	TIR-NBS protein 4 [<i>Malus baccata</i>] (AAQ93076.1)	29.1
NBS/E51-750	Eacag/Eacag	---	---	---
NBS/E67-650	CPL1/Eacag	TSMQIGLPLALK	R-like protein [<i>Aegilops ventricosus</i>] (CAC11104.1)	26.9
NBS/E109-335	CFL1/Eacta	HEGFIVKFLHFMFSFNTCFLDLGAF	NBS/LRR [<i>Pinus taeda</i>] (AAM28916.1)	29.5
NBS/E142-249	HD1-d/Eactc	YMGIIYDQHDVMR	R-like protein [<i>Pinus lambertiana</i>] (AAY78890.1)	26.5
NBS/E199-90	GKT3/Eaccg	GGVGKTT*	NS	---
NBS/E217-118	DVD2/Eacga	VILDDVDNCKDRE	R protein homolog [<i>Vigna unguiculata</i>] (AAD34882.1)	29.5

^a The second number in the marker indicates the marker DNA fragment length in base pairs.

^b * = stop codon.

^c NS = no RGA with significant homology was detected using “search for short, nearly exact matches” program in BLASTP (National Center for Biotechnology Information, Bethesda, MD).

LITERATURE CITED

- Belkadir, Y., Subramaniam, R., and Dangl, J. L. 2004. Plant disease resistance protein signaling: NBS-LRR proteins and their partners. *Curr. Opin. Plant Biol.* 7:391-399.
- Bennett, M. D., and Leitch, I. J. 2003. Plant DNA C-values database (release 2.0, January 2003). Royal Botanic Gardens, Kew, U. K.
- Chen, X. M., Line, R. F., and Leung, H. 1998. Genome scanning for resistance-gene analogs in rice, barley, and wheat by high-resolution electrophoresis. *Theor. Appl. Genet.* 97:345-355.
- Dangl, J. L., and Jones, J. D. 2001. Plant pathogens and integrated defense responses to infection. *Nature* 411:826-833.
- Flor, H. H. 1971. Current status of the gene-for-gene concept. *Annu. Rev. Phytopathol.* 9: 275-296.
- Hayes, A. J., and Maroof, M. A. S. 2000. Targeted resistance gene mapping in soybean using modified AFLPs. *Theor. Appl. Genet.* 100:1279-1283.
- He, L., Du, C., Covaleta, L., Xu, Z., Robinson, A. F., Yu, J. Z., Kohel, R. J., and Zhang, H. B. 2004. Cloning, characterization, and evolution of the NBS-LRR-encoding resistance gene analogue family in polyploid cotton (*Gossypium hirsutum* L.). *Mol. Plant-Microbe Interact.* 17:1234-1241.
- Hunt, R. S., Jensen, G. D., and Ekramoddoullah, A. K. M. 2004. Confirmation of dominant gene resistance (*Cr2*) in U.S. white pine selection to white pine blister rust growing in British Columbia. Pages 227–229 in: *Breeding and genetic resources of five-needle pines: Growth, adaptability, and pest resistance*. Proc. IUFRO Working Party 2.02.15 Int. Conf. R. A. Snieszko, S. Samman, S. E. Schlarbaum, and H. B. Kriebel, eds. Proc. RMRS-P32. USDA For. Serv., Rocky Mountain Res. Stn., Fort Collins, CO.
- Jermstad, K. D., Sheppard, L. A., Kinloch, B. B., Delfino-Mix, A., Ersoz, E. S., Krutovsky, K. V., and Neale, D. B. 2006. Isolation of a full-length CC–NBS–LRR resistance gene analog candidate from sugar pine showing low nucleotide diversity. *Tree Genet. Genomes* 2:76-85.
- Johnk, N., Hietala, A. M., Fosdall, 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.* 25:1533-1543.
- Kanazin, V., Marek, L. F., and Shoemaker, R. C. 1996. Resistance gene analogs are conserved and clustered in soybean. *Proc. Natl. Acad. Sci. USA.* 93(21):11746-11750.
- Kinloch, B. B., Snieszko, R. A., Barnes, G. D., and Greathouse, T. E. 1999. A major gene for resistance to white pine blister rust in western white pine from the western Cascade Range. *Phytopathol.* 89:861–867.
- Kumar, S., Tamura, K., and Nei, M. 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* 5:150-163.
- Liu, J.-J., and Ekramoddoullah, A. K. M. 2003. Isolation, genetic variation and expression of TIR-NBS-LRR resistance gene analogs from western white pine (*Pinus monticola* Dougl. ex. D. Don.). *Mol. Genet. Genomics* 270:432-441.
- Liu, J.-J., Ekramoddoullah, A. K. M., Hunt, R. S., and Zamani, A. 2006. Identification and characterization of RAPD markers tightly linked to a major gene (*Cr2*) for resistance to *Cronartium ribicola* (Fish.) in *Pinus monticola* (D. Don). *Phytopathol.* 4:395-399.
- Liu, J.-J., Ekramoddoullah, A. K. M., and Yu, Y. 2003. Differential expression of multiple PR10 proteins in western white pine following wounding, fungal infection, and cold-hardening. *Physiol. Plant.* 119:544-553.
- Liu, J.-J., Hunt, R., and Ekramoddoullah, A. K. M. 2004. Recent insights into genetic resistance of western white pine to white pine blister rust. *Recent Res. Devel. Biotech. Bioeng.* 6:65-76.
- Martin, G. B., Bogdanove, A. J., and Sessa, G. 2003. Understanding the functions of plant disease resistance proteins. *Ann. Rev. Plant Biol.* 54: 23-61.
- Martin, G. B., Brommonschenkel, S. H., Chunwongse, J., Fray, A., Ganai, M. W., Spivey, R., Wu, T., Earle, E. D., and Tanksley, S. D. 1993. Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science* 262:1432-1436.
- Meyers, B. C., Dickerman, A. W., Michelmore, R. W., Sivaramakrishnan, S., Sobral, B. W., and Young, N. D. 1999. Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J.* 20:317–332.
- Meyers, B. C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R. W. 2003. Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* 15:809-834.
- Ujino-Ihara, T., Yoshimura, K., Ugawa, Y., Yoshimaru, H., Nagasaka, K., and Tsumura, Y. 2000. Expression analysis of ESTs derived from the inner bark of *Cryptomeria japonica*. *Plant Mol. Biol.* 43:451–457.
- Vos, P., Hogers, R., Bleeker, M., Reijmans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., and Kuiper, M. 1995. AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Res.* 23:4407-4414.
- Yu, Y. G., Buss, G. R., Maroof, M. A. S. 1996. Isolation of a superfamily of candidate disease-resistance genes in soybean based on a conserved nucleotide-binding site. *Proc. Natl. Acad. Sci. USA.* 93:11751-11756.
- Zhou, T., Wang, Y., Chen, J.-Q., Araki, H., Jing, Z., Jiang, K., Shen, J., and Tian, D. 2004. Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Mol. Genet. Genomics* 271:402-415.