Chapter 2

In Silico Mining and PCR-Based Approaches to Transcription Factor Discovery in Non-model Plants: Gene Discovery of the WRKY Transcription Factors in Conifers

Jun-Jun Liu and Yu Xiang

Abstract

WRKY transcription factors are key regulators of numerous biological processes in plant growth and development, as well as plant responses to abiotic and biotic stresses. Research on biological functions of plant WRKY genes has focused in the past on model plant species or species with largely characterized transcriptomes. However, a variety of non-model plants, such as forest conifers, are essential as feed, biofuel, and wood or for sustainable ecosystems. Identification of WRKY genes in these non-model plants is equally important for understanding the evolutionary and function-adaptive processes of this transcription factor family. Because of limited genomic information, the rarity of regulatory gene mRNAs in transcriptomes, and the sequence divergence to model organism genes, identification of transcription factors in non-model plants using methods similar to those generally used for model plants is difficult. This chapter describes a gene family discovery strategy for identification of WRKY transcription factors in conifers by a combination of in silico-based prediction and PCR-based experimental approaches. Compared to traditional cDNA library screening or EST sequencing at transcriptome scales, this integrated gene discovery strategy provides fast, simple, reliable, and specific methods to unveil the WRKY gene family at both genome and transcriptome levels in non-model plants.

Key words: Conifer, gene discovery, in silico mining, non-model plant, PCR-based gene cloning, WRKY transcription factor.

1. Introduction

WRKY proteins constitute a superfamily of plant transcription factors involved in plant growth, development, and plant interactions with environmental factors (1). These proteins contain one or two copies of a DNA-binding domain, designated as a WRKY domain, which is composed of about 50–60 amino acids with the N-terminal conserved motif WRKYGXK and a zinc

L. Yuan, S.E. Perry (eds.), *Plant Transcription Factors*, Methods in Molecular Biology 754,
DOI 10.1007/978-1-61779-154-3_2,

finger motif (C-X_{4-8}-C-X_{22-28}-H-X_{1-2}-H/C) at the C-terminus (1). WRKY proteins regulate transcript expression of downstream genes through their interaction with the *cis*-element W-box, (C/T)TGAC(T/C), localized in the promoter regions of the target genes (1–4). Large WRKY families have been identified in *Arabidopsis thaliana* and in rice (*Oryza sativa*) with 72 and 105 members, respectively (5, 6). Biological functions of a few WRKY genes have been characterized in angiosperms (1). Those co-expressed WRKY genes play their biological roles through co-regulatory networks in *Arabidopsis* and rice (7). All angiosperm plants analyzed to date have numerous WRKY genes, and angiosperm WRKY genes are classified into three distinct groups and five subgroups (I, IIa+b, IIc, IId+e, and III) based on the copy number of the WRKY domain, intron positions and phases, and the structural features of the zinc finger motif (1, 6). As large-scale EST data became available from a limited number of other plant species, several WRKY-homologous sequences have also been reported in lower plants (alga, moss, and ferns) and in conifers by searching DNA databases (1, 4–6, 8). This gives rise to questions about how evolutionary expansion has occurred for this gene superfamily in the plant kingdom and how WRKY proteins have adapted to various biological functions by structural differentiation.

In model plants, such as *Arabidopsis* and rice, identification of transcription factor families usually involves cDNA library construction, followed by DNA sequencing and gene annotation at the transcriptome scale. These methods are expensive and time-consuming and require a large amount of genomic information of the target species. It is difficult to apply a similar approach in non-model plants, such as conifers (which have the largest plant genomes), because of very limited genomic information in these species, scarcity of regulatory mRNAs in transcriptomes, and a specific spatiotemporal pattern of the expressed target genes. We recently demonstrated that a combination of an in silico-based bioinformatic prediction strategy with PCR-based experimental approaches could identify the WRKY family in western white pine (*Pinus monticola* Dougl. ex D. Don) (8). This simple, efficient, and inexpensive strategy provides a practical method for gene discovery of the WRKY superfamily in other non-model plants.

A BLAST search (9) of EST databases of the Gene Index Project, 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/tgi/tgipage.html), identified 29 and 31 sequences in *Pinus taeda* and *Picea* species, respectively, with similarity to WRKY domains of the five subgroups from *Arabidopsis* and rice (**Table 2.1**). The amino acid motifs conserved among angiosperm and gymnosperm WRKY domains were determined by an alignment analysis and used to design a range of PCR primers

Table 2.1
Expressed members of the WRKY family in *Pinus* and *Picea* species

	Pinus taeda		*Picea*		
Group[a]	Gene No.	[EST sequence[b]]	Gene No.	[EST sequence[b]]	Species
I	1	[TC67599, TC10778, TC25556, TC31987, TC49174, TC5325]	1	[TC59872, TC2087, TC27577, EX331828]	*P. glauca*
I	2	[TC67819]	2	[TC53421, TC25771, DR588237]	*P. glauca*
I	3	[TC70314, TC11712, TC19803, TC37973, TC44417, TC6838]	3	[TC73002, ES665011]	*P. sitchensis*
I	4	[DR102534]		–	
IIa+b	5	[TC67611, TC29306, TC42772]	4	[TC72927, ES228321, ES228512, ES228759]	*P. abies*
IIa+b	6	[TC69912]	5	[EX371111]	*P. glauca*
IIa+b	7	[TC69221]	6	[TC59906, TC3742, TC20136, EX406580]	*P. glauca*
IIa+b	8	[CO198802]	7	[TC63325, TC24286, DR466155, CO207672]	*P. engelmannii* × *glauca*
IIa+b	9	[TC64319, TC45579]	8	[TC48570, ES261032, ES261221]	*P. sitchensis*
IIa+b	10	[TC75572, TC49622]	9	[TC49606, TC5670, TC37162, DR480521]	*P. sitchensis*
IIa+b	11	[TC62327]	10	[TC53881, TC17556, DR493091]	*P. sitchensis*
IIa+b	12	[TC68833, TC48794]		–	
IIa+b	13	[CO362998]		–	
IIc	14	[TC77638, TC44174]	11	[TC46708, EX396574, EX396867]	*P. glauca*

Table 2.1 (continued)

	Pinus taeda		*Picea*		
Group[a]	**Gene No.**	**[EST sequence[b]]**	**Gene No.**	**[EST sequence[b]]**	**Species**
IIc	15	[TC76195]	12	[TC57747, TC5412, TC28007, EX335523]	*P. glauca*
IIc	16	[TC69847]	13	[TC62191, TC23366, DV997756]	*P. glauca*
IIc	17	[TC77167]	14	[TC48310, TC30239, DR475708]	*P. engelmannii* × *glauca*
IIc	18	[TC71896]	15	[TC59853, TC2343, TC21165, DR475811]	*P. engelmannii* × *glauca*
IIc	19	[AW437880, ST73F02]	16	[ES260958]	*P. sitchensis*
IIc	20	[CF478959]	17	[ES871590]	*P. sitchensis*
IIc	21	[TC57307, TC20439, TC35508, TC47838][c]	18	[TC73769, TC6935, TC38386, FD731617]	*P. sitchensis*
IIc	22	[CF665435]		–	
IId+e	23	[DR694646, EST1084738]	19	[TC53444, ES875336]	*P. sitchensis*
IId+e	24	[TC59335, TC53663]	20	[TC52792, TC2588, TC20660, FD735167]	*P. sitchensis*
IId+e	25	[CF401103][d]	21	[DR503879]	*P. sitchensis*
IId+e	26	[TC66817, TC16416, TC26277, TC28002, TC41928, TC5870]	22	[DR495557]	*P. sitchensis*

Table 2.1
(continued)

	Pinus taeda		*Picea*		
Group[a]	**Gene No.**	**[EST sequence[b]]**	**Gene No.**	**[EST sequence[b]]**	**Species**
IId+e	27	[TC75206, TC36489, TC39130, TC48810, TC56850]	23	[DV982911]	*P. glauca*
IId+e	28	[TC75557]	24	[TC58586, EX386655, EX387020]	*P. glauca*
IId+e	29	[TC76519, TC55982]	25	[TC55698, EX328070, EX328428]	*P. glauca*
IId+e		–	26	[EX389580]	*P. glauca*
IId+e		–	27	[TC51682, EX307059]	*P. glauca*
IId+e		–	28	[TC80753, TC39381, DV986854]	*P. glauca*
IId+e		–	29	[TC66343, EX431002, EX402933]	*P. glauca*
IId+e		–	30	[TC83188, TC5604, DR581126]	*P. glauca*
IId+e		–	31	[DV979750]	*P. glauca*
III		–	32	[EX419977]	*P. glauca*

[a]Phylogenetic classification according to Zhang and Wang (6)
[b]GenBank accession numbers or the DFCI Gene Index (PGI) numbers (TCxxxxx)
[c]ORF interrupted by not processed intron sequence
[d]ORF interrupted by single nucleotide deletion

Table 2.2
PCR primer design for WRKY gene cloning

Primer name	Targeted a.a. motif	WRKY group	Nucleotide sequence (5′ to 3′)[a]	Reference/Accession No.[b]
WRKY-FP1	WRKYGQK	All groups	TGGMGIAARTAYGGNCARA	(11)
WRKY-FP2	RWRKYGQK	All groups	CGMTGGCGBAARTATGGACARAA	This study
WRKY-RP0	TEIVYKG	Group I	CCYTTGTAMACWAT TTCMGT	(8)/AK226301
WRKY-RP1	TTYEG(Q/V)H(N/T)H	Group I, IIa+b	TGRKTRTGYWSICCYTCRTAIGT	(11)
WRKY-RP2	SYLGRHNH	Group IIc	TGRTTRTGYCTNCCNAGRTANCT	(8)/DR024229
WRKY-RP3	TY(T/E)G(E/D)HNH	Group IId+e	TGRTTRTGYTCICCIKYRTAIGT	(8)/NM_122748, DAA05104
WRKY-RP4	TYXGEHTC	Group III	CAIGTRTGYTCICCIIIRTAIGT	(8)/DAA05110, ABF95809
WRKY-RP5	TYIGEHTC	Group III	CANGTRTGYTCNCCNATRTANGT	(8)/DAA05110
WRKY-RP6	TYYGHHTC	Group III	CANGTRTGRTGNCCRTARTANGT	(8)/CK109883
WRKY-RP7	TYSGVHSC	Group III	CARCTRTGNACNCCRCTRTANGT	This study, EX419977

[a]Degenerate IUB group codes: R=A+G, Y=C+T, M=A+C, K=G+T, W=A+T, S=G+C, D=G+A+T, B=T+C+G, N=A+T+C+G
[b]Primers were synthesized according to reference or designed in the present study based on amino acid sequences as indicated GenBank accession numbers

(**Table 2.2**). Using these WRKY domain primers, reverse transcription-polymerase chain reaction (RT-PCR) and rapid amplification of cDNA ends (RACE) were then performed to amplify expressed WRKY domain sequences and their respective full-length coding regions in various tissues of white pine. The primers were also used to amplify genomic DNA sequences of the WRKY domains and to detect DNA polymorphism for genetic mapping of the WRKY family in white pine populations. We identified 83 members of the *P. monticola* WRKY family (*PmWRKY*) (8). Using the neighbor-joining (NJ) method in the MEGA4 software package (10), the identified conifer WRKY genes were classified into groups I, IIa+b, IIc, IId+e, and III by a phylogenetic analysis at both nucleotide and amino acid sequence levels.

The WRKY gene superfamily is highly complex and divergent, suggesting its application for development of functional gene markers in genetic mapping (11, 12). A genetic mapping strategy modified from amplified fragment length polymorphism (AFLP) was used for genetic mapping of analogs of disease resistance genes (13–15) and transposable elements (16). We developed WRKY-AFLP markers by replacing a typical AFLP selective primer with one of the WRKY primers in the AFLP standard protocol. This WRKY-AFLP approach revealed 17–35% polymorphic bands, similar to the 26–40% found using a regular AFLP protocol in white pine populations (8). The identified WRKY-AFLP markers will advance our understanding of gene organization and evolution of the WRKY transcription factor family in a conifer species. Our research approach demonstrates a comprehensive and high-quality census of the WRKY transcription factors encoded within the white pine genome. The results provide a solid foundation for further systematic characterization of PmWRKY transcription factors at the level of either single genes or gene families.

2. Materials

2.1. Plant Tissue Samples

1. *Tissue for genomic DNA extraction.* In this study white pine seedlings were grown in the greenhouse and needle samples were collected and stored at –20°C.
2. *Tissue for RNA extraction.* In this study, samples of needles, stems, and roots were collected from pine seedlings, frozen in liquid nitrogen immediately, and stored at –80°C.

2.2. Genomic DNA Extraction

1. DNeasy Plant Mini Kit (Qiagen, Mississauga, ON, Canada).
2. NanoDrop 1000 UV-Vis Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA).

3. Agarose gel electrophoresis reagents and apparatus for DNA analysis.
4. Gel Doc 2000 image analysis system (Bio-Rad, Philadelphia, PA, USA).

2.3. Total RNA Extraction

1. RNA extraction buffer: 0.1 M Tris–HCl, pH 7.4, 0.5 M NaCl, 50 mM EDTA, 2% SDS, 2% PVP-40, 10 mM β-mercaptoethanol (added freshly before use). Autoclaving is not necessary.
2. Potassium acetate (KOAc) 3 M solution, pH 5.5, treated with diethylpyrocarbonate (DEPC), then autoclaved.
3. Phenol/chloroform/isoamyl alcohol (IAA) (25:24:1).
4. Chloroform/IAA (24:1).
5. Lithium chloride (LiCl) 8 M and 2 M solutions, treated with DEPC, then autoclaved.
6. Ethanol (70%), diluted from RNase-free ethanol with DEPC-treated ddH_2O.
7. TE buffer: 10 mM Tris–HCl, pH 7.5, 1 mM EDTA. Prepare from RNase-free chemicals and ddH_2O.
8. DEPC-treated ddH_2O.
9. Agarose gel electrophoresis reagents and apparatus for RNA analysis.

2.4. cDNA Synthesis

1. RQ1 RNase-free DNase (Promega, Madison, WI, USA).
2. Plant RNeasy extraction kit (Qiagen).
3. SMART cDNA library construction kit (Clontech, Palo Alto, CA, USA).
4. RNase-free ddH_2O.

2.5. WRKY Polymorphism by a Modified AFLP Method

1. AFLP analysis system I kit (Invitrogen, Carlsbad, CA, USA).
2. Polyacrylamide gel electrophoresis (PAGE) apparatus: adjustable nucleic acid sequencer (CBS Scientific Co., Del Mar, CA, USA) and PowerPac HV power supply (Bio-Rad, Hercules, CA, USA).
3. TBE buffer (5X): 0.445 M Tris, 0.445 M boric acid, 0.01 M EDTA, pH 8.0. Dissolve 54 g Tris and 27.5 g boric acid and add 20 mL of 0.5 M EDTA (pH 8.0) to a total volume of 1 L.
4. Formamide dye: 98% formamide, 10 mM EDTA, 0.025% bromophenol blue, and 0.025% xylene cyanol.
5. 40% acrylamide–bisacrylamide solution (20:1) (take appropriate safety precautions with these neurotoxic chemicals).

6. *N*,*N*,*N*′,*N*′-tetramethylethylenediamine (TEMED) (Bio-Rad).
7. Ammonium persulfate (AP), 10% solution, freshly prepared for daily use although it may be stable for 1 week in darkness at 4°C.
8. Gel fixing solution: 10% (v/v) acetic acid.
9. Gel staining solution: 200 μL of 37% formaldehyde in 200 mL of 0.1% $AgNO_3$.
10. Gel developing solution: 200 μL of 37% formaldehyde and 20 μL of 2.0% (w/v) $Na_2S_2O_3 \cdot 5H_2O$ in 200 mL of 2.5% (w/v) Na_2CO_3.
11. Gel preserving solution: 20% ethanol, 20% isopropanol, and 10% glycerol.
12. Gel diffusion buffer: 0.5 M ammonium acetate, 1 mM EDTA, pH 8.0, 0.1% SDS.
13. QIAquick gel extraction kit (Qiagen).

2.6. PCR Cloning and Sequencing of Amplified DNA Fragments

1. Perkin-Elmer Thermocycler (Perkin-Elmer Applied Biosystems, Foster City, CA, USA).
2. Taq PCR Master Mix Kit (Qiagen).
3. Hot Start DNA Polymerase mix, dNTP solution (10 mM), Hot Start polymerase buffer.
4. PCR primers (10 μM): WRKY domain primers (**Table 2.2**), Clontech's CDS III/3′ PCR primer, Clontech's 5′ PCR primers, and Invitrogen's AFLP primers.
5. Reagents and apparatus for DNA electrophoresis.
6. MinElute PCR purification kit (Qiagen).
7. pGEM®-T Easy Vector System (Promega).
8. LB medium for bacterium culture.
9. *Escherichia coli* competent cells.
10. Isopropyl-beta-thiogalactopyranoside (IPTG) (Sigma), 0.1 M solution. Dissolve 0.238 g IPTG in 10 mL of water and sterilize by filtration. Store at –20°C.
11. 5-Bromo-4-chloro-indoly-β-D-galactoside (X-gal) (Sigma), 20 mg/mL solution in dimethylsulfoxide (DMSO) or dimethylformamide, but not in ddH_2O. Sterilize by filtration and wrap in foil for protection from light. Store at –20°C.
12. Plasmid mini kit (Qiagen).
13. Restriction enzymes and buffers (Invitrogen).

14. Reagents and apparatus for DNA sequencing.
15. Standard DNA ladders: 1 kb DNA ladder (Gibco BRL) and 25 bp DNA ladder (Invitrogen).

3. Methods

3.1. In Silico Data Mining and Design of PCR Primers

1. Select representatives of each WRKY subgroup (I, IIa+b, IIc, IId+e, and III) from genomes of *Arabidopsis* and rice for BLAST searches. Use a similar approach for other transcription factor families of interest.
2. Search EST databases of the Gene Index Project as in silico resources, provided by the Computational Biology and Functional Genomics Laboratory at the Dana-Farber Cancer Institute (DFCI) and Harvard School of Public Health (http://compbio.dfci.harvard.edu/tgi/tgipage.html).
3. Perform a tBlastn search through the database of the DFCI Gene Index Project online (http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/Blast/index.cgi). For this study the DFCI Pine Gene Index (Release 6.0, July 19, 2005, total 45,557 output sequences) or the DFCI Spruce Gene Index (Release 3.0, July 11, 2008, total 80,494 output sequences) was searched with the representative WRKY domain sequences of the five subgroups from *Arabidopsis* or rice.
4. Retrieve all WRKY-homologous sequences from the Gene Index Project databases and store them in a local computer.
5. Perform alignment analysis of nucleotide or putative amino acid sequences online with the Clustal W network service at the European Bioinformatics Institute (EBI, Cambridge, UK; http://www.ebi.ac.uk/Tools/clustalw2/index.html) and identify those sequences with identities of 98% or more as the same gene. (The WRKY genes identified by this in silico data mining method are listed in **Table 2.1** with 29 WRKY family members in *P. taeda* and 31 members in *Picea* species.)
6. Construct phylogenetic trees using the neighbor-joining (NJ) method in the MEGA4 software package (10) based on sequence alignment analysis of the WRKY domains from a variety of plant taxa for classification of WRKY genes into distinct clusters; here the reliability of each tree is established by conducting 1,000 neighbor-joining bootstrap sampling steps. The gene clusters in the phylogenetic tree correspond to distinct WRKY gene groups (I, IIa+b, IIc, IId+e, and III) (*see* **Note 1**).

7. Search for conserved amino acid motifs in the WRKY domains based on alignment analysis of WRKY domain sequences from both angiosperms and gymnosperms; design PCR primers according to the identified amino acid motifs as listed in **Table 2.2** (*see* **Note 2**).

3.2. Genomic DNA PCR for WRKY Gene Discovery

1. Extract genomic DNA from plant tissues (white pine needles in this study) using a DNeasy plant mini kit following the manufacturer's instructions.
2. Determine the quality and concentration of extracted genomic DNA using a NanoDrop 1000 UV-Vis Spectrophotometer and verify the spectrophotometric results by 0.8% agarose gel electrophoresis.
3. Adjust genomic DNA concentration to 10 ng/μL with ddH_2O or TE buffer and store the genomic DNA samples at –20°C until further analysis.
4. Perform genomic DNA PCR using a Taq PCR Master Mix Kit (*see* **Note 3**) with one forward primer (FP1 or FP2) in combination with each of seven reverse primers (RP1–RP7) as listed in **Table 2.2**.
5. Prepare a PCR reaction in a total volume of 25 μL with 12.5 μL of Taq PCR Master Mix (2X), 1 μL of each primer (10 μM), 2 μL of genomic DNA (10 ng/μL), and 8.5 μL of ddH_2O (*see* **Note 4**).
6. Run the PCR on a Perkin-Elmer Thermocycler with 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 42–60°C for 1 min, and primer extension at 72°C for 1.5 min, and a final 10-min extension at 72°C (*see* **Note 5**).
7. Continue to **Section 3.5** for genomic PCR product separation and vector cloning. An example of genomic DNA amplification from WRKY domain primers is shown in **Fig. 2.1a**. Genomic PCR fragments (or smear regions) larger than 160 bp are subjected to further PCR fragment purification and vector cloning (*see* **Note 6**).

3.3. RT-PCR and RACE for Expression Profiling of WRKY Genes

3.3.1. RNA Isolation

1. Grind the tissue sample (~1.0 g of white pine tissue in this study) into a fine powder in liquid nitrogen and transfer the tissue powder into 10 mL of RNA extraction buffer (*see* **Note 7**).
2. Shake the extraction mixture thoroughly and incubate it at 65°C for 20–30 min.
3. Remove plant cellular debris by centrifugation at 12,000×*g* for 15 min at room temperature.

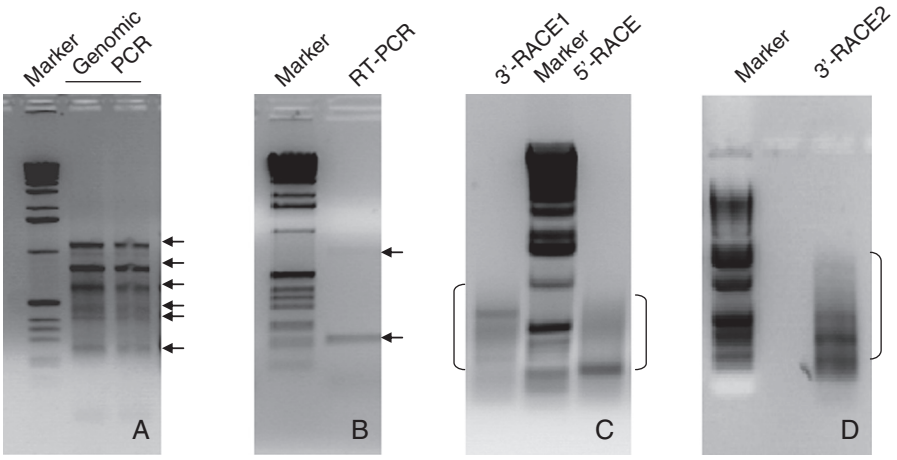

Fig. 2.1. Agarose gel electrophoresis of PCR products using WRKY domain primers. Marker standard 1 kb DNA ladder (Gibco BRL) is included in each gel. (**a)** Genomic DNA amplification using primers WRKY-FP1 and RP1. Six amplified fragments range from 0.25 to 1.15 kb and are shown by *arrows*. (**b)** cDNA fragments amplified by RT-PCR from total RNAs pooled from vegetative organs (shoots, stems, needles, and roots) using primers WRKY-FP1 and RP1. Two fragments are indicated by arrows with sizes about 0.16 and 0.70 kb. (**c)** PCR products from 5′RACE using WRKY-RP1 primer and Clontech's 5′ PCR primer and 3′RACE using WRKY-FP1 primer and Clontech's CDS III/3′ PCR primer with an annealing temperature of 45°C. (**d**) PCR products from 3′RACE using WRKY-FP2 primer and Clontech's CDS III/3′ PCR primer with an annealing temperature of 60°C.

4. Add 1/3 volume of 3 M KOAc, pH 5.5, to the supernatant and mix well; then incubate on ice for 30 min.
5. Remove the pellet with centrifugation at 12,000×*g* at 4°C for 10 min (*see* **Note 8**).
6. Extract the supernatant once with an equal volume of phenol/chloroform/IAA (25:24:1) and extract the supernatant once again with an equal volume of chloroform/IAA (24:1).
7. Precipitate total RNA from the supernatant by adding 1/3 volume of 8 M LiCl and incubate overnight at 4°C (*see* **Note 9**).
8. Recover total RNA by centrifugation at 12,000×*g* at 4°C for 30 min.
9. Wash total RNA pellet once with 2 M LiCl and then wash it once again with 70% ethanol.
10. Air-dry the RNA pellet and re-suspend it in 50 μL of TE buffer or DEPC-treated ddH_2O, then store RNA samples at –80°C until further analysis.
11. Measure RNA concentration by UV spectrophotometry for each individual tissue sample, verify RNA integrity by

agarose gel electrophoresis, and visualize the gel by ethidium bromide staining.

3.3.2. cDNA Synthesis

A common strategy for plant gene discovery is reverse transcription of plant messenger RNA into cDNA, followed by construction of a plant cDNA library or PCR cloning of targeted cDNA using gene-specific primers. A cap-switching approach is widely used for generating cDNA with high potential for full-length cDNA cloning from a small amount of mRNA or total RNA. The first strand (ss) cDNA is synthesized using a modified oligo d(T) primer. As the newly synthesized ss-cDNA extends to the 5′-cap structure at the end of the mRNA templates, a short non-template sequence, called a cap-switch oligonucleotide, is integrated to the cDNA 3′-end by the terminal transferase activity of the MMLV reverse transcriptase (17, 18). Therefore, additional sequences are attached at both 3′- and 5′-ends of the synthesized ss-cDNA that can be used for PCR primers and for the generation of double strand (ds) cDNA by PCR amplification. This cap-switching technology is very useful for cDNA library construction as well as for cDNA template production in targeted gene cloning by RT-PCR or RACE and is utilized in a variety of commercial kits. The SMART^{TM} (*s*witching *m*echanism *a*t 5′ end of *R*NA *t*emplate) cDNA library construction kit (Clontech) is the one we use here.

1. Treat total RNA samples with RQ1 RNase-Free DNase to eliminate any contaminating genomic DNA, following the manufacturer's instructions.
2. Remove the enzyme from RNA samples by RNA re-purification using a Plant RNeasy extraction kit, following the manufacturer's instructions.
3. Verify total RNA integrity by agarose gel electrophoresis and measure RNA concentration using a NanoDrop 1000 UV-Vis Spectrophotometer.
4. Reverse-transcribe total RNA templates into ss-cDNA and synthesize ds-cDNA using the SMART cDNA library construction kit, following the manufacturer's instructions (*see* **Note 10**).
5. Dilute both ss-cDNA and ds-cDNA reactions appropriately and store them at –20°C for use as templates in RT-PCR and RACE.

3.3.3. RT-PCR for Amplification of WRKY Domain Sequences

The procedures are similar to genomic DNA PCR for *WRKY* discovery described in steps 4–6 of **Section 3.2**, but the PCR templates are replaced with diluted ss-cDNA or ds-cDNA. An example of RT-PCR amplification of two expressed WRKY domain fragments is shown in **Fig. 2.1b**. Usually one (~0.16 kb) or two

cDNA fragments (~0.16 and ~0.70 kb) are amplified using each of the 14 combinations of the WRKY domain primers listed in **Table 2.2**. The two RT-PCR fragments, or smear regions with similar sizes, are subjected to further PCR fragment purification and vector cloning (*see* **Note 11**).

3.3.4. RACE for cDNA Cloning of Full-Length Coding Regions

1. Mix the following reagents in a sterile 0.2 mL tube: 5 μL of Hot Start polymerase buffer (10X), 1 μL of dNTP solution (each 10 mM), 2.5 μL of WRKY primer (FP1 or FP2 for 3′RACE or one of RP0–RP7 for 5′RACE), 1 μL of one of Clontech's primers (CDS III 3′ PCR primer for 3′RACE or 5′ PCR primer for 5′RACE), 2.5 units of Hot Start polymerase mix, 1 μL of diluted ss-cDNA or ds-cDNA (from step 5 of **Section 3.3.2**), and ddH_2O to a total volume of 50 μL (*see* **Note 12**).
2. Heat the mixture in a thermocycler at 95°C for 3 min, then run 30 cycles of RACE using a PCR program as follows: template denaturation at 95°C for 20 s, primer annealing at 45–68°C for 1 min, and primer extension at 72°C for 3 min (*see* **Note 13**).
3. Continue to **Section 3.5** for RACE product separation and vector cloning. Examples of RACE using WRKY domain primers are shown in **Fig. 2.1c**, **d**. The RACE products are usually in a continuous size range, so at least a few hundred recombinant plasmids are necessarily selected for DNA sequencing. A sampling of recombinant clones from one RACE experiment is shown in **Fig. 2.2.**

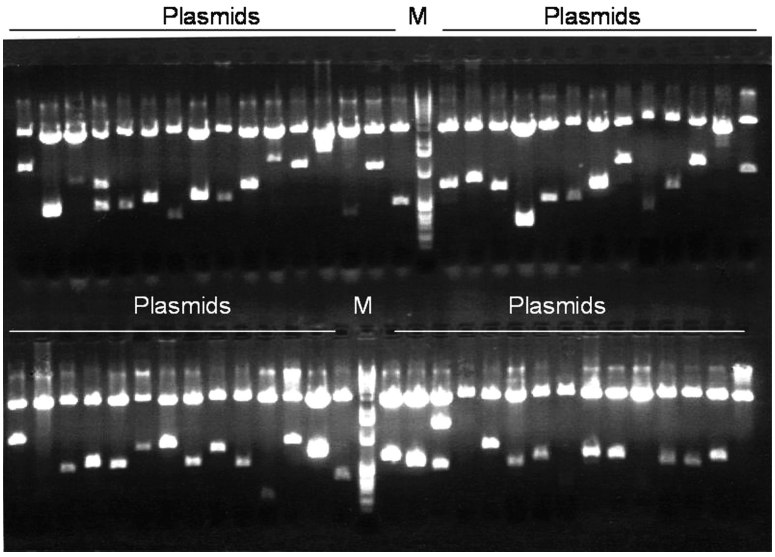

Fig. 2.2. Restriction enzyme analysis of the recombinant plasmids. 3′RACE products as shown in **Fig. 2.1d** were ligated into the pGEM-T Easy Vector (Promega). Plasmids were digested with *EcoR*I to show insert size variation.

3.4. A Modified AFLP for WRKY Polymorphism Detection

To detect DNA polymorphism and map WRKY genes, a modified AFLP method was developed based on a method described previously (8, 13). In this method, the WRKY-AFLP markers are first detected by PAGE followed by silver staining. The polymorphic DNA fragments of interest are then extracted, re-amplified, and sequenced to confirm their gene identities.

3.4.1. A Modified AFLP Protocol

Genomic DNA digestion and adapter ligation are performed using the AFLP analysis system I kit or another commercial AFLP kit following the manufacturer's instructions.

1. Add the genomic DNA (250 ng) with 5 μL of reaction buffer (5X), 2 μL of *Eco*RI/*Mse*I, and ddH_2O to a final volume of 25 μL.
2. Incubate at 37°C for 4 h, then at 70°C for 15 min to inactivate restriction enzymes.
3. Add 24 μL of adapter ligation solution and 1 μL of T4 DNA ligase to the 25 μL of genomic DNA digestion mixture for adapter ligation.
4. Incubate adapter ligation mixture at room temperature (20–22°C) for 2 h.
5. Dilute the ligation mix in a 1:10 ratio as follows: add 10 μL of ligation mix to 90 μL of TE buffer.
6. Store diluted and undiluted ligation mixes at –20°C.
7. Prepare AFLP pre-amplification reactions in 0.2 mL PCR tubes by adding the following: 5 μL of diluted template genomic DNA, 5 μL of PCR buffer plus Mg^{++} (10X), 2.5 μL of *Eco*RI-AC primer (10 μM), 2.5 μL of *Mse*I-CC primer (10 μM), 5 μL of dNTPs (each 2 mM), 1 μL of Taq DNA polymerase (5U/μL), and ddH_2O to a total volume of 50 μL (*see* **Note 14**).
8. Run AFLP pre-amplification with the following conditions: pre-denaturation at 94°C for 3 min, followed by 40 cycles of denaturation at 94°C for 30 s, annealing at 56°C for 60 s, and primer extension at 72°C for 60 s with a final extension at 72°C for 7 min.
9. Dilute pre-amplification PCR reactions to a ratio of 1:50 or 1:100 and store both diluted and undiluted reactions at –20°C until further analysis.
10. Perform WRKY-AFLP profiling by selective AFLP amplification reactions: prepare the reaction mixtures in 0.2 mL PCR tubes by adding 1 μL of the diluted pre-amplification reaction, 12.5 μL of 2X Taq Master Mix, 1 μL of *Eco*RI-ACNN or *Mse*I-CCNN primer (10 μM), 1 μL of WRKY primers (10 μM), and ddH_2O to a total volume of 25 μL (*see* **Notes 14** and **15**).

11. Run AFLP selective amplification under the following PCR conditions: pre-denaturation at 94°C for 3 min, followed by 30 cycles of denaturation at 94°C for 30 s, annealing at 56°C for 60 s, and primer extension at 72°C for 60 s with a final extension at 72°C for 7 min (*see* **Note 16**).

3.4.2. PAGE Separation of DNA Fragments

An adjustable nucleic acid sequencer (CBS Scientific) is used for PAGE analysis here. Advanced DNA sequencer systems can be used for high-throughput AFLP profiling.

1. Prepare a 6% polyacrylamide gel (0.75 mm thick, size 30 × 20 cm with 60 wells) with 60 mL of gel solution by adding the following: 27 g urea, 9 mL of 40% acrylamide–bisacrylamide stock solution, 6 mL of 5X TBE, 15 mL of ddH_2O, 0.5 mL of 10% APS, and 0.03 mL of TEMED.
2. Set up the adjustable nucleic acid sequencer apparatus and add 800 mL of 0.5X TBE buffer to the upper and lower chambers of the gel unit.
3. Run pre-electrophoresis for 20–30 min at 500–700 V for stabilization of the gel temperature at 50°C.
4. Prepare DNA samples while the gel is pre-running by adding equal volumes (25 μL) of formamide dye to each selective PCR reaction and mix well.
5. Denature DNA samples by incubation at 94°C for 3 min and keep the denatured DNA samples on ice until sample loading on the gel.
6. Load 10 μL of the denatured DNA samples onto the gel including one well at both sides with 50 ng of a 25 bp DNA standard for estimation of the amplified DNA fragment sizes.
7. Run AFLP electrophoresis with 0.5X TBE buffer until the xylene cyanol band is within 2–3 cm of the bottom edge of the gel.

3.4.3. PAGE Silver Staining

1. Fix the gel for 30 min by adding 200 mL of 10% acetic acid (v/v) (gel fixing solution) for each gel.
2. Rinse the gel in ddH_2O three times, each for 2 min; then stain the gel with 200 mL of gel staining solution for 20–30 min.
3. Rinse the gel on both surfaces with ddH_2O for 10–20 s.
4. Incubate the gel at room temperature in 200 mL of pre-cooled (4°C) gel developing solution until DNA bands are clearly visible (2–10 min).
5. Stop the developing reaction by incubating the gel in 10% acetic acid solution for 2–3 min, then rinse the gel with ddH_2O.

6. Scan the AFLP profile and save the image into a computer using the Quantity One software package (Bio-Rad). An example of a resulting WRKY-AFLP profile is shown in **Fig. 2.3**.
7. Incubate the gel in gel preserving solution overnight with slight shaking; then dry the gel for long-term storage if necessary.

3.4.4. Elution of DNA Fragments from Polyacrylamide Gel

Following treatment of a gel with gel diffusion buffer, a QIAquick gel extraction kit is used to elute DNA fragments from the polyacrylamide gel, following the manufacturer's instructions.

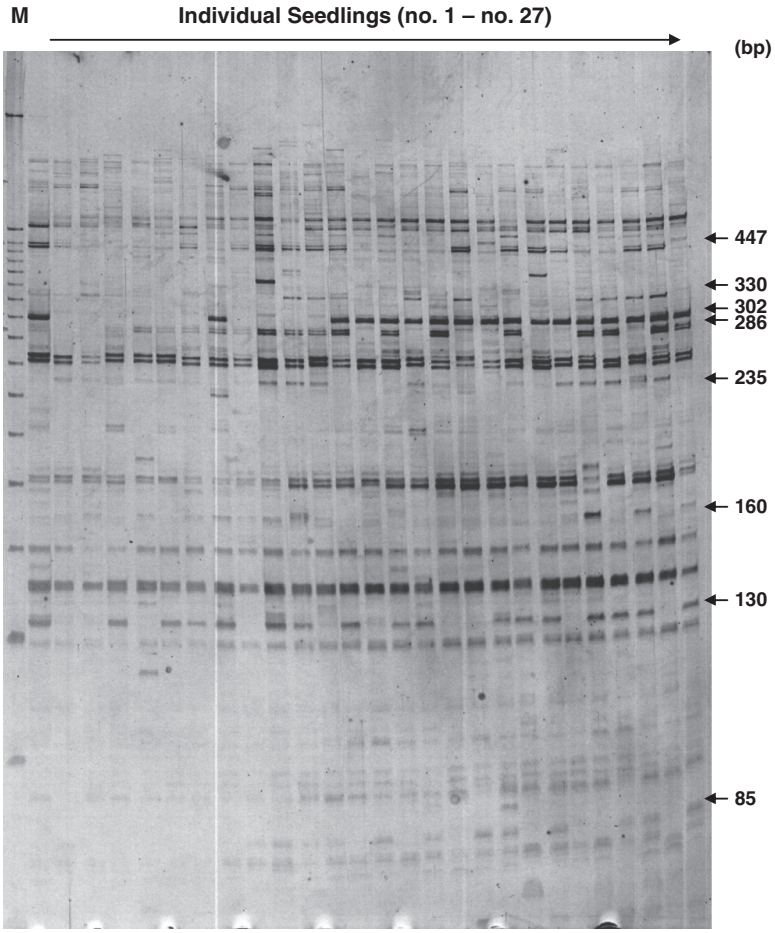

Fig. 2.3. An example of the white pine WRKY-AFLP polymorphic profiles. Marker standard 25 bp DNA ladder (M) is included in the left well. AFLP selective amplification was performed using one WRKY primer (RP0) coupled with one AFLP selective primer (*EcoR*I-ACTC). Eight WRKY-AFLP markers ranging from 85 to 447 bp are indicated by *arrows*. Amplicons of 27 individual seedlings (no. 1–no. 27) from a seed family are shown here.

1. Excise the gel slice containing the DNA band of interest from the polyacrylamide gel.
2. Weigh the gel slice and add 1–2 volumes (v/w) of gel diffusion buffer into the gel slice tube.
3. Incubate at 50°C for 30 min and centrifuge the sample tube at 12,000×*g* for 1 min.
4. Transfer the supernatant to a new tube and process it immediately with a QIAquick gel extraction kit.
5. Add three volumes of QG buffer to the supernatant and ensure the color of the mixture is yellow, which indicates the correct pH.
6. Apply the sample mixture to a QIAquick spin column for a quick spin (30–60 s). The DNA fragment is now bound to the QIAquick spin column. Wash the column with 0.75 mL of PE buffer and remove the PE buffer trace by centrifugation at 12,000×g for 2 min.
7. Add 50 μL of EB buffer to the column for a quick spin to elute DNA.
8. Re-amplify the eluted DNA fragment using the same pair of primers and the same PCR conditions used for WRKY-AFLP profiling.
9. Verify the size of the re-amplified DNA fragment with PAGE.
10. Continue to **Section 3.5** for DNA fragment separation and vector cloning.

3.5. Cloning, Sequencing, and Identity Confirmation of DNA Fragments

Cloning of DNA fragments amplified from genomic DNA or cDNA is necessary to confirm gene identity by DNA sequencing and sequence data analysis. There are a variety of commercial kits available for DNA fragment purification from agarose gel, PCR cloning, and subsequent DNA sequencing. Researchers can choose their own preferred kits and methods. We use Qiagen's gel extraction kit for DNA fragment purification and the pGEM-T Easy Vector for PCR cloning.

1. Perform agarose gel electrophoresis to separate PCR products from genomic DNA amplification, RT-PCR, RACE, or WRKY-AFLP.
2. Visualize the electrophoresis profiles of PCR fragments by ethidium bromide staining and record PCR profiles digitally using a Gel Doc 2000 image analysis system.
3. Excise agarose gel slices containing the DNA fragments of interest under UV light (*see* **Note 17**).
4. Extract each DNA fragment from the agarose gel slices using a gel extraction kit, such as Qiagen's QIAquick gel

extraction kit or the MinElute PCR purification kit (*see* **Note 18**), following the manufacturer's instructions.

5. Ligate each purified DNA fragment (or smear region) into Promega's pGEM-T Easy Vector at room temperature or 4°C (*see* **Note 19**), following the manufacturer's instructions.
6. Transform the ligation mixture into appropriate *E. coli* competent cells using white-blue selection with β-galactosidase under X-gal and IPTG in the LB-agar medium.
7. Select white *E. coli* colonies, grow overnight, and extract plasmids using a Plasmid mini kit.
8. Digest plasmids with *EcoR*I to verify insert sizes of recombinant clones using agarose gel electrophoresis.
9. Sequence inserts of recombinant plasmids with vector primers (T7 and SP6) and additional gene-specific primers if long inserts are present in the recombinant plasmids (*see* **Fig. 2.2**).
10. Analyze the DNA sequence data and annotate WRKY genes as described in **Section 3.1**.

4. Notes

1. This gene classification based on phylogenetic analysis is consistent with other group-specific features of the WRKY family, such as intron positions and phases when genomic sequences are available.
2. The PCR forward primers FP1 and FP2, targeted at the N-terminal region of the WRKY domain, are universal for both angiosperms and gymnosperms. In contrast, the C-terminal regions of the WRKY domains are less conserved, with group specificity or even gene specificity. Numerous reverse primers are required to cover a range of potential members of the WRKY family in a non-model plant. Reverse primers RP0–RP7 in **Table 2.2** are only a few oligonucleotides used in white pine. Other reverse primers with more diverse amino acid sequences may be needed to target the C-terminal regions of the WRKY domains in other non-model plants.
3. Researchers may wish to use another PCR kit or other thermostable DNA polymerases for genomic DNA amplification. For example, Hot Start DNA polymerase helps eliminate non-specific DNA amplification. *Pfu* DNA polymerase

is a better choice if nucleotide misincorporation is a concern in PCR cloning. *Tth* DNA polymerase may improve PCR results if there is a problem resulting from a high degree of secondary structure in the templates.

4. Because white pine has one of the largest genomes (~28.25 pg/1C), we used 20–50 ng of genomic DNA for WRKY gene amplification. The amount of genomic DNA template is adjustable, depending on the genome sizes of plant species, copy numbers of the targeted gene, and quality of the purified DNA samples.
5. Annealing temperature (*T*m) could be adjusted from 42 to 60°C in the case of white pine with related primer combinations. A higher *T*m will not only decrease non-specific amplification but also decrease the cloning coverage of the gene family. The primer extension time may be longer if a large intron (for example, 1 kb or longer) is predicted to be in the WRKY domain sequence.
6. An intron is predicted in the genomic DNA sequences of the WRKY domains. Because of the uncertainty regarding the size of this WRKY domain intron in uncharacterized plant genomes, we suggest collecting all amplified DNA fragments larger than 160 bp (a predicted size for WRKY domain mRNA) for genomic DNA cloning of the WRKY domains. For each genomic DNA fragment, multiple recombinant clones are needed for DNA sequencing to check gene diversity within a specific fragment because each DNA band probably contains PCR products amplified from multiple genes with very similar sizes.
7. No commercial RNA extraction kit is suitable for extraction of RNA from conifers. The method of conifer RNA extraction we describe here is adapted from a protocol reported previously (19, 20) and is effective for total RNA extraction from various conifer tissues. At this step, a high tissue-buffer ratio (w/v=1:10) helps eliminate polysaccharides in subsequent steps.
8. Many types of plant samples are rich in polysaccharides and polyphenols that are easily co-purified with RNA; this can seriously reduce RNA quality. Before the final RNA precipitation with 8 M LiCl or ethanol, this step of using 1/3 volume of 3 M KOAc, pH 5.5, is crucial to precipitate the sugars, although some RNA may be also lost. As an indicator of the increased purity, the RNA sample should have a higher ratio of A_{260}/A_{280}, which far outweighs the loss of RNA to the polysaccharide plugs.
9. If RNA extraction must be finished in 1 day, total RNA could be precipitated optionally by adding three volumes of

95% ethanol. Contaminating genomic DNA in total RNA usually is not a serious problem because each total RNA sample will be treated by RNase-free DNase before reverse transcription for cDNA synthesis. However, alcohol precipitation may result in co-purification of polysaccharides and polyphenolics along with RNA and genomic DNA, especially for plant tissues rich in secondary metabolites.

10. Because only a few cDNA syntheses are available from this kit, a mixture of total RNA pooled from multiple tissue samples is preferably used as the reverse-transcription template. In our study, total RNA was pooled equally from different tissues (roots, stems, needles, and immature male and female cones) or needle tissues with different treatments (wounding, disease infection, etc.) for optimal discovery of the expressed WRKY genes.
11. Even if an RT-PCR fragment is very sharp on an agarose gel, it is necessary to select at least 10 recombinant clones for DNA sequencing. If DNA sequencing reveals high gene diversity in a particular DNA fragment, a larger number of recombinant clones from the fragment should be sequenced in case it contains additional gene members (*see* **Note 6**).
12. When a ds-cDNA mixture is used as a RACE template, only 0.1–1.0 ng of ds-cDNA is added in an amplification reaction. In addition, Hot Start DNA polymerase is preferred for RACE because there is non-specific amplification in most cases (*see* **Note 3**). To increase the potential of cloning multiple genes simultaneously, we used degenerate WRKY domain primers (**Table 2.2**) to perform 3′RACE PCR. Once the sequence of the 3′-untranslated region is determined, a gene-specific primer could be designed for 5′RACE to clone a specific WRKY gene. A special commercial RACE kit may be used for this work.
13. A higher annealing temperature is recommended, especially for 5′RACE using WRKY reverse primers with high degeneracy, which easily cause non-specific amplification in RACE (*see* **Note 5**).
14. There are two options to modify the standard AFLP procedure for WRKY-AFLP profiling. One option is to modify AFLP at the step of AFLP selective amplification. Another option is to modify AFLP at the step of AFLP preamplification where one WRKY primer and one standard AFLP primer (*Eco*RI-AC or *Mse*I-CC) are used. Then, the same WRKY primer and one of 32 standard AFLP primers (*Eco*RI-ACNN or *Mse*I-CCNN) should be used at the next step for AFLP selective amplification.

15. When a WRKY primer is considered for WRKY-AFLP profiling, it will be designed to have an annealing temperature consistent with that of regular AFLP primers. In addition, for each of the 10 WRKY primers listed in **Table 2.2**, there are 32 primer combinations from standard AFLP selective primers (16 *Eco*RI-ACNN and 16 *Mse*I-CCNN). A selection of 320 primer combinations provides the WRKY-AFLP profiling with a high potential to discover WRKY polymorphism in a population.
16. The number of cycles in the AFLP selective amplification step may be adjusted according to the final dilution of pre-amplified products and at which step (pre-amplification or selective amplification) the WRKY primer is integrated into the AFLP protocol.
17. Ensure exposure to ultraviolet (UV) light is as brief as possible when DNA fragments are cut out from an agarose gel. The PCR product is preferably visualized only with a long-wave UV source. Otherwise, the efficiency of DNA cloning can decrease greatly because of UV-caused DNA damage.
18. There are various types of commercial kits for DNA fragment purification from agarose gel. Qiagen's MinElute PCR purification kit uses a small volume (10 μL) to elute a DNA fragment from a column at a relatively high concentration, which potentially increases efficiency of subsequent DNA ligation and cloning.
19. DNA ligation buffer (2X) contains ATP that might be degraded during temperature fluctuations. To avoid this, make single-use aliquots of the ligation buffer. Incubation of DNA ligation reactions overnight at 4°C often produces the maximum number of transformants.

Acknowledgments

The authors would like to thank Steve Glover for his critical review on the manuscript. This work was supported in part by Natural Resources Canada, the Canadian Forest Service, and Agriculture and Agri-Food Canada, the Pacific Agri-Food Research Centre.

References

1. Eulgem, T., Rushton, P. J., Robatzek, S., and Somssich, I. E. (2000) The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* **5**, 199–206.
2. Rushton, P. J., Torres, J. T., Parniske, M., Wernert, P., Hahlbrock, K., and Somssich, I. E. (1996) Interaction of elicitor-induced DNA-binding proteins with elicitor response

elements in the promoters of parsley PR1 genes. *EMBO J.* **15**(**20**), 5690–5700.
3. Du, L. and Chen, Z. (2000) Identification of genes encoding receptor-like protein kinases as possible targets of pathogen- and salicylic acid-induced WRKY DNA-binding proteins in *Arabidopsis*. *Plant J.* **24**(**6**), 837–847.
4. Ülker, B. and Somssich, I. E. (2004) WRKY transcription factors: from DNA binding towards biological function. *Curr. Opin. Plant Biol.* **7**, 491–498.
5. Wu, K. L., Guo, Z. J., Wang, H. H., and Li, J. (2005) The WRKY family of transcription factors in rice and *Arabidopsis* and their origins. *DNA Res.* **12**, 9–26.
6. Zhang, Y. and Wang, L. (2005) The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants. *BMC Evol. Biol.* **5**, 1.
7. Berri, S., Abbruscato, P., Faivre-Rampant, O., Brasileiro, A. C. M., Fumasoni, I., Satoh, K., Kikuchi, S., Mizzi, L., Morandini, P., Pè, M. E., and Piffanelli, P. (2009) Characterization of WRKY co-regulatory networks in rice and Arabidopsis. *BMC Plant Biol.* **9**, 120.
8. Liu, J.-J. and Ekramoddoullah, A. K. M. (2009) Identification and characterization of the WRKY transcription factor family in *Pinus monticola*. *Genome* **52**, 77–88.
9. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1999) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.
10. Kumar, S., Tamura, K., and Nei, M. (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* **5**, 150–163.
11. Borrone, J. W., Kuhn, D. N., and Schnell, R. J. (2004) Isolation, characterization, and development of WRKY genes as useful genetic markers in *Theobroma cacao*. *Theor. Appl. Genet.* **109**, 495–507.
12. Trognitz, F., Manosalva, P., Gysin, R., Niño-Liu, D., Simon, R., del Herrera, M. R., Trognitz, B., Ghislain, M., and Nelson, R. (2002) Plant defense genes associated with quantitative resistance to potato late blight in *Solanum phureja* × Dihaploid *S. tuberosum* hybrids. *Mol. Plant Microbe Interact.* **15**, 587–597.
13. Hayes, A. J. and Maroof, M. A. (2000) Targeted resistance gene mapping in soybean using modified AFLPs. *Theor. Appl. Genet.* **100**, 1279–1283.
14. 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**(**6**), 728–763.
15. Liu, J.-J. and Ekramoddoullah, A. K. M. (2008) Development of LRR polymorphism, AFLP, and SCAR markers to the *Cronartium ribicola* resistance gene *Cr2* in western white pine (*Pinus monticola*). *Tree Genet. Genomes* **4**, 601–610.
16. Teunissen, H. A., Rep, M., Houterman, P. M., Cornelissen, B. J., and Haring, M. A. (2003) Construction of a mitotic linkage map of *Fusarium oxysporum* based on Foxy-AFLPs. *Mol. Genet. Genomes* **269**, 215–226.
17. Schmidt, W. M. and Mueller, M. W. (1999) CapSelect: a highly sensitive method for 5′cap-dependent enrichment of full-length cDNA in PCR-mediated analysis of mRNAs. *Nucleic Acids Res.* **27**, e31.
18. Schramm, G., Bruchhaus, I., and Roeder, T. (2000) A simple and reliable 5′-RACE approach. *Nucleic Acids Res.* **28**, e96.
19. Liu, J.-J., Goh, C.-J., Loh, C.-S., Liu, P., and Pua, E.-C. (1998) A method for isolation of total RNA from fruit tissues of banana. *Plant Mol. Biol. Rep.* **16**, 87a–f.
20. Liu, J.-J., Ekramoddoullah, A. K. M., and Yu, X. (2003) Differential expression of multiple PR10 proteins in western white pine following wounding, fungal infection and cold-hardening. *Physiol. Plant* **119**, 544–553.

http://www.springer.com/978-1-61779-153-6

Plant Transcription Factors
Methods and Protocols
Yuan, L.; Perry, S.E. (Eds.)
2011, XI, 347 p., Hardcover
ISBN: 978-1-61779-153-6
A product of Humana Press