

J.-J. Liu · A. K. M. Ekramoddoullah

Isolation, genetic variation and expression of TIR-NBS-LRR resistance gene analogs from western white pine (*Pinus monticola* Dougl. ex. D. Don.)

Received: 24 June 2003 / Accepted: 30 September 2003 / Published online: 28 October 2003
© Springer-Verlag 2003

Abstract Western white pine (*Pinus monticola* Dougl. ex. D. Don., WWP) shows genetic variation in disease resistance to white pine blister rust (*Cronartium ribicola*). Most plant disease resistance (R) genes encode proteins that belong to a superfamily with nucleotide-binding site domains (NBS) and C-terminal leucine-rich repeats (LRR). In this work a PCR strategy was used to clone R gene analogs (RGAs) from WWP using oligonucleotide primers based on the conserved sequence motifs in the NBS domain of angiosperm NBS-LRR genes. Sixty-seven NBS sequences were cloned from disease-resistant trees. BLAST searches in GenBank revealed that they shared significant identity to well-characterized R genes from angiosperms, including *L* and *M* genes from flax, the tobacco *N* gene and the soybean gene *LM6*. Sequence alignments revealed that the RGAs from WWP contained the conserved motifs identified in angiosperm NBS domains, especially those motifs specific for TIR-NBS-LRR proteins. Phylogenetic analysis of plant R genes and RGAs indicated that all cloned WWP RGAs can be grouped into one major branch together with well-known R proteins carrying a TIR domain, suggesting they belong to the subfamily of TIR-NBS-LRR genes. In one phylogenetic tree, WWP RGAs were further subdivided into fourteen clusters with an amino acid sequence identity threshold of 75%. cDNA cloning and RT-PCR analysis with gene-specific primers demonstrated that members of 10 of the 14 RGA classes were expressed in foliage tissues, suggesting that a large and diverse NBS-LRR gene family may be functional in conifers. These results provide evidence for the

hypothesis that conifer RGAs share a common origin with R genes from angiosperms, and some of them may play important roles in defense mechanisms that confer disease resistance in western white pine. Ratios of non-synonymous to synonymous nucleotide substitutions (Ka/Ks) in the WWP NBS domains were greater than 1 or close to 1, indicating that diversifying selection and/or neutral selection operate on the NBS domains of the WWP RGA family.

Keywords Disease resistance gene analog (RGA) · *Cronartium ribicola* · *Pinus monticola* · NBS-LRR genes · Selection pressure

Introduction

Western white pine (*Pinus monticola* Dougl.) is an important species of the forest ecosystem in western North America. White pine blister rust (*Cronartium ribicola* J.C. Fisch.) has caused widespread mortality in western white pine (WWP) and other five-needle pines since it was introduced to North America around 1910. The rust has had a severe impact on forest ecosystems and has resulted in reduced planting of *P. monticola* and other five-needle pines. Resistance to *C. ribicola* and maintenance of genetic diversity in production populations are very important for WWP breeding. Currently genetic resistance is the key management tool that provides an approach to re-establishing five-needle pines in North America. In western white pine, several types of resistance to white pine blister rust have been identified, involving, e.g., slow canker growth (SCG) (Hunt 1997) and dominant disease resistance genes (*Cr2*) (Kinloch et al. 1999). These resistance resources are currently being used in white pine breeding programs in British Columbia (BC). An SCG resistance is expressed in the stems of western white pine seedlings that restricts infection area, and sometimes the canker heals over. Different types of SCG have been observed, suggesting that this disease resistance may be controlled by multiple

Communicated by R. Hagemann

J.-J. Liu · A. K. M. Ekramoddoullah (✉)
Pacific Forestry Centre,
Canadian Forest Service,
Natural Resources Canada,
506 West Burnside Road,
V8Z 1M5, Victoria,
British Columbia, Canada
E-mail: aekramodoul@pfc.forestry.ca

genes (Ekramoddoullah and Hunt 2002). On the other hand, *Cr2*-mediated resistance is a qualitative trait and occurs in the needles, setting off a classical hypersensitive reaction (HR), preventing stem infection by *C. ribicola*.

Plant disease-resistance genes (R genes) are postulated to encode receptors for the specific recognition of avirulence gene (Avr) products synthesized by the pathogen. The interaction of R genes and avirulence genes from pathogens is presumed to trigger a network of signal-transduction cascades that result in a series of plant defense processes, including the induction of pathogenesis-related (PR) proteins, which ultimately lead to disease resistance (Innes 1998). In recent years, several approaches have been adopted to investigate plant defense mechanisms in conifer pathosystems during fungal infection (Ekramoddoullah and Hunt 2002). An analysis of protein profiles was used to search for differentially expressed proteins, including some PR proteins, from resistant and susceptible conifer seedlings during pathogen infection (Ekramoddoullah and Hunt 1993; Ekramoddoullah et al. 1998). A significantly increased level of PR10 protein was observed in needles of resistant sugar pine (carrying the dominant resistance gene *Cr1*) following infection by *C. ribicola* (Ekramoddoullah and Tan 1998). A more detailed investigation revealed that multiple PR10 proteins and genes are differentially expressed in WWP in response to fungal infection, wounding and low temperatures (Liu et al. 2003). Immunocytochemical analysis demonstrated the localization of PR10 protein around the cells of the blister rust fungus *C. ribicola* in infected spots of the susceptible WWP needles. PR10 protein was also found, but the pathogen itself was no longer detectable in the hypersensitive infection spots on sugar pine. Histochemical examination revealed an extra layer of endodermal cells surrounding the vascular bundle at the hypersensitive infection spots in the white pine needles (Ekramoddoullah and Hunt 2002). This type of physical barrier induced by infection is proposed to contribute to rust resistance in conifers. Other PR proteins such as endochitinases, thaumatin-like proteins, and a putative anti-fungal protein have also been identified in conifers following fungal infection (Davidson and Ekramoddoullah 1997, Robinson et al. 2000, and unpublished data).

Over 30 disease resistance genes have now been cloned from model and crop species of angiosperms; they confer resistance to various plant pathogens, including viruses, bacteria, oomycetes, fungi, nematodes and insects (Ellis and Jones 1998, Ferrier-Cana et al. 2003). R genes are grouped into five types based on the structure of their protein products (Dangl and Jones 2001). Most R genes encode proteins containing a nucleotide-binding site (NBS) domain and a stretch of leucine-rich repeats (LRR) at the C-terminus. Other R proteins include those with only an LRR domain and no NBS domain, an LRR domain and a serine-threonine protein kinase domain without a NBS domain, a protein kinase domain without an NBS or an LRR domain, or a coiled-coil (CC) domain plus a putative signal anchor at the N-terminus. Based on

motifs located N-terminal to the NBS domain, the NBS-LRR superfamily can be further subdivided into the TIR (*Drosophila* Toll and mammalian interleukin-1 receptor homology regions) subfamily and the non-TIR subfamily, members of which typically contain N-terminal CC domains (Meyers et al. 1999; Pan et al. 2000a; Young 2000; Cannon et al. 2002). The NBS-LRR genes are abundant in plants. A total of 166 NBS-LRR genes are present in the genome of *Arabidopsis thaliana* ecotype Col-0, 70% of which belong to the TIR-NBS-LRR subfamily (Richly et al. 2002).

Eight conserved motifs have been identified in the NBS domains (Meyers et al. 1999), some of which are specific to the non-TIR or the TIR-NBS-LRR subfamily. These motifs have been used to synthesize degenerate primers for the isolation of disease resistance gene analogues (RGAs) containing NBS sequences from both dicot and cereal genomes (Cannon et al. 2002). This PCR cloning strategy has been employed to clone RGA sequences from soybean (Kanazin et al. 1996, Yu et al. 1996, Peñuela et al. 2002), potato (Leister et al. 1996), lettuce (Shen et al. 1998), bean (Creusot et al. 1999, Rivkin et al. 1999), Arabidopsis (Aarts et al. 1998, Speelman et al. 1998), maize (Collins et al. 1998), rice and barley (Leister et al. 1998), and other angiosperm species.

Our long-term objective is to isolate the *Cr2* gene, which confers resistance against *C. ribicola* to western white pine. This has prompted us to clone RGAs from this gymnosperm species. In the present study, we used a PCR cloning strategy with degenerate primers to successfully amplify NBS sequences from the white pine genome. Sequence determinations and phylogenetic analyses were performed to elucidate their relationships with well-known R genes in angiosperms. The RGAs from western white pine were investigated for their diversity and expression, in order to gain insights into the evolution of TIR-NBS-LRR genes and their potential for use in disease resistance breeding.

Materials and methods

Plant material

Western white pine seed lots #3277 and #3278, both known to contain the *Cr2* gene, were obtained from Dorena, USA (Kinloch et al. 1999; Registration Nos. 119-15045-845 X OP and 119-15045-845 X 15045-841 X OP), sown and grown at the Pacific Forestry Centre in Victoria (British Columbia) in 1994. They were inoculated in the *ribes* (disease) garden. Survivors were confirmed as resistant by re-inoculation the following year. Needles from these survivors were collected in 2001 for the isolation of genomic DNA and total RNA.

PCR and molecular cloning

Six oligonucleotide primers were designed for PCR cloning of RGA sequences from western white pine. These primers targeted three highly conserved motifs (P-loop/kinase-1, Kinase-2, GLPL) and one less conserved motif (DMGRDL) in the NBS domain of well-characterized resistance genes from angiosperm plants (Table 1). Because highly degenerate primer sets, such as Gkt2/Pal2, amplified

Table 1 Nucleotide sequences of primers used for the amplification and cloning of RGA sequences from western white pine

Conserved motif	Primer name	Primer sequence (5' to 3')	Degree of degeneracy
GGVGKTT (P-loop)	Gkt1 (Forward)	GGNGTNGGNAARACNAC	1/512
	Gkt2 (Forward)	GGIGGIGTIGGIAARACIAC	1/2048
VILDDVD (Kinase-2)	Dvd1 (Forward)	GTCATTTTAGATGATGTAGA	1/1
	Pal1 (Reverse)	CAAAGATAAGGGTAACCC	1/1
GLPL(S/A)L	Pal2 (Reverse)	IARIGCIARIGGIARICC	1/32768
	Dmg1 (Reverse)	AAGATCTCGTCCCATATC	1/1
DMGRDL			

a high proportion of non-specific fragments of about the same sizes as the RGAs, the primers Dvd1, Pal1 and Dmg1 were designed as unique sequences to avoid non-specific amplification (Table 1). A nested primer strategy was applied in some PCR experiments using primers for more than two conserved motifs. In the experiments with the nested primer approach, the diluted first-round PCR products were used as templates for the second-round PCR with every possible primer combination for two internal motifs, or for one internal motif and one motif targeted in the first-round PCR.

Genomic DNA was prepared from current-year needles of individual seedlings using a DNeasy kit (Qiagen, Mississauga, Ont., Canada). PCR was performed using a PCR Master Mix Kit (Qiagen) in a final volume of 50 µl, containing 100 ng of genomic DNA on a Perkin-Elmer Thermocycler (Perkin-Elmer Applied Biosystems, Foster City, Calif. USA). Thermal cycling conditions consisted of an initial denaturation step at 94°C for 3 min, followed by 35 cycles of denaturation at 94°C for 30 s, primer annealing at 42°C for 1 min and primer extension at 72°C for 1 min, with a final 10-min extension at 72°C.

After agarose gel fractionation, the amplified DNA fragments of expected sizes (assuming no introns) were excised and purified using a MinElute gel extraction kit (Qiagen) and cloned into the pGEM-T easy vector (Promega, Madison, Wis., USA). Selection and manipulation of recombinant plasmid clones followed standard methods (Sambrook et al. 1989). To discriminate between different sequences derived from PCR, multiple clones were analyzed by restriction enzyme digestion with *Taq* I or *Rsa* I, and grouped according to their digestion patterns. For each group, at least two clones were picked for DNA sequence analysis. In most cases, the two clones from a given group had the same sequence. If not, more clones from these groups were analyzed to confirm DNA sequences and identify possible errors that had resulted from misincorporation during PCR.

DNA sequence analysis

DNA sequences of RGA clones were determined on both strands using an ABI310 DNA sequencer (Applied Biosystems), a Thermocycle sequence kit (Amersham), and T7, SP6 and other internal

primers as needed, according to the manufacturer's instructions. DNA sequence data were compiled and analyzed using BLAST (Altschul et al. 1997) and ClustalW network services at the National Center for Biotechnology Information (NCBI). A phylogenetic tree based on the NBS sequence alignment was constructed by the Fitch-Margoliash and least-squares methods with evolutionary clock (Kuhner and Felsenstein 1994) using the BioEdit package (Hall 1999). The reliability of the tree was established by conducting 1000 neighbor-joining bootstrap sampling steps. The RGA sequences from western white pine have been deposited in the GenBank database under the Accession Nos. AY294058–AY294124.

cDNA preparation and amplification

RNA was isolated from pooled foliage of multiple rust-resistant trees using the procedure described previously (Liu and Ekramoddoullah 2003). Total RNA was treated with RQ 1 RNase-Free DNase (Promega) and purified again using a Plant RNeasy extraction kit (Qiagen). cDNA was reverse-transcribed using a SMART cDNA library construction kit (Clontech Laboratories, Palo Alto, Calif., USA) following the manufacturer's instructions. PCR was carried out with 1 µl of cDNA in 25 µl of PCR mix using Taq PCR Master Mixture (Qiagen). The procedure for RT-PCR cloning of RGAs was the same as that used for RGA cloning from genomic DNA, except that cDNA was used as the template for PCR.

To examine the expression of RGAs cloned from genomic DNA, RT-PCR was performed with 18 pairs of RGA-specific primers based on the analysis of RGA nucleotide sequence alignments (Table 2). RT-PCR was carried out with 1 µl of cDNA in 25 µl of PCR mix using Taq PCR Master Mixture (Qiagen). Each RT-PCR was performed with negative controls (DNA and cDNA-free) and pooled genomic DNA from 10 resistant trees. The PCR products from templates of cDNA and genomic DNA were compared for the size and intensity. RGA expression levels were compared taking the genomic PCR product as reference. RT-PCR conditions were as described above for the NBS motif primers, except that the annealing temperature was 48°C and PCR was performed for 30 cycles.

Table 2 RGA-specific primer pairs used for RT-PCR and genetic linkage analysis

RGA clone	Forward primer (5' to 3')	Reverse primer (5' to 3')
G9034	TTGTGTGCATCTGATCAG	CTTATAAGATATATGTAGTGC
G9036	ATGCCACAAAGCATGTC	CTCTCAGTCTCATCAATG
G8029	TGGATGCATTTTTGTCAATG	TCTCAGTAGGAAGGTTTAC
G8106	CAGTAAATGCACTTCTCC	CATCACAGGCTTTTAAATC
G8501	AGAGATAATGCGGACAAG	CGAACTTCTGTACCAAGG
G8540	TGAGGTAAGAGACAAACC	CTACTTCTCAAATCCTCC
G6207	GCGGATCAGGTTTATGAA	ATGACCGAATCCACGTC
G6209	CGATTTTAGTTACCTCTTG	GAAGACACTGTTTTGTGTC
G6215	TCAGCTTTTGCCTGATCA	TTTTCCATCATTCACTCT
G6227	TTTGCCAGTTCTAAGAGGA	TGTACATTTCTATTTCATCCG
G6237	GTACATGCACTTTTGCTG	GCTCTCATGACTCTCTCAC
G6243	TAGATGCACTTCTCAGGC	TCAGTCCCATCATAAGGG
G6244	ATGAACTTTTGTAGGTCG	CTGTGTTATAACTATTCACC
G6253	ATGATGTCCTTCCTTTCA	ACATGTTTTGTAGATTCCTCA
G6262	CTTTTTCTCATGGATCTCC	TTCCATTGACTTACCATAC
GM15	GAAGGATAGGTAGATAATCT	GATCGTGCACGCGTATAT
G5004	AGGACATGAATGGTGTAG	TTCTGGAGGTGGATATAG
GI5033	AGCAGTTAGAAGCTCTCG	CTCTGGATCTGAACAGCT

Results

Analysis of WWP RGAs by PCR cloning

PCR using primer combinations corresponding to P-loop/DMG motifs yielded two major DNA bands of approximately 850 bp and 900 bp. PCRs using P-loop/GLPL, Kinase-2/DMG, or Kinase-2/GLPL primer pairs each produced only one DNA band of roughly 500 bp, 620 bp or 270 bp, respectively.

Sequence analysis revealed that only a small proportion of the clones obtained from the first-round PCR fragments displayed homology to R genes or RGAs of angiosperms. Only one out of 48 clones was identified as an RGA from the first-round of PCR with the primer combination Gkt2/Pal2. The nested primer strategy increased the specificity of the amplification of WWP RGAs. The second round of PCR with nested primers produced sharper DNA bands of the expected size assuming no introns. All sequenced clones from the nested PCR, including those performed with the primer sets Gkt1/Pal1, Dvd1/Dmg1 and Dvd1/Pal1, were RGAs, indicating that the nested PCR strategy was effective in increasing the efficiency of RGA cloning. In total 455 recombinant plasmids were identified with DNA inserts of expected sizes. Restriction enzyme analysis enabled the recombinant plasmids to be grouped into several classes. All unique groups were given a name based upon their cloning template and fragment length. At least two clones from each group were sequenced. BLAST searches led to the identification of 67 WWP RGAs, 21 from cDNA and 46 from genomic DNA (Table 3).

Analysis of RGA sequences revealed the presence of stop codon and/or frame shift mutations in 18 of 46 RGA genomic clones and in 14 of 21 RGA cDNA clones, while others contained an uninterrupted ORF (Fig. 1 and Table 3). Alignment of RGA nucleotide sequences showed that mutations were localized at the same position in several clones which were otherwise identical. The consistency of these RGA clones containing nonsense mutations therefore may not result from mistakes in PCR amplification or random errors in DNA sequence analysis, but may in fact indicate the presence of RGA pseudogene families in the WWP genome.

Comparison of NBS domain RGAs from WWP with angiosperm R genes

BLASTP searches of the GeneBank database demonstrated that the WWP RGAs shared significant homologies with well-characterized R genes from angiosperms, as indicated by the E values ranging from 4e-6 to 9e-44. Comparison of the deduced amino acid sequences of WWP RGAs with those predicted from R genes, such as tobacco *N* (A54810), flax *M* and *L5* (AAB47618 and

Table 3 Characteristics of the RGA sequences from western white pine

Cluster	Clone name ^a	Primers	ORF ^b	RT-PCR ^c	
1	C608 (13)	Dvd1/Dmg1	*/▲	+	
	C618 (3)	Dvd1/Dmg1	+	+	
	GM01, G6202, G6229, G6236	Dvd1/Dmg1	+	nt	
	GS02, GS22	Dvd1/Dmg1	*	nt	
	G6204	Dvd1/Dmg1	▲	nt	
	G6209, G6253	Dvd1/Dmg1	+	-	
	G6264	Dvd1/Dmg1	**/*	nt	
	G8540	Gkt1/Pal1	▲	+	
	2	C601 (2), G9036	Dvd1/Dmg1	+	+
		3 G6237	Dvd1/Dmg1	*	-
G9034 (2)		Gkt1/Dmg1	+	-	
3	C62	Dvd1/Dmg1	*	+	
	4 GM05	Dvd1/Dmg1	*	nt	
	G6262	Dvd1/Dmg1	**/*	-	
4	C202	Dvd1/Pal1	+	+	
	GM09, G6205	Dvd1/Dmg1	+	nt	
	5 GM10	Dvd1/Dmg1	*	nt	
	GM12	Dvd1/Dmg1	**/*	nt	
	G8106 (2)	Gkt1/Dmg1	+	+	
	G8503 (2)	Gkt1/Pal1	+	nt	
5	G6243	Dvd1/Dmg1	+	-	
	C201	Dvd1/Pal1	*	+	
6	GM201	Dvd1/Pal1	+	nt	
	G8029	Gkt1/Dmg1	+	+	
7	G9044	Gkt1/Dmg1	+	nt	
	GM06	Dvd1/Dmg1	+	nt	
8	G6244	Dvd1/Dmg1	+	+	
	G6269	Dvd1/Dmg1	*	nt	
	C204	Dvd1/Pal1	▲	+	
	G6207, G6261	Dvd1/Dmg1	+	nt	
9	G6215	Dvd1/Dmg1	▲	-	
	G8501	Gkt1/Pal1	+	+	
	G8502, G8506	Gkt1/Pal1	▲	nt	
	G8508	Gkt1/Pal1	+	-	
	G8533	Gkt1/Pal1	▲/▲	nt	
	GM15	Dvd1/Dmg1	*/▲	-	
10	G5004	Gkt1/Pal1	*	+	
	G5028	Gkt1/Pal1	*	nt	
11	G6227	Dvd1/Dmg1	+	-	
	G15033	Gkt2/Pal2	+	-	

^aClones beginning with "G" were from genomic DNA, and those beginning with "C" were from cDNA. The numbers in parentheses indicate the numbers of RGA clones in the cluster that showed 99% nucleotide sequence identity

^bContinuous ORF (+), stop codon generated by substitution (*), frameshift generated by insertion or deletion (▲), double variations from stop codon and/or frame-shift (*/▲, */*, ▲/▲)

^cPresence of expected fragments assuming no intron (+), absence of expected fragment (-), or not tested (nt)

AF093645), and soybean *LM6* (AAG09951), showed identities in the range of 24% to 36%. BLASTP searches also detected one homologous sequence from *Pinus radiata* (AF038649) and four others from *P. taeda* in GenBank (AY091558, AY091560, AY091562, and AY091566) (Meyers et al. 2002).

To establish the origin and elucidate the evolution of WWP RGAs, angiosperm R genes (*AtRPS2*, *AtRPS5*, *BoRPS2*, *Lu-M*, *Lu-L*, *Nt-N*) and other homologous conifer RGAs were included in this phylogenetic analysis. One phylogenetic tree was constructed from aligned sequences and contained two

1
G9034 GVGKTTLAK ELFNRSK-SR YSKSCFLID- --VRE-KA-K SSIPSLQRKL LKDLGASDQE --PIDSKDEG IEILRRHFS- ----- --VL SAPV LVLDDVDHQ
C601 I. N. R. EK. I. S. V.
G9036 HF. M.K-Y P. P. --IKD-G-- -LT. ARI.G- V. KAM.IE.LKK VNAPDTGKEM PTKH.E.SKF I. E.
G8029D.GG...F--A.AN C.LN...S. .G.TQM.R- -Q...IA. .M.TNLK- -FHA .V. N. V
G8106VQ-N.HR...S--N.G. I.LH...K. .NS.TQL.E- -Q.ENVS. .L.KT.L- -COA .VI. N. V
G6237 VI.
G6262 VI. F
G6243 VI. PV
G6202 VI. I
G6209 VI. I
C608 VI. I
G6204 VI. I
G6229 VI. I
G6236 VI. I
C618 VI. I
G6264 VI. I
G6253 VI. I
GM15 VI. YE
G6227 VI. QT
G6244 VI. KV
GM06 VI. KV
G6269 VI. KV
Lu-M G.I...T... AVY.KI.-.H FDRC.VDN- --.AMQEQ. DG.FI...K. VSEILRM.S- -VGFTNDSG. RKMIKERV.- -KSKI .V. EK
Lu-L5 G.I...T... AVY.KI.-.C FDRC.IDN- --I.TQDQ. DGVVV...K. VYELRI.SG SVGFNNDSG. RKMIKERV.- -RFKI .V. EK
Nt-N G.....I.R AI.DTL-L-G. MDS.YQFDGA CFLKDI.EN. RGMH...NA. .SE.LREKA- --NYNNEED. KHQMASRLR- -KKI .I. I.NK
G8501 G.....I.R AI.DTL-L-G. MDS.YQFDGA CFLKDI.EN. RGMH...NA. .SE.LREKA- --NYNNEED. KHQMASRLR- -TPN .I. KA
G5004 G.....M... F...NK-Y .KTRS.FN- --GGNA.T. N-LN...SH. I.MMGV.V- -Q.CNP...S. .KYY. NYHA .VI. NV
G85040 G.....I.R AI.DTL-L-G. MDS.YQFDGA CFLKDI.EN. RGMH...NA. .SE.LREKA- --NYNNEED. KHQMASRLR- -CKA .I. I
GI5033 G.....I.R AI.DTL-L-G. MDS.YQFDGA CFLKDI.EN. RGMH...NA. .SE.LREKA- --NYNNEED. KHQMASRLR- -GKKI .L. IQSL
GM201 VI. L
Cons. GVGKTTLak efnrk.S. ...&fL.d. .vrdna.k .slnsLok.\$ Lksl1.d. !.s.deG i.lkKh.s. S... LvlDDVDH.
P-LOOP RNBS-A-TIR Kin-2

111
G9034 DQVYALLPVT DKGILTLPSS SLILITSRDT NVLTRSGVQ- EPSIYKLTGL SREHSRELFCS SHAF-CQPHP LSGFEHLVDQ FLEACSGLPL SLKVFALLC GKTNKSYWKE
C601 ..D.F... ..H. F. K. I. S. F. Y. R
G9036 M E. K. E. R. P. V. G. RN.D. K
G8029 ..LD.F.S-M KD---V.RSDV...K H...S.IG- .S...R...N.Q...Q...L.S...H.Y. P...Y.E .T.G... ..RE.D.C.EE
G8106 N.LN...RPI RS---V.H.HK D...R.E- .S...R...NTHQ.Q...L.P.G... .P...E.EN D.K.D... ..Y. NHD.F.EE
G6237 ..H...L...D... ..C... ..K... ..R... ..-W... ..G... ..N.H... ..
G6262FLMD ---V.RHDLC.Q .I.K...E- .S...R...S.NT.C... ..-R.MQ...N. .NS*N... ..G.H.N... ..E
G6243 N.LD...RPI RD---VILSN G.....NR D.K.R.E- .S.F.T... .PQQ.QQ...L...T... FL... ..K .K.DR... ..Y. NDD.F.EH
G6202 Q.MN...SP. NG---V.SGV...NK D.I.W.IV- .S...K... DMQQ.K...R.T...H.S... DV...EV.L .KT.D...L. .M.L.HMH .RHLN.EE
G6209 Q.MN...SP. NG---V.SGV...CNK D.I.W.IV- .S...K... DLQQ.K...C...H.S.L DV...EV.L .KTYD... ..MVI.HMH .QKALK.EA
C608 Q.MN...SP. NG---V.SGF.V...HNK D.I.W.IV- .S...K... DLQ*LK...C...H.S... DV...EVA.L .KT.D.FS. .MVLV.HMH .QKALK.EA
G6204 Q.MN...P. NV---V.SGV...NK D.I.W.IV- .S...A... DPO.K...C...H.S... DV...EV.L .T.D... ..T.L.HMH .KILK.EA
G6229 Q.MN...LPA KD---V.GV...NK D.I.W.IV- .S...D... DPO.K...C...H.S... DV...EV.L .T.D... ..T.L.HMH .QKHLN.EA
G6236 Q.MN...SP. ND---V.SGV...HNK D.I.W.IV- .S...D... DPOQ.K...W...H.S... HV...EV.L .KT.D... ..M.L.HMH .QKHLK.EA
C618 Q.MN...SP. ND---V.SGV...NK D.I.W.IV- .S...D... DPOQ.K...W...H.SR. DE...KV.NL V.T.G... ..M.L.HMH .QIHLK.EA
G6264 Q.MN...SP. ND---V.SGV...HNK D.VI.W.IV- .S...D... DLQQ.K.I...*...H.S... HV...E.NL .KT.G... ..M.L.VMH .EK.LK.EA
G6253 Q.MNP...SP. ND---V.F.V...CNK D.I.W.II- .S...D... DPOQ.K...W.T...H.S... DV...KVFNL L.KT.G...L. .M.L.HMH .EKYLK.EA
GM15 G*.DN...NRV...Q.R D.RK...E- .I...I...M PLK...D...Y... .AY. K...R.NE .L.Q... ..F...E.F.ND.TLVE
G6227 S.MD.F...L RG---G.NV...NK D.G.IVP DA...E... NAPPY.QQ...F...GHS... VP...P...ER .V.T.QA.S .LS.E.V.DP...E
G6244 ..DE.LGR I---V.HSN C...Y...RK D.K.L.E- .S...HN.D. .S...L...Y LYS.-N.AD. .PE...S.AIK .IKSS... ..IN .E...A.EA
GM06 .K.DE.LGR I---V.HSN C...Y...RK D.K.L.E- .S...HN.N. .S...L...Y LYS.-N.AD. .PE...S.AIK .IKSSG... ..N .E...A.EA
G6269 ..DEI.LGR I---V.HSD C...Y...RK D.K.L.E- .S...HN.ND. .S...L...Y LYS.-N.AD. .PE...Y.AIK .IKSSG... ..IN .E...A.EA
Lu-M PKFEDI.GCP KD---FDSG TRPI...NQS.LNEN- QCKL.EVGSM .EQ.L...S K...KKN.T. P.DY.T.AND IVSTTG... T...T.SF.F- .RQIAGV.EE
Lu-L5 PKFEDM.GSP KD---FISQ .RPI...SM R.G.TLNEN- QCKL.EVGSM .KPR.L...S K...KKN.T. P.DY.T.AND VVDTTA... T...I.SI.F- .QIAGV.EE
Nt-N .HYLEY.AGD LD---WFGNG .R.I.T...K HLI---EK NDI.EV.A. PDHE.IQ.K Q...GKEV. NEN.K.SLE VVNAYK... A...W.S.H -NRLRTE.SS
G8501 ...E...PD LT---L.H.D T.V...YR D.I.S...E- .S...M... TTQ...H...L...-NR.A APA...S.QK .V...G... ..
G5004 .ELDVF...PI KG---V.H.NV...R G.KSA.IV- .L...H.R...*...-S.LY. PPE.RN.ADR .IDE.Y... ..
G85040 Q.MD...LPV KD---V.Q.GV...NK D.KSA.IV- KS...T.V. D.Q.K...K...-Q.S.Y DG...EV.EG .N...D... ..
GI5033 E.LE.GGNY SD---LGEG .CL...NQ HIRKLAN.D- --VHEVR. PH.YAMQ.K F...PSSQS DPQLQT.SND IVS...G... ..
GM201 ..LD.F...IK LD---V.H.KV...K GL.I.ARIA- .S...H.S... N.RY.QQ... ..S.LD. PLEYQC...G .VK...D... ..
Cons. ##.all.p.vl... slILITsr.k dvL.rsg!.. #sI8.ltgL .q.s.#Lfc.HaF .q.p .g%#lv.. fl.ac.GLPL SLkv.g... g...yw..
RNBS-B/Kin3 RNBS-C GLPL

221
G9034 ELKELRKT-L HEDIQKSLQV SYDALRREEQ QIFLDIACFF IGESRDTAIR VWDASGWNCS RVFQSLSKC LVEMHI-GGE SP--HIYILI RMHDHLRDMG RDL
C601 Q.K...K... ..K... ..EE.V.D.L F... ..D.-ET ESRS.N.V.
G9036 QV.K...K... ..K... ..EE.KN... ..D.L L... ..D.-DET ESRS.N.V.
G8029 Q.GR.QQVN. PTE..RT.KING.K E... ..NK.M.L I.NG...K. LGLRN.KD.VDS-KNRL
G8106 Q.DR.EQI-. PDE..R.I .Y..QID.KV... ..K... ..I.GS.K.A LG...QN.SVD.-ENQH
G6237K.K.EK...*... ..L... ..P... ..D.SESH ESPRR... ..
G6262 ..ER.EEI-. PDE..R.*IHID.K H... ..R... ..YGL.K... .G...H... ..VQYMY.K SME---HC. T... ..
G6243 Q.ERVEQI-. P...KTR... ..L.QDD.KRKY... ..K.ER...K.W LG.RR.QD.SVDSGVDP YDGTG--IR.I
G6202 K.RKISNV-. PT..CR.KIS.DQQ.K N.L.TT... R.KD...M. I.GFD.E.E LS.RN.QN...L.VNDKNE-V
G6209 .HKISNV-F PM..RCR.KIS.DQQ.K N.L.T... REKD...V. I.GFD.E.E LS.RNRQN...L.VNDKNE-V
C608 .RKISNV-. PI.M.CR.NIS.DQQ.K N.L.T... R.KD...T. I.VFDSE.E LS.N.QNR...L.VNDKNE-V
G6204 .RKISNV-. PT..RCR.KIS.DQQ.N N.L.T.Y... R.KG...V. I.G.D.E.E LS.N.QNR...L.VNYKNE-I.N
G6229 .HKISNV-. PT..RCR.KIS.DQQ.K N.L.T... R.KD...T. I.G.D.E.E LS.RN.QNR...L.VNDKNE-G
G6236 .RKISNV-. PT..CR.KIS.DQQ.K NM...TP... R.KD...T. I.G.D.E.E LS.RN.QNR...L.VNDKNE-T.N
C618 .RKPSNV-. T..RCR.KIS.DQQ.K N.L.T... R.KD...T. I.G.D.E.E LN.RN.QN...L.VKDKNE-V GI
G6264 Q.RKISNV-. PT..RCR.KIS.DK* K N.L.T... R.KD...M. I.GYD.E.E LS.RN.QN...L.VNDENE-
G6253 ELQITISNI-. PT.M.SRMKII.DQQ.K N.L.T... K.KY...T.I V.R.D.E.E LS.WN.QNM...L.VNDKNE-M
GM15 AHLHFQVHEV PDV.K.T.RIHEKV... ..K...ITVE I.EP.RK...LS.N.E... ..VDSDNV-RV
G6227 Q.DI.PQN- VRDMRDR.ISL.D.K DV...T.S... K.RAEMS...L.LRLK...LLL.A.E.RSLDSNGR-E.YNSV...
G6244 HIDKIQHI-. P.E.K.R.I .I.C...DM.DR H...YL.LD.RV... IC.G.A.IL LG.RN.D.RVNSYNT-V I...Q...
GM06 HIDKIQHI-. P.E.K.R.I .I.C...DM.DR H...SL.LD.RV... IC.G.A.IL LG.RN.D.R V...VNSYNT-I...Q... ..H
G6269 RMDK.QHI-. P.E.K.R.I .I.C...DM.DR H...S*... ..D.RV.K IC.G.A.IL LG.RN.YN... ..VNSYNT-I...Q... ..
Lu-M T.EQ...LD LDEYDR.KIKA.AK E... ..RNKEMPYV M.SECKFPYK SNII.F.IQR. MIQVGGDD.V-L.E.Q... ..EI
Lu-L5 T.QQ...LN LDEYDR.KIKP.AK E... ..KNKEVPYV M.TDCNFYPA SNII.F.IQR. MIQVGGDDK-F.K...Q... ..EI
Nt-N AIEHMKNNS YSG.IDK.KIGLEPKQ. EM...L R...EK.YILQ ILESCHIGAE YGLRI.ID.SFISEYNQ-Q...LIQ... KYI
G8501
G5004
G85040
GI5033
GM201
Cons. .l.....l p#.i.rLi SydaL...ek .iFLDIacFF iGe.rDt.ir !wd.s.w.g. l.f.nLq.kc Lvev...n.i rMhdhlrDMG Rdl
RNBS-D-TIR

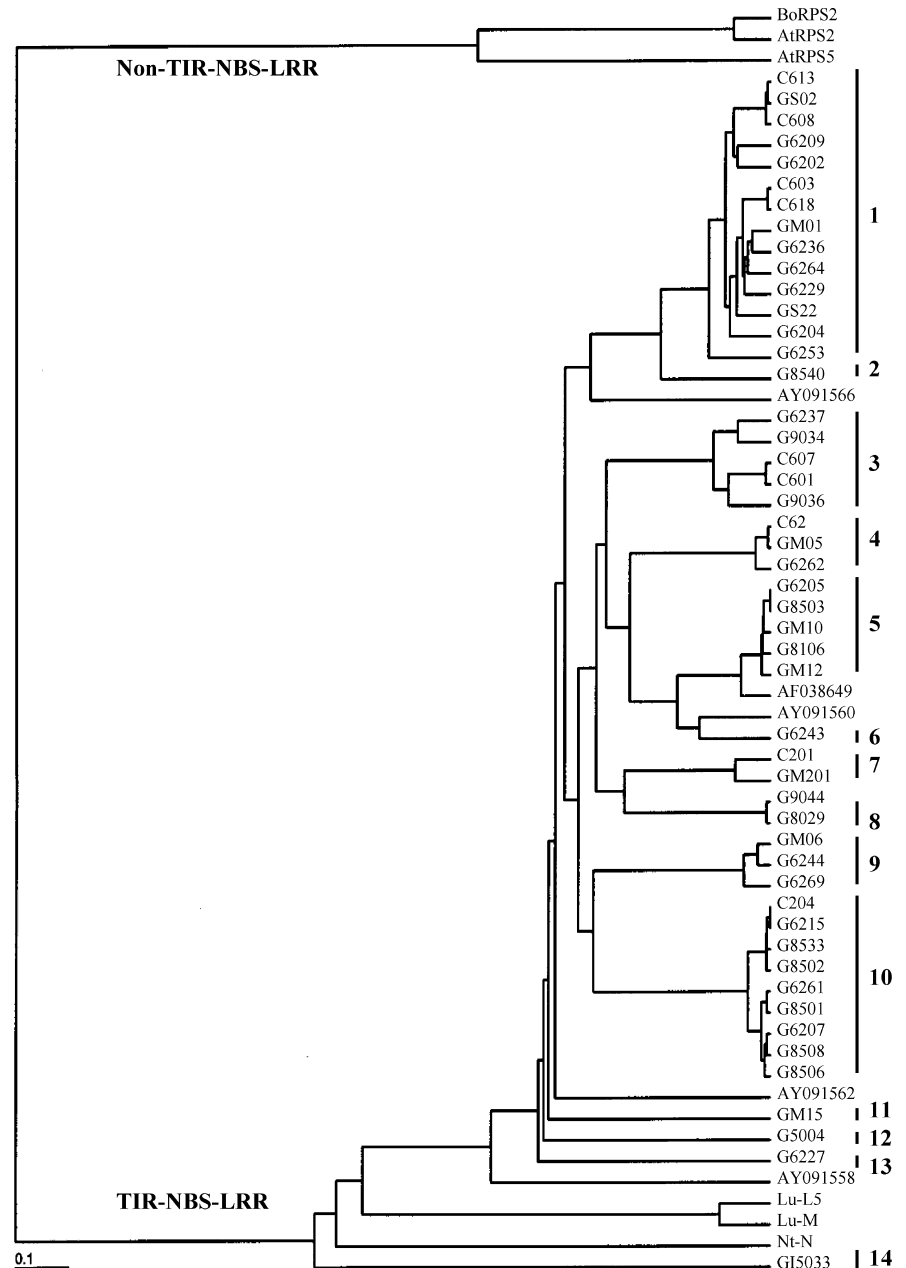


Fig. 1 Alignment of the predicted amino acid sequences of NBS sequences from western white pine. Tobacco *N* (A54810) and flax *L5* (AF093645) and *M* (AAB47618) are included in this analysis. Only 28 representative WWP RGAs that show $\leq 90\%$ amino acid sequence identity are presented. Amino acid sequence identities are indicated by the *dots*. Premature stop codons and frame-shifts are indicated by *asterisks* and *black triangles*, respectively. The conserved motifs defined by Meyers et al. (1999) are *highlighted*

major branches (Fig. 2). As expected, *AtRPS2*, *AtRPS5* and *BoRPS2* were grouped together in one branch representing the non-TIR subfamily. All WWP RGAs identified in the current study were present in one branch containing the *Nt-N*, *Lu-L5*, and *Lu-M* genes, suggesting that these WWP RGAs may belong to the

TIR-NBS-LRR subfamily. This classification is further supported by the presence of internal conserved motifs in the NBS domain. Alignments of deduced amino acid sequences demonstrated that the conifer RGAs also contained the conserved motifs in NBS domains that were previously identified in R genes and RGAs of angiosperms (Meyers et al. 1999, Pan et al. 2000a, Peñuela et al. 2002), including the P-loop (Kinase-1), RNBS-A-TIR, Kinase-2, RNBS-b/Kinase-3, RNBS-C, GLPL, and RNBS-D-TIR signatures (Fig. 1). Among these conserved motifs, the P-loop and GLPL motifs are conserved in both TIR- and non-TIR-NBS-LRR proteins, while RNBS-A-TIR (LQKKLLSKLL) and RNBS-D-TIR (FLHIACFF) have been found only in the TIR-NBS-LRR proteins (Meyers et al. 1999). The

Fig. 2 Phylogenetic tree based on the amino acid sequences of RGAs from western white pine and selected resistance proteins from angiosperms. RGA sequences from western white pine were compared to additional NBS-LRR type sequences with known disease resistance functions, including tobacco *N* (*Nt-N*), flax *L5* (*Lu-L5*) and *M* (*Lu-M*), Arabidopsis *AtRPS2* (AF487823) and *AtRPS5* (AF074916), and Brassica *BoRPS2* (AAF19803) proteins. Only 48 representative WWP RGAs that show $\leq 98\%$ amino acid sequence identity are presented. The TIR-NBS-LRR branch is grouped into 14 classes of RGAs. Members of each class display $\leq 75\%$ amino acid sequence identity with members of all other classes. The scale at the *bottom* indicates genetic distance proportional to the number of substitutions per site



discovery of a large number of RGAs in the WWP genome and their significant homologies with angiosperm TIR-NBS-LRR family provides evidence for the hypothesis that the TIR-NBS-LRR subfamily originated before the divergence of angiosperms and gymnosperms (Meyers et al. 1999, Pan et al. 2000a).

Variation and evolutionary analysis of WWP RGA sequences

WWP RGAs are highly divergent, sharing 4–99% identity at the nucleotide level and 18–99% identity at the amino acid level in the NBS domain, indicating that, like angiosperms, conifers contain a large and diverse family of genes for proteins with the NBS domain. WWP RGAs appear as short-separated branches that form different clusters defined by levels of amino acid sequence identity ranging from 77% to 99% within clusters (Fig. 2). Fourteen clusters are defined at a sequence identity threshold of 75% (Fig. 2 and Table 3), with identity scores for pairwise comparisons of members from different classes ranging from 18% to 74%. Only one WWP RGA was identified in each of clusters 2, 6, 11, 13 and 14 (Table 3). RGA clone GI5033 was positioned at the base of the TIR-branch. It has higher identity to angiosperm R genes than to conifer RGAs. Its identity to TIR-type R or RGA sequences from both conifers and angiosperms ranges from 18% to 36%. This range is comparable to its identity (24% to 28%) to members of the non-TIR subfamily. This suggests that WWP RGAs might have arisen from a common ancestor of both TIR- and non-TIR gene families.

Thirty RGA cDNA clones obtained with primer sets based on Kin2 and MEG motifs were randomly picked for sequence analysis, yielding 18 different RGA sequences. They were divided into three types representing clones C601, C608 and C618, respectively (Table 3). RGAs of types C601 and C608 contained a continuous reading frame. However, RGA type C608, defined by 13 clones, had both premature stop codon mutations and frame-shift mutations. The reason for such a high rate of pseudogene recovery from a transcribed RGA type is not clear. C608-type RGAs were 99% homologous, which may in part reflect single nucleotide polymorphisms of allelic or paralogous pseudogenes in the WWP population, because the cDNA template was pooled from the foliage of multiple trees. However, several RGA genomic clones from a single seedling also showed nucleotide differences, suggesting that RGA gene duplication might have occurred in the WWP genome. Apart from single base differences, other types of mutations were also identified within the NBS domains. In 12 RGA clones of the C608 type a single nucleotide insertion resulted in a shift in the reading frame. A multiple base deletion at the same position was found in RGA clones C24, G6215, G8502 and G8533, which were grouped together in class 10 (Fig. 2).

To examine the evolutionary forces acting on NBS domains of WWP RGAs, the ratio of non-synonymous (Ka) to synonymous nucleotide substitutions (Ks) was calculated for members of five RGA classes (Table 4). The majority of nucleotide variations between RGA members represented non-synonymous substitutions. Ka/Ks values for pairwise comparisons ranged from 0.4 to ∞ . One extreme example was the G6244/GM06 comparison, where every nucleotide change was non-synonymous (Table 4). Ka/Ks ratios in many RGA comparisons were significantly larger than 1, while others showed no significant difference from the value expected for neutral selection (Ka/Ks=1) (Table 4). Pairwise comparisons involving the RGA pseudogene

Table 4 Ka/Ks ratios for pairwise comparisons of the NBS domains of RGAs from western white pine

RGA pair ^a	Degree of nucleotide sequence identity	Degree of amino acid sequence identity	Ka/Ks ^b	X ^{2c}	Probability
G6202/G6209	0.93	0.88	1.50	1.27	0.254
G6202/G6229	0.92	0.87	1.50	1.48	0.221
G6202/G6236	0.92	0.89	1.20	0.21	0.647
G6202/G6253	0.89	0.81	2.31	9.63	0.001*
G6202/GM01	0.92	0.87	1.61	2.03	0.157
G6202/C618	0.90	0.85	1.67	3.16	0.073
G6209/G6229	0.92	0.86	1.75	2.85	0.088
G6209/G6236	0.92	0.86	1.93	3.98	0.045*
G6209/G6253	0.89	0.78	2.42	10.98	0.001*
G6209/GM01	0.92	0.85	1.93	3.98	0.045*
G6209/C618	0.91	0.83	2.25	6.98	0.008*
G6229/G6236	0.95	0.91	1.80	1.79	0.179
G6229/G6253	0.90	0.80	3.38	16.55	0.000*
G6229/GM01	0.94	0.91	1.36	0.49	0.488
G6229/C618	0.94	0.89	1.85	2.79	0.094
G6236/G6253	0.91	0.83	2.53	9.54	0.002*
G6236/GM01	0.95	0.93	1.08	0.00	1.000
G6236/C618	0.93	0.88	1.67	2.09	0.147
G6253/GM01	0.92	0.84	3.18	11.94	0.001*
G6253/C618	0.90	0.81	3.30	15.73	0.000*
GM01/C618	0.95	0.91	1.64	1.27	0.254
C608/C618	0.90	0.82	2.00	6.03	0.014*
C608/G6209	0.92	0.86	2.21	6.03	0.014*
C608/G6229	0.91	0.84	1.55	1.02	0.317
C608/G6236	0.92	0.88	1.31	0.59	0.442
C608/G6253	0.89	0.79	2.44	11.29	0.001*
C608/GM01	0.92	0.85	1.57	2.12	0.147
C601/G9034	0.92	0.84	1.86	4.99	0.025*
C601/G9036	0.86	0.79	1.96	7.82	0.005*
G9034/G9036	0.86	0.77	2.56	15.67	0.000*
G8106/G8503	0.98	0.98	0.75	0.00	1.000
G8106/GM09	0.98	0.99	0.40	0.58	0.446
G8106/G6205	0.98	0.96	1.75	0.37	0.543
GM09/G6205	0.98	0.97	1.00	0.10	0.920
G6244/GM06	0.98	0.95	∞	7.16	0.007*
G8501/G8508	0.97	0.95	2.00	0.76	0.383
G6207/G6261	0.96	0.94	1.44	0.42	0.517

*Significantly different ($P < 0.05$) from a Ka/Ks ratio of 1

^aThe Ka/Ks ratios were calculated for RGA pairs with a maximum nucleotide sequence identity of 98% between members of clusters 1, 3, 5, 7 and 10

^bRatio of non-synonymous to synonymous substitutions

^cChi-square (X^2) values were calculated with Yates's adjustment using a 2X2 contingency table

C608 did not change the pattern of Ka/Ks ratios. The divergent or neutral selection acting on the NBS domains of WWP RGAs shown by this analysis contrasts with the NBS domain of angiosperm RGAs, which are considered to be under conservative selection pressure (Michelmore and Meyers 1998, Graham et al. 2002).

Expression analysis of WWP RGAs

The expression profiles of selected RGAs are summarized in Table 3. RT-PCR showed that seven of 18 RGAs were expressed (Fig. 3). For the other 11 RGAs, the corresponding mRNAs were not detectable in foliage tissues. The control PCR produced the expected DNA fragments from the genomic DNA template using all 18 RGA-specific primer pairs. The different staining densities of RT-PCR products on agarose gels demonstrated that the expression levels of RGAs were different, because the same amount of cDNA template was added to each RT-PCR tube. As shown in Fig. 3, G9036 was expressed at the lowest level, G8106 and G6244 were expressed at the highest levels, and other three RGAs (G8029, G5004, and G8540) were expressed at intermediate levels.

Discussion

In the first study aimed at elucidating the diversity and evolution of RGAs in a conifer genome, we report here the characterization of NBS sequences from western white pine. In conifers, RGAs encode TIR-NBS-LRR proteins with the structural features that are also characteristic of angiosperm R proteins. This is consistent with previous studies (Cannon et al. 2002, Meyers et al. 2002), in which RGA searches in GenBank revealed the presence of NBS-LRR sequences in other gymnosperm species. The identification of an NBS-LRR superfamily

in both conifers and angiosperms suggests that its common ancestor was present prior to the split between these seed plants about 275 to 300 million years ago (Cannon et al. 2002, Meyers et al. 2002).

Our results indicate that the PCR approach is also useful for systematically searching for RGAs in a conifer genome. Despite the fact that all of the cloned RGAs belong to the TIR-NBS-LRR subfamily, there is great diversity among them at both the nucleotide and amino acid sequence levels (Fig. 1 and Fig. 2). PCR with more primers based on different conserved motifs in the NBS domain should allow us to recover all NBS sequences from the WWP genome. A classical genetic study has demonstrated that at least one R-locus (*Cr2*) is present in the WWP genome (Kinloch et al. 1999). Although many RGAs might have no functional activity due to disruption of the reading frame, the WWP RGAs cloned in this study resemble angiosperm R genes and encode proteins with the expected structure (Fig. 1 and Table 3). Therefore, it is reasonable to speculate that some WWP RGAs described here may act as functional R genes against pathogens.

The current study finds that almost half of the cloned RGAs are pseudogenes, and some of these are even transcribed into mRNA (Table 3). Sequence comparisons suggested that the WWP RGA pseudogenes originated from point mutations, insertions, or deletion of nucleotides. In a similar study in tomato, only 10% of cloned RGAs turned out to be pseudogenes (Pan et al. 2000b)—a much lower fraction than reported even for other angiosperm species. This may reflect a difference in RGA organization between genomes of angiosperms and conifers. Pseudogenes are nearly as abundant as functional genes in the human genome (Harrison et al. 2002). They have also been identified in the paralogs of the R genes *Xa21*, *Cf9*, *Pto* and *Dm3*. Probably serving as reservoirs of potential variation, pseudogenes are believed to be advantageous because they may allow for recombination and gene conversion between alleles or paralogs of functional R genes (Michelmore and Meyers 1998). A recent report has demonstrated that a mouse pseudogene is expressed and regulates the stability of the mRNA expressed from the corresponding protein-coding gene (Hirotsune et al. 2003). Many WWP RGA pseudogenes were found to be expressed in foliage tissues; therefore, investigation of their biological roles may help elucidate the mechanisms of R gene regulation.

In many angiosperm species, RGAs are found clustered around known R loci. Within a cluster of RGAs, there may be more than one gene conferring resistance to different isolates of a specific pathogen or to biologically diverse pathogen taxa (Cooley et al. 2000, Van der Vossen et al. 2000). Some RGAs are genetically linked to R loci in *Arabidopsis* (Aarts et al. 1998). Therefore, RGA cloning and identification in the WWP genome offers an excellent starting point for a map-based approach to the cloning of target RGA genes, as well as other members in these clusters, including potential R genes. Considering their high level of polymorphism and diversity, and the

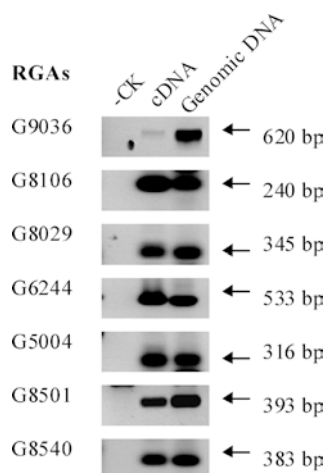


Fig. 3 RT-PCR analysis to determine the expression profiles of WWP RGAs. Total RNA was purified from foliage tissues and reverse-transcribed. Only positive results obtained with the indicated gene-specific primer pairs are shown

putatively functional character of their transcription activity, these WWP RGAs will provide a source for the development of molecular markers, especially for the selection of resistance to white pine blister rust (*C. ribicola*) in forest breeding. For a more definitive elucidation of the proximity of RGA sequences to known R genes in the WWP genome, we intend to construct a genetic map of RGAs, *Cr2* and other R genes, merged with general RAPD and AFLP markers.

The proteins encoded by R genes are thought to recognize specific Avr proteins produced by pathogens and activate signal transduction cascades for complex defense responses (Innes 1998, Grant et al. 2000). In these signaling processes, NBS domains play an important role. Diversity among NBS sequences is believed to be critical for the specificity of interaction either with a pathogen elicitor or with proteins downstream in the various signaling pathways that have been identified by genetic analysis (Aarts et al. 1998). Other studies using precise domain swapping have shown a number of LRRs to be essential for the interaction with avirulent factors (Van Der Hoorn et al., 2001; Wulff et al., 2001). Based on NBS sequences, the cloned RGAs in the WWP genome can be grouped into 14 classes (Fig. 2 and Table 3). The high degree of variation in NBS domains suggests that it might contribute in part to the evolutionary fitness of conifers.

R genes are under constant evolutionary pressure because they function to recognize novel Avr gene products produced by the evolving pathogen. Elucidation of the evolution of R genes is very important if we are to uncover the mechanisms by which plants maintain and adapt their defense to pathogens (Pan et al. 2000a). The evolution of R genes remains largely unexplored, but useful information has recently been gained from molecular genetic analyses. Various genetic mechanisms have been proposed to account for the evolution of R genes (Michelmore and Meyers 1998, Noël et al. 1999). It has been hypothesized that point mutations are the major evolutionary mechanism that leads to novel polymorphisms at R-loci (Michelmore and Meyers 1998). Sequence mispairing and unequal crossing-over can also account for R gene evolution (Noël et al. 1999). These analyses refer to crop and modern plants with short life cycles, and so far little data have been obtained from perennial plants like conifers. Our study shows that single nucleotide differences, insertions, deletions and duplications are also found in WWP RGAs, which is consistent with previous observations in angiosperms that revealed that R genes exhibit relatively high rates of mutations and recombination. For long-lived species such as conifers, with a life span of hundreds to thousands of years, the generation of new resistances via somatic mutations may be particularly advantageous (Michelmore and Meyers 1998). Comparisons of variability in the R gene domains that determine pathogen specificity among long-lived individuals need to be performed.

The ratio of nonsynonymous to synonymous changes (Ka/Ks) is an indicator of the evolutionary pressures on a class of genes. Within coding regions, a Ka/Ks ratio

greater than 1 reflects diversifying selection and a value lower than 1 suggests purifying selection (Parniske et al. 1997). According to this theory, when under strong selection to generate new variants, R genes should display a pattern in which nonsynonymous substitutions occur more frequently. In the R protein domains, Ka/Ks ratios greater than 1 are generally observed in the β -strand/ β -turn motifs of the LRR domain, providing evidence for diversifying selection. This is consistent with the prediction that the amino acids of these LRR motifs might be solvent-exposed and involved in ligand binding for R gene function (Graham et al. 2002). Other regions of resistance genes, like the NBS domains, typically do not show evidence of diversifying selection. Ka/Ks for NBS domains of known angiosperm R genes is about 0.37–0.44, suggesting that the NBS region is subject to purifying selection (Michelmore and Meyers 1998). In contrast to these observations, some Ka/Ks ratios in WWP NBS domains are significantly higher than 1, while others are not significantly different from 1 (Table 4). This evidence for different selection pressures on NBS domains suggests that the evolution of RGAs or R genes might differ between angiosperms and conifers.

Acknowledgements This research was funded by the Canadian Forestry Service, Canadian Biotechnology Strategy and Forest Innovation Investment awarded to A.K.M.E. We thank Miss Summer Lane for her technical assistance on plasmid extraction and restriction analysis (supported by the Government of Canada Youth Internship Program) and two anonymous reviewers for thoughtful comments on the manuscript.

References

- Aarts MGM, Hekkert BT, Holub EB, Beynon JL, Stiekema WJ, Pereira A (1998) Identification of R-gene homologous DNA fragments genetically linked to disease resistance loci in *Arabidopsis thaliana*. *Mol Plant-Microbe Interact* 11:251–258
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Cannon SB, Zhu H, Baumgarten AM, Spangler R, May G, Cook DR, Young ND (2002) Diversity, distribution, and ancient taxonomic relationships within the TIR and non-TIR NBS-LRR resistance gene subfamilies. *J Mol Evol* 54:548–562
- Collins NC, Webb CA, Seah S, Ellis JG, Hulbert SH, Pryor A (1998) The isolation and mapping of disease resistance gene analogs in maize. *Mol Plant-Microbe Interact* 11:966–978
- Creusot F, Macadre C, Ferrier-Cana E, Riou C, Geffroy V, Sevignac M, Dzron M, Langin T (1999) Cloning and molecular characterization of three members of the NBS-LRR subfamily located in the vicinity of the *Co-2* locus for anthracnose resistance in *Phaseolus vulgaris*. *Genome* 42:254–264
- Cooley M, Pathirana S, Wu HJ, Kachroo P, Klessig D (2000) Members of the *Arabidopsis* *HRT/RPP8* family of resistance genes confer resistance to both viral and oomycete pathogens. *Plant Cell* 12:663–676
- Dangl JL, Jones DG (2001) Plant pathogens and integrated defence responses to infection. *Nature* 411:826–833
- Davidson JJ, Ekramoddoullah AKM (1997) Analysis of bark protein in blister rust-resistant and susceptible western white pine (*Pinus monticola*). *Tree Physiol* 17:663–669

- Ellis J, Jones D (1998) Structure and function of proteins controlling strain-specific pathogen resistance in plants. *Curr Opin Plant Biol* 1:288–293
- Ekrמודدوللا AKM, Hunt RS (1993) Changes in protein profile of susceptible and resistant sugar pine foliage infected with the white pine blister rust fungus *Cronartium ribicola*. *Can J Plant Pathol* 15:259–264
- Ekrמודدوللا AKM, Hunt RS (2002) Challenges and opportunities in studies of host-pathogen interactions in forest tree species. *Can J Plant Pathol* 24:408–415
- Ekrמודدوللا AKM, Tan Y (1998) Differential accumulation of proteins in resistant and susceptible sugar pine (*Pinus lambertiana*) seedlings inoculated with white pine blister rust fungus (*Cronartium ribicola*). *Can J Plant Pathol* 20:308–318
- Ekrמודدوللا AKM, Davidson JJ, Taylor D (1998) A protein associated with frost hardiness of western white pine is up-regulated by infection in the white pine blister rust pathosystem. *Can J For Res* 28:412–417
- Ferrier-Cana E, Geffroy V, Macadre C, Creusot F, Imbert-Bollere P, Seignac M, Langin T (2003) Characterization of expressed NBS-LRR resistance gene candidates from common bean. *Theor Appl Genet* 106:251–261
- Graham MA, Marek LF, Shoemaker RC (2002) PCR sampling of disease resistance-like sequences from a disease resistance gene cluster in soybean. *Theor Appl Genet* 105:50–57
- Grant M, Brown I, Adams S, Knight M, Ainslie A, Mansfield J (2000) The *RPM1* plant disease resistance gene facilitates a rapid and sustained increase in cytosolic calcium that is necessary for the oxidative burst and hypersensitive cell death. *Plant J* 23:441–450
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41:95–98
- Harrison PM, Hegyi H, Balasubramanian S, Luscombe NM, Bertone P, Echols N, Johnson T, Gerstein M (2002) Molecular fossils in the human genome: identification and analysis of the pseudogenes in chromosomes 21 and 22. *Genome Res* 12:272–280
- Hirotsune S, Yoshida N, Chen A, Garrett L, Sugiyama F, Takahashi S, Yagami K-I, Wynshaw-Boris A, Yoshiki A (2003) An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature* 423:91–96
- Hunt RS (1997) Relative value of slow-canker growth and bark reaction as resistance responses to white pine blister rust. *Can J Plant Pathol* 19:352–357
- Innes RW (1998) Genetic dissection of R gene signal transduction pathways. *Curr Opin Plant Biol* 1:229–304
- Kanazin V, Marek LF, Shoemaker RC (1996) Resistance gene analogs are conserved and clustered in soybean. *Proc Natl Acad Sci USA* 93:11746–11750
- Kinloch BB, Sniezko RA, Barnes GD, Greathouse TE (1999) A major gene for resistance to white pine blister rust in western white pine from the western Cascade Range. *Phytopathology* 89:861–867.
- Kuhner MK, Felsenstein J (1994) A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Mol Biol Evol* 11:459–68
- Leister D, Ballvora A, Salamini F, Gebhardt C (1996) A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. *Nature Genet* 14:421–429
- Leister D, Kurth J, Laurie DA, Yano M, Sasaki T, Devos K, Graner A, Schulze-Lefert P (1998) Rapid reorganization of resistance gene homologues in cereal genomes. *Proc Natl Acad Sci USA* 95:370–375.
- Liu J-J, Ekrמודدوللا AKM (2003) Root-specific expression of a western white pine *PR10* gene is mediated by different promoter regions in transgenic tobacco. *Plant Mol Biol* 52:103–120.
- Liu J-J, Ekrמודدوللا AKM, Yu X (2003) Differential expression of multiple PR10 proteins in western white pine following wounding, fungal infection and cold-hardening. *Physiologia Plantarum*, in press
- Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND (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 BC, Morgante M, Michelmore RW (2002) TIR-X and TIR-NBS proteins: two new families related to disease resistance TIR-NBS-LRR proteins encoded in Arabidopsis and other plant genomes. *Plant J* 32:77–92
- Michelmore RW, Meyers BC (1998) Clusters of resistance genes evolve by divergent selection and a “birth-and-death” process. *Genome Res* 8:1113–1130
- Noël L, Moores TL, van der Biezen EA, Parniske M, Dzaniels MJ, Parker JE, Jones JDG (1999) Pronounced intraspecific haplotype divergence at the *RPP5* complex disease resistance locus of Arabidopsis. *Plant Cell* 11:2099–2111
- Pan Q, Wendel J, Fluhr R (2000a) Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. *J Mol Evol* 50:203–213
- Pan Q, Liu Y-S, Budai-Hadrian O, Sela M, Carmel-Goren L, Zamire D, Fluhr R (2000b) Comparative genetics of nucleotide binding site-leucine rich repeat resistance gene homologues in the genomes of two dicotyledons: tomato and Arabidopsis. *Genetics* 155:309–322
- Parniske M, Hammond-Kosack KE, Golsten C, Thomas CM, Jones DA, Harrison K, Wulff BBH, Jones IDG (1997) Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the *Cf-4/9* locus of tomato. *Cell* 91:821–832
- Peñuela S, Danesh D, Young ND (2002) Targeted isolation, sequence analysis, and physical mapping of non TIR NBS-LRR genes in soybean. *Theor Appl Genet* 104:261–272
- Richly E, Kurth J, Lerster D (2002) Mode of amplification and reorganization of resistance genes during recent *Arabidopsis thaliana* evolution. *Mol Biol Evol* 19:76–84
- Rivkin MI, Vallejos CE, McClean PE (1999) Disease-resistance related sequences in common bean. *Genome* 42:41–47
- Robinson R, Sturrock RR, Davidson JJ, Ekrמודدوللا AKM, Morrison D (2000) Detection of endochitinase-like protein in roots of Douglas-fir infected with *Armillaria ostoyae* and *Phellinus weirii*. *Tree Physiol* 20:493–502
- Sambrook J, Fritsch E F, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Shen KA, Meyers BC, Islam-Faridi MN, Chin DB, Stelley DM, Michelmore RW (1998) Resistance gene candidates identified by PCR with degenerate oligonucleotide primers map to clusters of resistance genes in lettuce. *Mol Plant-Microbe Interact* 11:815–823
- Speelman E, Bouchez D, Holub EB, Beynon JL (1998) Disease resistance-gene homologs correlate with disease resistance loci of *Arabidopsis thaliana*. *Plant J* 14:467–474
- Van der Hoorn RA, Roth R, Joosten MHJ (2001) Identification of distinct specificity determinants in resistance protein Cf-4 allows construction of a Cf-9 mutant that confers recognition of avirulence protein AVR4. *Plant Cell* 13:273–285
- Van der Vossen EA, van der Voort JN, Kanyuka K, Bendahmane A, Sandbrink H, Baulcombe DC, Bakker J, Stiekema WJ, Klein-Lankhorst RM (2000) Homologues of a single resistance-gene cluster in potato confer resistance to distinct pathogens: a virus and a nematode. *Plant J* 23:567–576
- Wulff BB, Thomas CM, Smoker M, Grant M, Jones JD (2001) Domain swapping and gene shuffling identify sequences required for inducing of an Avr-dependent hypersensitive response by the tomato Cf-4 and Cf-9 proteins. *Plant Cell* 13: 255–272.
- Young ND (2000) The genetic architecture of resistance. *Curr Opin Plant Biol* 3:285–290.
- Yu YG, Buss GR, Maroof MA (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.