Identification and characterization of the WRKY transcription factor family in *Pinus monticola*

Jun-Jun Liu and Abul K.M. Ekramoddoullah

**Abstract:** The WRKY gene family represents an ancient and highly complex group of transcription factors involved in signal transduction pathways of numerous plant developmental processes and host defense response. Up to now, most WRKY proteins have been identified in a few angiosperm species. Identification of WRKY genes in a conifer species would facilitate a comprehensive understanding of the evolutionary and function-adaptive process of this superfamily in plants. We performed PCR on genomic DNA to clone WRKY sequences from western white pine (*Pinus monticola*), one of the most valuable conifer species endangered by white pine blister rust (*Cronartium ribicola*). In total, 83 *P. monticola* WRKY (*PmWRKY*) sequences were identified using degenerate primers targeted to the WRKY domain. A phylogenetic analysis revealed that *PmWRKY* members fell into four major groups (1, 2a+2b, 2c, and 2d+2e) described in *Arabidopsis* and rice. Because of high genetic diversity of the *PmWRKY* family, a modified AFLP method was used to detect DNA polymorphism of this gene family. Polymorphic fragments accounted for 17%–35% of total PCR products in the AFLP profiles. Among them, one WRKY AFLP marker was linked to the major resistance gene (*Cr2*) against *C. ribicola*. The results of this study provide basic genomic information for a conifer WRKY gene family, which will pave the way for elucidating gene evolutionary mechanisms in plants and unveiling the precise roles of *PmWRKY* in conifer development and defense response.

**Key words:** genetic map, phylogenetic analysis, transcriptional factor, western white pine, WRKY protein.

**Mots-clés:** carte génétique, analyse phylogénétique, facteur de transcription, pin argenté, protéines WRKY.

**Introduction**

As transcription factors, WRKY proteins play extensive roles in plant growth and development as well as in plant–pathogen interactions. They contain one or two copies of a DNA-binding domain that is composed of about 60 amino acids with the N-terminal sequence WRKYGXK followed by a zinc finger motif (C-X4-5-C-X22-23-H-X-H or C-X5-8-C-X25-28-H-X1-2-C) at the C-terminus (Eulgem et al. 2000; Zhang and Wang 2005). WRKY proteins regulate a number of defense-related genes, including pathogenesis-related (PR) genes, through interaction with the W-box (C/T)TGAC(T/C) in the promoter regions of these genes (Rushton et al. 1996; Du and Chen 2000; Eulgem et al. 2000; Ülker and Somssich 2004). WRKY genes appear to be unique to plants and comprise a large gene family with 74 members in *Arabidopsis thaliana*, 105 members in rice (*Oryza sativa*), and at least 109 members in soybean (*Glycine max*) (Wu et al. 2005, Zhang and Wang 2005). All of the angiosperm plants analyzed to
date have numerous members classified into three groups (Eulgem et al. 2000). Recently, these three groups were subdivided into five major groups (groups 1, 2a+2b, 2c, 2d+2e, and 3) based on the number of WRKY domains, intron positions, and the structural features of the zinc finger motif (Zhang and Wang 2005).

The roles of WRKY genes in regulating host defense response against pathogen infection have been elucidated in several investigations. In Arabidopsis, more than two thirds of AtWRKY members are regulated by bacterial infection and treatment with salicylic acid, suggesting a key role of this family in biotic stress (Dong et al. 2003; Kalde et al. 2003; Eulgem 2005). Ectopic expression of the AtWRKY18 gene led to enhanced expression of PR genes and resistance to the bacterial pathogen Pseudomonas syringae (Chen and Chen 2002). Overexpression of OsWRKY45 also enhanced resistance to rice blast fungus (Shimono et al. 2007). In contrast, analysis of both T-DNA insertion mutants and transgenic overexpression lines showed that AtWRKY7 and AtWRKY25 function as a negative regulator of salicylic acid mediated defense responses to P. syringae (Kim et al. 2006; Zheng et al. 2007). Pathogen-induced AtWRKY33 regulates the antagonistic relationship between defense pathways mediating responses to P. syringae and necrotrophic pathogens (Zheng et al. 2006). Arabidopsis protein (AtWRKY52 or RRS1), which has a nucleotide-binding site and leucine-rich repeats in addition to the WRKY domain, confers genetic resistance to bacterial wilt (Ralstonia solanacearum) by its physical interaction with PopP2, a type III pathogenic effector targeted to the plant nucleus (Deslandes et al. 2003). Recently, Shen et al. (2007) found a molecular mechanism of barley R (MLA) protein resistance to barley powdery mildew where the MLA protein directly interacts with a WRKY transcription factor in the nucleus, derepressing basal defense genes.

In addition to regulating biotic stresses in plants, a microarray study demonstrated that some AtWRKYs play important roles in Arabidopsis responses to cold, drought, and salt stresses (Seki et al. 2002). A barley gene (HvWRKY38) is also involved in cold and drought response (Maré et al. 2004). RNAi silencing analysis revealed that AtWRKY75 plays a modulator role in phosphate acquisition and root development in Arabidopsis (Devaiah et al. 2007).

Western white pine (Pinus monticola) is a very important forest species in western North America. This conifer is tolerant to frost and drought and is especially valuable because of its nonhost resistance to laminated root rot. These biological characteristics make P. monticola a good candidate species for forest restoration in western North America. Despite its economic and ecologic importance and long-term efforts over the past 50 years in breeding programs, western white pine is still not used for timber production because of its susceptibility to Cronartium ribicola, the causal agent of white pine blister rust (WPBR). To search for genetic resistance materials and to maintain the diversity of the genetic resources of WPBR resistance in western white pine populations is a difficult task in breeding programs. The most promising breeding strategy is involved in a major dominant resistance gene (Cr2) against WPBR (Kinloch et al. 1999). In the defense response triggered by WPBR infection, a variety of families of PR proteins are regulated in western white pine (Liu et al. 2004). One western white pine PR10 promoter contains W-boxes and directs a pathogen-induced gene expression (Liu et al. 2005). A W-box has also been found in the western white pine gene promoter for an antimicrobial peptide (PmAMP1) (Ekramoddoullah et al. 2006). Being transcription factors, WRKY proteins may take part in the white pine defense response against C. ribicola infection. Characterization of WRKY genes and analysis of their association with disease resistance will be significant for our understanding of host resistance mechanism in the WPBR pathosystem.

Although WRKY genes have been reported in more than 10 angiosperm species, little is known about the presence and function of this family in a gymnosperm species (Zhang and Wang 2005). To further understand the functions and evolution of the WRKY family in higher plants, here, we carried out a sequence search of WRKY genes in a conifer genome. The aim of this study was to characterize gene organization and polymorphism in the western white pine WRKY family (PmWRKY). A comparative phylogenetic analysis was performed on WRKY members from western white pine, loblolly pine (Pinus taeda), Arabidopsis, and rice to reveal genomic organization of gene family.

**Materials and methods**

**Plant materials**

*Pinus monticola* seed lot Nos. 3277, 3278, and 3566 were used in the present study, all of which are known to possess the *Cr2* gene conferring resistance to *C. ribicola*. Seed lot Nos. 3277 and 3278 were obtained from the Dorena Genetic Resources Center, USDA Forest Service (Liu and Ekramoddoullah 2007). Both of these seed lots were open-pollinated seeds from two mother resistant trees with USDA registration Nos. 119-15045-845 and 119-15045-845 × 15045-841, respectively (Kinloch et al. 1999). Seed lot No. 3566 was open-pollinated seeds from one mother tree, originally from a bulked seed orchard collection from the Dorena Genetic Resources Center. This mother tree was previously identified as heterozygous resistance (*Cr2/cr2*) (Liu et al. 2006).

**Genomic DNA searching of *PmWRKY* sequences**

Genomic DNA was extracted from needles of family Nos. 3277 and 3278 resistant seedlings using a Qiagen DNeasy kit (Qiagen Inc., Mississauga, Ontario). A targeted PCR approach was used to clone WRKY domain sequences from the western white pine genome. The conserved WRKY domain was used to design degenerate PCR primers for *WRKY* gene cloning and molecular marker development (Borrone et al. 2004). Degenerate primers were designed based on conserved motifs in the WRKY domain of different groups of WRKY proteins (Table 1). 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, California). 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 40 s, primer annealing at 42 °C for 1 min, and primer extension at 72 °C for 1.5 min with a final 10-min extension at 72 °C.

**DNA sequence analysis**

The DNA fragments were amplified by PCR and resolved...
by agarose gel electrophoresis. The DNA bands with expected sizes were excised and purified using a MinElute gel extraction kit (Qiagen) and cloned into the pGEM-T easy vector (Promega). Insert DNA sequences of recombinant clones were determined on both strands on an ABI310 DNA sequencer (Applied Biosystems) using a Thermo-cycle DNA sequencer kit (Amersham, Baie d’Urfe, Quebec). Nucleotide sequence data were compiled and analyzed using ExPASy Proteomics tools (Swiss Institute of Bioinformatics, Berne, Switzerland) and BLAST programs (National Center for Biotechnology Information, Bethesda, Maryland) (Altschul et al. 1997).

**Phylogenetic analysis**

Alignment analysis of nucleotide or putative amino acid sequences was performed online with the Clustal W network service at the European Bioinformatics Institute (Cambridge, UK). Based on the sequence alignment analysis, phylogenetic trees were constructed using the neighbor-joining method in the MEGA3 software package (Kumar et al. 1997). The reliability of each tree was established by conducting 1000 neighbor-joining bootstrap sampling steps. The PmWRKY sequences reported here have been registered in the GenBank database under accession Nos. EU69775 to EU69857 that correspond to PmWRKY1 to PmWRKY83 in number order.

**DNA polymorphism of WRKY sequences in western white pine**

To detect WRKY polymorphism, a modified AFLP method was used following a method described by Hayes and Maroof (2000). Megagametophyte genomic DNA from seed lot No. 3566 was digested with the restriction enzymes EcoRI and MseI and ligated with adaptors according to standard AFLP protocol (Vos et al. 1995). Primers MseI-CC and EcoRI-AC were used for preamplification in a PCR mixture of 25 μL total volume with 5 μL of diluted restriction/ligation DNA mixture as template, 0.4 μmol/L E-AC and M-CC primers, and 12.5 μL of 2× Taq Master Mix (Qiagen). Three WRKY primers, WRKY-F1, WRKY-R1, and WP-W-R (Table 1), were coupled with one of 16 EcoRI-AC + 2 selective primers, respectively, in the AFLP selective amplification. AFLP marker cloning was performed as described previously (Liu and Ekramoddoullah 2008).

To detect WRKY AFLP polymorphic markers in the Cr2 linkage, WRKY AFLP primer combinations were screened using bulked segregant analysis and haploid segregation analysis as performed previously (Liu et al. 2006). WRKY AFLP amplification products were separated in 7 mol/L urea – 6% polyacrylamide gels and the DNA fragment patterns were visualized by silver staining. Each marker was tested for Mendelian segregation by $\chi^2$ ($\alpha = 0.05$) using the locus genotype frequency analysis of JoinMap version 3.0 software (Van Ooijen and Voorrips 2001) and mapped on the Cr2 linkage using JoinMap version 3.0 with the Kosambi mapping function provided by the software.

### Results

**Characterization of WRKY genomic DNA sequences**

Six combinations of degenerate primers (F1/R1, F1/R2, F1/R3, F1/R4, F1/R5, and F1/R6) were used to clone PmWRKY genomic sequences (Table 1). These primer pairs produced different amplicon patterns under the same PCR conditions. In total, five genomic DNA fragments were amplified to sizes of about 0.27, 0.43, 0.49, 0.64, and 1.0 kb. Following purification and ligation, these DNA fragments were cloned into the pGEM-T easy vector. Based on restriction enzyme analysis, 246 genomic clones were selected for nucleotide sequence analysis. DNA sequences encoding for WRKY domains were obtained. In case of possible PCR integration errors, only genomic sequences with nucleotide identities of 98% or less were considered as distinct members of the PmWRKY family for further analysis. Genomic clones without significant similarity to angiosperm WRKY sequences were highly divergent, with nucleotide identities ranging from 6% to 98%. In a phylogenetic analysis based on nucleotide sequence alignment, these PmWRKY genomic clones were grouped into four monophyletic clusters with general nucleotide identity below 18% between groups (Fig. 1). These monophyletic clusters corresponded to angiosperm WRKY groups 1, 2a+2b, 2c, and 2d+2e (Zhang and Wang 2005). PmWRKY group 2a+2b was the most complex and consisted of 74 members (PmWRKY1 to PmWRKY74) with nucleotide identities ranging from 28% to 98%, and it was subdivided into 13 classes (1 to XIII) with a general nucleotide identity

### Table 1. PCR primers for WRKY cloning and WRKY AFLP polymorphism detection in Pinus monticola.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Targeted amino acid motif</th>
<th>WRKY group</th>
<th>Nucleotide sequence (5’ to 3’)</th>
<th>Reference/accession No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>WRKY-F1</td>
<td>WRKYGGK</td>
<td>All groups</td>
<td>TGGGMGIAARTAYGGGNCARA</td>
<td>Borrono et al. 2004</td>
</tr>
<tr>
<td>WRKY-R1</td>
<td>TTYEG(Q/V)H(N/T)H</td>
<td>Group1, 2</td>
<td>TGRKTRTGYSICCYCTRAITA</td>
<td>Borrono et al. 2004</td>
</tr>
<tr>
<td>WRKY-R2</td>
<td>TY(T/E)G(E/D)IH</td>
<td>Group2d+2e</td>
<td>TGRTRTGYTCIICKYRTAIGT</td>
<td>NM_122748, DAA05104</td>
</tr>
<tr>
<td>WRKY-R3</td>
<td>SYLGRNHN</td>
<td>Group2c</td>
<td>TGRTRTGYCNCCNAGRTANCT</td>
<td>DR024229</td>
</tr>
<tr>
<td>WRKY-R4</td>
<td>TXYGEHTC</td>
<td>Group3</td>
<td>CAIGTRGTYCICIIIRTAIGT</td>
<td>DAA05110, ABF95809</td>
</tr>
<tr>
<td>WRKY-R5</td>
<td>TYGEHCTC</td>
<td>Group3</td>
<td>CANGTRGTYCICIIIRTAIGT</td>
<td>DAA05110</td>
</tr>
<tr>
<td>WRKY-R6</td>
<td>TYYGHHTC</td>
<td>Group3</td>
<td>CANGTRGTYCICIIIRTAIGT</td>
<td>DAA05110</td>
</tr>
<tr>
<td>WP-W-R</td>
<td>TEIVYKG</td>
<td>Group1</td>
<td>CCYTTGTAAMCWA TTGMC</td>
<td>AK226301</td>
</tr>
</tbody>
</table>


*Primers were synthesized according to the reference or designed in the present study based on amino acid sequences as indicated by GenBank accession numbers.
Fig. 1. Phylogenetic analysis of *PmWRKY* genomic sequences. The phylogenetic tree was generated based on Clustal W alignment analysis of 83 *PmWRKY* genomic nucleotide sequences, including intron sequences, using a neighbour-joining method. A bootstrap of 1000 replications was employed to evaluate the reliability of the tree branching. The numbers I–XIII indicate classification of *PmWRKY* sequences into 13 similarity classes with an 82% identity threshold. The scale bar at the bottom indicates genetic distance proportional to the nucleotide substitutions per site.
threshold of 82% between classes. In contrast, only two members were identified in group 1 (\textit{PmWRKY82} and \textit{PmWRKY83}) and group 2c (\textit{PmWRKY75} and \textit{PmWRKY82}), respectively. There were five members in group 2d+2e (\textit{PmWRKY80} to \textit{PmWRKY83}) that were further divided into two classes with a nucleotide identity threshold of 65% between themselves.

Comparison of western white pine and angiosperm WRKY domains

Alignment analysis of genomic DNA sequences of \textit{PmWRKY} genes with \textit{AtWRKY} and \textit{OsWRKY} sequences from \textit{Arabidopsis} and from rice demonstrated that the intron positions in WRKY domains are conserved between gymnosperms and angiosperms (Fig. 2). The positions and phases of the introns in the WRKY domains of \textit{PmWRKY} genes were the same as those identified in WRKY groups 1, 2a+2b, and 2d+2e from \textit{Arabidopsis} and rice (Xie et al. 2005; Zhang and Wang 2005). In group 2a+2b, the phase 0 intron was found after the fifth codon downstream of CX5C, and in group 2c and 2d+2e, the phase 2 intron was localized after the second nucleotide of the fifth codon upstream CX4–5C. As in the angiosperms, the intron position in the C-terminal WRKY domain of \textit{PmWRKY} group 1 members was also conserved and localized at the same position and phase as those in groups 2c and 2d+2e. All 5\textsuperscript{'-}exon/intron and 3\textsuperscript{'-}intron/exon boundaries conformed to the well-known GT/AG donor/acceptor site rule except in the case of \textit{PmWRKY80} where a rare
Further investigation is needed to confirm this rare donor/acceptor site GT/AA because of possible PCR error.

After putative pre-mRNA splicing, 81 of 83 genomic DNA clones contain continuous open reading frames (ORF) encoding WRKY domains with homologies to those from angiosperms (Fig. 2). PmWRKY28 had a site mutation that resulted in a premature stop codon just ahead of the intron, and PmWRKY62 had two single base insertions, leading to ORF shifts, suggesting that these two WRKY sequences may be pseudogenes or PCR artifacts.

Translation of 83 PmWRKY DNA sequences produced 62 distinct WRKY domain sequences with amino acid identities from 40% to 98%. Fourteen PmWRKY DNA sequences shared the same amino acid sequence as the following: PmWRKY2, PmWRKY9, PmWRKY13, and PmWRKY16; PmWRKY3, PmWRKY8, PmWRKY10, and PmWRKY11; PmWRKY6 and PmWRKY18; PmWRKY30 and PmWRKY31; PmWRKY33 and PmWRKY35; PmWRKY37, PmWRKY39, and PmWRKY40; PmWRKY46, PmWRKY47, and PmWRKY48; PmWRKY52 and PmWRKY56; PmWRKY53, PmWRKY54, and PmWRKY55; PmWRKY57 and PmWRKY58; PmWRKY60 and PmWRKY61; PmWRKY67 and PmWRKY70; PmWRKY68 and PmWRKY69; and PmWRKY77 and PmWRKY79. The WRKY domain is the dominant structural feature of WRKY proteins and is used to distinguish groups of this superfamily (Fig. 2). Generally, PmWRKY members shared higher identities inside the group (63%–98%) than among groups (40%–64%). The amino acid replacements in the WRKY domain in each group may reflect evolutionary divergence of individual family members. Those distinct DNA sequences sharing identical WRKY domains were either different gene members or alleles of the same gene members.

To comprehensively understand the relationship between angiosperm and gymnosperm WRKY proteins, alignment analysis of the putative amino acid sequences was performed to construct a phylogenetic tree (Fig. 3). Apart from representative members from Arabidopsis (AtWRKY) and rice (OsWRKY), this phylogenetic analysis also included 29 P. teada WRKY genes that we searched out from the DFCI Pine Gene Index (release 6.0, 19 July 2005, a total of 45,557 output sequences) provided by the Computational Biology and Functional Genomics Laboratory at the Dana-Farber Cancer Institute and Harvard School of Public Health (http://compbio.dfci.harvard.edu/cgi-bin/tgi/Blast/index.cgi) (Table 2). PmWRKY proteins were assigned to different groups based on phylogenetic analysis with well-classified WRKY proteins from Arabidopsis (Eulgem et al. 2000)
and rice (Zhang and Wang 2005). According to the WRKY classification in Arabidopsis and rice, PmWRKYs were clustered into four WRKY groups: 1, 2a+2b, 2c, and 2d+2e (Fig. 3). The WRKY proteins were clustered in a similar pattern to that of the gene nucleotide sequences. The phylogenetic tree indicates that a common gene may have duplicated to give rise to the ancestor genes of each group before the split of gymnosperms and angiosperms. In groups 1 and 2d+2e, most conifer genes were diversely distributed in the clades containing angiosperm members. In groups 2a+2b and 2c, most conifer sequences formed monophyletic clades separable from Arabidopsis and rice homologues, suggesting that in large part, these WRKY genes might have evolved independently in the conifers.

Polymorphism of WRKY-related sequences

Three WRKY primers (WRKY-F1, WRKY-R1, and WP-W-R) were combined with 16 EcoRI-AC+2 selective primers to perform selective PCR for WRKY AFLP genotyping. The detectable number of DNA fragments amplified by each of 48 primer combinations ranged from 53 (WRKY-R1/EcoRI-ACCG) to 67 (WRKY-R1/EcoRI-ACAC) with an average of 60. The sizes of amplified bands ranged from 60 to 800 bp. The polymorphic bands accounted for 17%–35% of the total bands derived from these primer combinations.

To explore potential WRKY markers linked to the Cr2 gene against WPBR (C. ribicola), bulked segregant analysis was used to reveal polymorphism between resistant and susceptible bulks. Among the 48 primer combinations for WRKY AFLP amplification, only one pair of primers (WPW-R/EcoRI-ACTC) detected a polymorphic DNA fragment of 302 bp linked to Cr2 in the repulsion phase (Fig. 4). Segregation of this Cr2-specific WRKY AFLP marker, W/E13-302r, was tested in family No. 3566 and was found in a 1:1 ratio for the presence and absence of the marker. This WRKY AFLP marker was mapped onto the Cr2 linkage with the five RAPD loci that we identified previously (Liu et al. 2006). In this genetic map of the Cr2 gene, W/E13-302r was 2.9 cM away from Cr2 (Fig. 5).

The DNA marker W/E13-302r and the other 12 WRKY AFLP polymorphic fragments were cloned and sequenced (Table 3). Sequence analysis demonstrated that those 12 fragments contained WRKY-specific and AFLP selective primers at their 5’ and 3’ ends. Thus, all but one of the markers contained amino acid sequences corresponding to WRKY primers. The polymorphic DNA fragment W-E5-283 was a nonspecific AFLP fragment amplified from one single AFLP selective primer, EcoRI-ACAC (Table 3). The BLASTP program of “search for short, nearly exact matches” found that eight DNA fragments, including W/E13-302r, encoded addi-
tional amino acid sequences extended from corresponding WRKY primers that were significantly similar to WRKY sequences, whereas the other five fragments showed no homology with any WRKY proteins (Table 3). The low level of WRKY AFLP marker homology with known WRKY genes may result from short sequences with limited ORFs and the presence of introns in the WRKY domains. Long-distance genomic DNA walking is necessary to definitively identify western white pine WRKY AFLP polymorphic fragments.

Discussion

A superfamily of WRKY proteins has been characterized in numerous angiosperm species. However, database search revealed only a few members in lower plants (algae, moss, and fern) and conifers (Eulgem et al. 2000; Ulker and Somssich 2004; Wu et al. 2005; Zhang and Wang 2005). A previous survey of WRKY genes from 81,802 Pinus ssp. ESTs revealed only four sequences (three in group 1 and one in group 2d+2e), suggesting a lower WRKY member abundance in conifers than in angiosperms (Zhang and Wang 2005). This led to questions about when and how evolutionary expansion has occurred for this family in the plant kingdom. The present study presents the first comparative genomic study and evolutionary analysis of the WRKY family of genes in a gymnosperm species. Through extensive searching of its genome, we identified 83 members of the PmWRKY family in western white pine. Based on the phylogeny of these sequences at both the nucleotide and amino acid levels, PmWRKY genes were clustered into four WRKY groups (1, 2a+2b, 2c, and 2d+2e). Our search of a P. taeda EST databank with 45,557 output sequences found only 29 PtWRKY genes, and these were distributed in a similar grouping pattern as those in P. monticola. No group 3 member was found in the P. taeda transcriptome.

The distribution of the cloned sequences demonstrated that not all WRKY genes and alleles are amplified equally in western white pine by the primer sets that we used. Although reverse primers R4, R5, and R6 (Table 1) were based on group 3 WRKY members from Arabidopsis, rice, and tree species (Populus), respectively, no member was isolated from group 3 in western white pine. Borrone et al. (2004) reported a preferential amplification of group 1 and group 2 WRKY genes in Theobroma cacao. We characterized about 83 members of the PmWRKY family using six

Fig. 4. WRKY AFLP polymorphic profiles. AFLP selective amplification was performed using one WRKY primer coupled with one AFLP selective primer. The arrows point to the polymorphic DNA fragments amplified by primer pair WP-W-R/EcoRI-CTC. One WRKY AFLP marker with size 302 bp (W/E13-302r) was linked to Cr2 in repulsion. Samples of 15 resistant individuals and 16 susceptible individuals are shown here.
the WRKY family has undergone gene duplication events and subsequent divergences (Eulgem et al. 2000; Dong et al. 2003; Zhang and Wang 2005). Intron number and position in the WRKY domain are important gene structural characteristics for gene classification in rice and *Arabidopsis*. The intron positions of gene members of the *PmWRKY* family in each group were the same as their homologs in angiosperms. Zhang and Wang (2005) proposed that genes of group 1 with only a C-terminal WRKY domain are ancestors of other descendant genes in the other four groups (2a+2b, 2c, 2d+2e, and 3). In each group, *PmWRKY* genes showed higher identities with their corresponding WRKY genes from *Arabidopsis* and rice than with western white pine members in other groups, suggesting that the duplication event for group differentiation might have occurred before the separation of gymnosperms and angiosperms. However, the ancestor gene numbers of each group should be limited. With a databank search, Zhang and Wang (2005) found one gene of group 1 in a green alga (*C. reinhardtii*) and fern (*Ceratopteris richardii*). The moss *Physcomitrella* genome contains only 38 WRKY genes (three in group 1, seven in group 2a+2b, 17 in group 2c, five in group 2d+2e, five in group 3, and one other) (Rensing et al. 2008). The later evolution of ancestor genes of each group may have given rise to different members, leading to a rapid and great expansion of each group in seed plants.

In each group, the subgrouping pattern into classes showed that in general the *PmWRKY* members were more similar to each other and to *P. taeda* members than to those in *Arabidopsis* and rice (Fig. 3). Most conifer WRKY genes were clustered together, phylogenetically distinct from the homologous members in *Arabidopsis* and rice. During evolutionary processes, orthologs are usually under selective pressure to maintain gene functions, whereas paralogs tend to accumulate mutations for deviation of gene function or regulation. In the transcription factor MADS-box gene family, an orthologous relationship was detected between angiosperms and gymnosperms (Liu et al. 2003). Here, we did not detect orthologous relationship between angiosperm and gymnosperm WRKY members, suggesting that the members in each WRKY group might have expanded independently in gymnosperms and angiosperms. This type of gene divergence between angiosperms and gymnosperms was previously observed in another western white pine superfamily related to disease/pest resistance, the nucleotide-binding site – leucine-rich repeat (NBS-LRR) gene family (Liu and Ekramoddoullah 2003, 2007). Both tandem duplications and large-scale duplications were implicated in the expansion of the NBS-LRR gene family (Meyers et al. 2003). It awaits further investigation to determine whether a similar expansion pattern has happened to the WRKY family after the split of gymnosperms and angiosperms. It is believed that the intra- and intergenic recombination or structural chromosome rearrangements play an important role in the gene divergence of the angiosperm WRKY family (Xie et al. 2005; Zhang and Wang 2005).

Several WRKY members were found to play a variety of developmental and physiological roles in plants, including plant immune responses to abiotic and biotic stresses (Ülker and Somssich 2004; Eulgem 2005), regulating embryogenesis (Lagace and Matton 2004), senescence (Robatzek and Somssich 2002), and nutrient starvation response and root
Table 3. WRKY AFLP polymorphic DNA fragments in *Pinus monticola* detected using a modified AFLP procedure.

<table>
<thead>
<tr>
<th>Marker name&lt;sup&gt;a&lt;/sup&gt;</th>
<th>PCR primers</th>
<th>Putative amino acid sequence&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Homologous WRKY (species) (GenBank accession No.)&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Homology score (bits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>W-E1-275</td>
<td>WRKF/EcoRI-ACCC</td>
<td>WRKYGGIKVTVDTS</td>
<td>WRKY25 (<em>Arabidopsis thaliana</em>) (NP_180584.1)</td>
<td>29.9 (63)</td>
</tr>
<tr>
<td>W-E2-266</td>
<td>WRKF/EcoRI-ACAC</td>
<td>KYGQREK&lt;sup&gt;a&lt;/sup&gt;</td>
<td>NS</td>
<td>—</td>
</tr>
<tr>
<td>W-E3-408</td>
<td>WRKF/EcoRI-ACCT</td>
<td>WRKYGQTVNLEIHQ</td>
<td>WRKY46 (<em>Arabidopsis thaliana</em>) (AAK96020.1)</td>
<td>29.5 (62)</td>
</tr>
<tr>
<td>W-E4-175</td>
<td>WRKR/EcoRI-ACAC</td>
<td>NNLMQATTYEGKHT</td>
<td>WRKY33 (<em>Arabidopsis thaliana</em>) (NP_181381.2)</td>
<td>27.8 (58)</td>
</tr>
<tr>
<td>W-E5-283</td>
<td>WRKR/EcoRI-ACAC</td>
<td>AFLP primer at both ends</td>
<td>NS</td>
<td>—</td>
</tr>
<tr>
<td>W-E6-345</td>
<td>WRKR/EcoRI-ACCG</td>
<td>ASQFRRLISNLF</td>
<td>WRKY3 (<em>Nicotiana attenuata</em>) (AAS13439.1)</td>
<td>27.4 (57)</td>
</tr>
<tr>
<td>W-E7-376</td>
<td>WRKR/EcoRI-ACCG</td>
<td>ERYAPIHPSRTYEGKH</td>
<td>WRKY2 (<em>Nicotiana benthamiana</em>) (AAS55706.1)</td>
<td>27.4 (57)</td>
</tr>
<tr>
<td>W-E8-173</td>
<td>WRKR/EcoRI-ACCG</td>
<td>*NLTYEYKHT</td>
<td>NS</td>
<td>—</td>
</tr>
<tr>
<td>W-E9-376</td>
<td>WRKR/EcoRI-ACCG</td>
<td>ERYAPIHPSRTYEGKH</td>
<td>WRKY2 (<em>Nicotiana benthamiana</em>) (AAS55706.1)</td>
<td>27.4 (57)</td>
</tr>
<tr>
<td>W-E10-286</td>
<td>WPWR/EcoRI-ACTC</td>
<td>GTIISKYENKLTEIVYKG</td>
<td>NS</td>
<td>—</td>
</tr>
<tr>
<td>W-E11-765</td>
<td>WPWR/EcoRI-ACTC</td>
<td>KIESFITEIYK</td>
<td>WRKY1 (<em>Pimpinella brachycarpa</em>) (AAC31956.1)</td>
<td>28.2 (59)</td>
</tr>
<tr>
<td>W-E12-447</td>
<td>WPWR/EcoRI-ACTC</td>
<td>*KNFTEIYKKG</td>
<td>NS</td>
<td>—</td>
</tr>
<tr>
<td>W-E13-302&lt;sup&gt;r&lt;/sup&gt;</td>
<td>WPWR/EcoRI-ACTC</td>
<td>GEKJQVKVRSSRERIKTK</td>
<td>WRKY30 (<em>Oryza sativa</em>) (AAW63719.1)</td>
<td>31.2 (66)</td>
</tr>
</tbody>
</table>

<sup>a</sup>The second number in the marker indicates the DNA fragment length (bp).

<sup>b</sup>A stop codon is represented by an asterisk.

<sup>c</sup>NS, no homologous WRKY was detected using the “search for short, nearly exact matches” program in PLASTP.
Furthermore, the abundance and diversity of the motif primers has led to the isolation of WRKY sequences. A special phenotype could provide an accessible pathway to identify potential candidate markers for conifer breeding. The diversity and variations of the PmWRKY class of WRKY genes in each class suggests that paralogous evolution might have given rise to those related members, contributing to functional diversity in this superfamily. The results of the current study provide the necessary genomic information for further investigations into the molecular functions of PmWRKYs and their evolutionary expansion in a conifer genome. Future research is required to reveal the full complement of PmWRKY sequences. Identification of full-length and additional genes in conifers would facilitate a more comprehensive understanding of the evolutionary and function-adaptive process of this gene superfamily in plants.

Because of the large and divergent family of WRKY genes with multiple regulatory functions in different plant processes, it is possible to develop functional markers of candidate genes to understand their roles. WRKY genes are potential regulatory candidates for plant defense response against pathogen attack and other environmental stresses (Cheong et al. 2002). A ligation-mediated PCR revealed that genetic loci related to WRKY regulatory genes and other defense genes were significantly associated with potato quantitative disease resistance (Trognitz et al. 2007). The candidate gene approach is an effective method to investigate molecular mechanisms underlying complex plant traits such as disease resistance. For conifer species with large genomes (>107 kb), the association of a functional marker with a special phenotype could provide an accessible pathway leading to the identification of key genetic components.

Characterization of PCR products amplified by conserved motif primers has led to the isolation of WRKY sequences. Furthermore, the abundance and diversity of the WRKY gene family have been used for gene linkage mapping (Borrone et al. 2004). One WRKY-related primer in combination with a linker-specific primer was used to generate WRKY AFLP markers and map WRKY-related loci in potato (Trognitz et al. 2002). The present study developed WRKY AFLP markers in western white pine using one WRKY primer coupled with one typical AFLP selective primer. This modified AFLP strategy has been used for genetic mapping of disease resistance gene analogs (Hayes and Maroof 2000; Liu and Ekramoddoullah 2007) and transposable elements (Teunissen et al. 2003). The proportion of polymorphic bands revealed in WRKY AFLP (17%–35%) was similar to that of the standard AFLP (26%–40%) in western white pine. The PmWRKY-related DNA markers will help in understanding gene organization and evolution of this transcription factor family in a conifer species.

In conclusion, we characterized 83 sequences of the PmWRKY family through searching the western white pine genome. Our investigation further demonstrated that the diversity and variations of the PmWRKY family could provide potential candidate markers for conifer breeding.

Acknowledgements

This research was supported by Forest Genomics funding from Natural Resources Canada, Canadian Forest Service. The authors appreciate the critical review of the manuscript by Dr. M. Elliott and H. Williams.

References


