

## Research Article

# The WRKY Transcription Factor Genes in *Lotus japonicus*

Hui Song,<sup>1</sup> Pengfei Wang,<sup>2</sup> Zhibiao Nan,<sup>1</sup> and Xingjun Wang<sup>2</sup>

<sup>1</sup> College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou 730000, China

<sup>2</sup> Bio-Tech Research Center, Shandong Academy of Agricultural Sciences, Shandong Provincial Key Laboratory of Crop Genetic Improvement, Ecology and Physiology, Jinan 250100, China

Correspondence should be addressed to Zhibiao Nan; zhibiao@lzu.edu.cn and Xingjun Wang; xingjunw@hotmail.com

Received 26 September 2013; Revised 15 December 2013; Accepted 27 January 2014; Published 16 March 2014

Academic Editor: John Parkinson

Copyright © 2014 Hui Song et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

WRKY transcription factor genes play critical roles in plant growth and development, as well as stress responses. WRKY genes have been examined in various higher plants, but they have not been characterized in *Lotus japonicus*. The recent release of the *L. japonicus* whole genome sequence provides an opportunity for a genome wide analysis of WRKY genes in this species. In this study, we identified 61 WRKY genes in the *L. japonicus* genome. Based on the WRKY protein structure, *L. japonicus* WRKY (LjWRKY) genes can be classified into three groups (I–III). Investigations of gene copy number and gene clusters indicate that only one gene duplication event occurred on chromosome 4 and no clustered genes were detected on chromosomes 3 or 6. Researchers previously believed that group II and III WRKY domains were derived from the C-terminal WRKY domain of group I. Our results suggest that some WRKY genes in group II originated from the N-terminal domain of group I WRKY genes. Additional evidence to support this hypothesis was obtained by *Medicago truncatula* WRKY (MtWRKY) protein motif analysis. We found that LjWRKY and MtWRKY group III genes are under purifying selection, suggesting that WRKY genes will become increasingly structured and functionally conserved.

## 1. Introduction

Transcription factors are crucial in regulating gene expression. Transcription factors present sequence-specific DNA binding sites and are able to modulate the transcription rate of downstream target genes [1]. WRKY genes have primarily been located in plants, where they are one of the most important transcription factor families [2]. WRKY genes are defined by having a unique WRKY domain of approximately 60 amino acid residues [2]. The WRKY domain contains a highly conserved amino acid sequence WRKYGQK at the N-terminal and a metal chelating zinc finger motif (C-X<sub>4-5</sub>-C-X<sub>22-23</sub>-H-X-H, (C<sub>2</sub>H<sub>2</sub>) or C-X<sub>5-8</sub>-C-X<sub>25-28</sub>-H-X<sub>1-2</sub>-C, (C<sub>2</sub>HXC)) at the C-terminal end [2–5]. In some WRKY genes, the WRKY domain can be characterized as WRRY, WSKY, WKRY, WVKY, or WKKY [6]. WRKY transcription factors interact with the W-box (TTGAC[T/C]) sequence in promoter regions to modulate gene expression [4, 7, 8]. In addition, WRKY transcription factors bind SURE, a novel *cis*-element in higher plants, to regulate sugar response [9].

Members of the WRKY family can be classified into three groups according to the number of WRKY domains and the pattern of the zinc finger motif [2]. Generally, group I WRKY transcription factors contain two WRKY domains with distinct functions. Previous studies have demonstrated that the C-terminal WRKY domain mediates sequence-specific binding to the target DNA [8, 10, 11]. It has been proposed that the N-terminal WRKY domain increases the affinity or specificity of these proteins to the target sites [2]. Group II and III WRKY transcription factors contain one WRKY domain with a C<sub>2</sub>H<sub>2</sub> zinc finger motif and C<sub>2</sub>HXC zinc finger motif [2]. Based on a phylogenetic analysis of the WRKY family, the members of group II can be divided into five subgroups: IIa, IIb, IIc, IId, and IIe [2].

Since the cloning a WRKY gene cDNA from *Ipomoea batatas* [11], a large number of WRKY protein genes have been cloned from different plant species [3, 4, 12–24]. So far, only two WRKY homologues have been identified from nonplant species, *Giardia lamblia* [5] and *Dictyostelium discoideum* [25].

Plant WRKY genes regulate plant growth and development under normal and stressful conditions [26, 27]. Early studies found that WRKY genes play an important role in gene expression responses to sucrose [11]. Studies in *Arabidopsis* [17, 28, 29], rice [30], tobacco [31, 32], and parsley [33] have indicated that WRKY proteins play key roles in plant responses to pathogens [27, 34]. In addition, previous studies revealed the involvement of WRKY proteins in abiotic stress responses [27], for example, to high temperatures [35], low temperatures [13], salt and drought [24], H<sub>2</sub>O<sub>2</sub> [36], and UV radiation [37]. WRKY proteins have also been reported to upregulate in response to herbivory [38], nematode damage [39], and wounding [40]. Moreover, WRKY genes may be involved in seed development [9, 41], dormancy and germination [42–44], plant senescence [45, 46] and regulation of metabolic pathways [9], trichome morphogenesis [47], and plant growth [48].

*Lotus japonicus*, an important forage crop in the legume family, is planted in many parts of the world. It has been used extensively in plant research as a model legume, due to its short life cycle (2–3 months), self-fertility, and relatively simple diploid genome [49]. Since the release of the whole genome sequence of *L. japonicus*, it is now possible to compare transcription factors in this plant with other plants. In this study, we analyzed 61 putative WRKY genes from the *L. japonicus* genome. We conducted a phylogenetic analysis to evaluate gene duplications, chromosomal localization, motif analysis, gene structure, and selection pressure analysis of group III WRKY genes to provide information about WRKY gene family evolution in *L. japonicus*.

## 2. Materials and Methods

**2.1. Sequence Database Search.** The *L. japonicus* genome sequence (build 2.5) was downloaded from <http://www.kazusa.or.jp/lotus/>. The complete set of WRKY gene sequences was identified using a deliberative process. First, a Hidden Markov Model (HMM) profile of the WRKY domain (PF03106) was downloaded from the Pfam database (<http://pfam.sanger.ac.uk/>) [50]. We employed the WRKY domain as a query to identify all possible WRKY gene sequences in the *L. japonicus* genome database using the BLASTp program ( $P$  value = 0.001). Subsequently, a search on the Pfam database was used to confirm and classify each putative WRKY sequence. We located overlapping genes by aligning all of the candidate WRKY gene sequences using Clustal W [51]. Only the nonoverlapping WRKY sequences were used for further analysis.

**2.2. Gene Duplication and Chromosomal Locations of WRKY Genes.** To detect potential gene duplications, we aligned and calculated all of the relevant genes identified in *L. japonicus* genomes. We defined gene duplication between any two loci such that [52]: (1) the alignable nucleotide sequence covered >70% of the longer sequence; (2) the amino acid identity between the sequences was >70% identical.

In order to determine the physical locations of WRKY genes in chromosomes, we blasted each WRKY gene as

a query against the *L. japonicus* genome (<http://www.kazusa.or.jp/lotus/>) to determine the initiation site of each gene. MapInspect software was used to draw the location images of the WRKY genes ([http://www.plantbreeding.wur.nl/uk/software\\_mapinspect.html](http://www.plantbreeding.wur.nl/uk/software_mapinspect.html)).

**2.3. Multiple Sequence Alignments, Phylogenetic, and Gene Structure Analysis.** To examine the domain organization of WRKY proteins in detail, multiple sequence alignments of WRKY domain sequences were performed using Clustal W [51]. Phylogenetic and molecular evolution analyses of WRKY proteins were conducted in MEGA 4.0 [53]. Phylogenetic trees were estimated with the Neighbor-Joining method. Bootstrap analyses of 1,000 repetitions were obtained for each tree to analyze statistical support for nodes. The complete amino acid sequences of group III WRKY genes were analyzed for evidence of selection pressure [54]. Gene structure display server (GSDS) software [55] was used to illustrate exon-intron organization for individual WRKY genes by comparing the cDNA sequence with the corresponding genomic DNA sequence.

**2.4. Identification of Conserved Motifs.** The program MEME 4.9 [56] was used to predict motifs in the WRKY domain with the following parameters: (1) any number of repetitions, (2) an optimum motif width between 6 and 200 residues, (3) and a maximum of 20 motifs. Structural motif annotation was performed using the Pfam database.

**2.5. Selection Pressure in Group III WRKY Proteins.** The amino acid sequences of group III *Medicago truncatula* (MtWRKY) and *L. japonicus* WRKY (LjWRKY) proteins were used to estimate phylogenetic trees and then the trees were used to detect evidence of selection. The PAL2NAL program [57] was used for conversion of a protein sequence into the corresponding nucleotide sequence. We used PAML 4.7 [58] to analyze codon substitution patterns in a maximum likelihood framework, implementing a site-specific model. The program CODEML was employed to calculate the  $d_N/d_S$  ratio (or  $\omega$ ), the ratio of nonsynonymous/synonymous distances. Generally,  $\omega = 1$ ,  $>1$ , and  $<1$  indicate neutral, positive, and purifying selection, respectively. We detected variation in  $\omega$  among sites by employing a likelihood ratio test (LRT) between M0 versus M3 and M7 versus M8 [54]. The nodes were considered to have undergone positive selection, if they satisfied the following criteria [59]: (1) an estimate of  $\omega > 1$  under M8, (2) sites identified to be under positive selection by Bayes Empirical Bayes (BEB) analysis, (3) and a statistically significant LRT.

## 3. Results

**3.1. Identification and Classification of WRKY Genes.** In this study, a total of 71 WRKY genes in the *L. japonicus* genome (build 2.5) were identified (Table 1). Among these sequences, 10 WRKY genes were excluded from this study due to the divergent structures of the putative proteins in these genes, a lack of specific domains or motifs, and the short length

TABLE 1: WRKY family genes in *Lotus japonicus* genome.

Annotation ID <sup>a</sup>	Gene name	Group	Chr	Location	Annotation ID <sup>a</sup>	Gene name	Group	Chr	Location
CM0012.480.r2.a	LjWRKY1	I	1	58,226,335	<i>CM0432.2010.r2.d</i>	<i>LjWRKY37</i>	<i>Iix</i>	4	7,529,554
CM0032.140.r2.d	LjWRKY2	Iic	1	17,061,228	CM0536.110.r2.d	LjWRKY38	Iid	4	13,585,366
CM0032.170.r2.d	LjWRKY3	Iic	1	17,093,440	CM1622.200.r2.a	LjWRKY39	Iic	4	37,600,439
CM0105.740.r2.a	LjWRKY4	III	1	65,847,050	<i>CM0040.550.r2.d</i>	<i>LjWRKY40</i>	<i>Ila</i>	5	5,190,496
CM0122.1300.r2.m	LjWRKY5	Iib	1	62,472,860	<i>CM0040.670.r2.d</i>	<i>LjWRKY41</i>	<i>Ila</i>	5	5,268,779
CM0122.2160.r2.d	LjWRKY6	Iie	1	63,001,936	CM0239.70.r2.d	LjWRKY42	Iic	5	25,979,555
CM0123.240.r2.m	LjWRKY7	Iib	1	6,329,403	CM0852.250.r2.d	LjWRKY43	Iid	5	2,801,268
CM0133.780.r2.m	LjWRKY8	Iic	1	23,799,861	CM1574.880.r2.m	LjWRKY44	Iid	5	18,726,557
CM0147.250.r2.m	LjWRKY9	Iic	1	42,532,993	CM0037.660.r2.m	LjWRKY45	I	6	14,396,956
CM0315.160.r2.m	LjWRKY10	Iib	1	35,572,836	CM1034.270.r2.d	LjWRKY46	Iie	6	19,990,091
CM0320.110.r2.d	LjWRKY11	III	1	15,934,171	LjT11E14.30.r2.m	LjWRKY47	I	6	9,459,959
CM0433.90.r2.d	LjWRKY12	Iia	1	24,711,963	LjT46F23.30.r2.d	LjWRKY48	Iid	6	17,672,968
<i>CM1413.480.r2.d</i>	<i>LjWRKY13</i>	<i>I</i>	<i>1</i>	<i>34,039,015</i>	<i>LjSGA_012799.2</i>	<i>LjWRKY49</i>	<i>Iic</i>	<i>nd</i>	<i>nd</i>
CM1413.70.r2.d	LjWRKY14	Iie	1	33,660,455	<i>LjSGA_014003.1</i>	<i>LjWRKY50</i>	<i>I</i>	<i>nd</i>	<i>nd</i>
CM0058.50.r2.m	LjWRKY15	I	2	34,643,392	<i>LjSGA_015261.1</i>	<i>LjWRKY51</i>	<i>I</i>	<i>nd</i>	<i>nd</i>
CM0060.250.r2.a	LjWRKY16	III	2	31,851,611	LjSGA_020468.1	LjWRKY52	I	nd	nd
CM0120.570.r2.d	LjWRKY17	Iic	2	20,356,545	LjSGA_021425.1	LjWRKY53	III	nd	nd
CM0191.300.r2.a	LjWRKY18	I	2	43,680,266	LjSGA_025444.2	LjWRKY54	Iie	nd	nd
CM0608.50.r2.m	LjWRKY19	Iix	2	20,505,843	LjSGA_038153.1	LjWRKY55	nd	nd	nd
CM0608.70.r2.m	LjWRKY20	Iic	2	20,514,487	LjSGA_038411.1	LjWRKY56	nd	nd	nd
CM1895.20.r2.m	LjWRKY21	III	2	34,148,852	LjSGA_046833.1	LjWRKY57	Iia	nd	nd
LjB06N21.120.r2.a	LjWRKY22	Iib	2	42,171,450	LjSGA_050544.1	LjWRKY58	nd	nd	nd
CM0243.470.r2.m	LjWRKY23	Iib	3	39,562,863	LjSGA_055568.1	LjWRKY59	Iie	nd	nd
CM0936.60.r2.d	LjWRKY24	I	3	18,082,486	<i>LjSGA_058050.1</i>	<i>LjWRKY60</i>	<i>Iie</i>	<i>nd</i>	<i>nd</i>
LjT41F16.60.r2.d	LjWRKY25	III	3	33,152,607	LjSGA_058052.1	LjWRKY61	nd	nd	nd
LjT45M09.130.r2.d	LjWRKY26	Iia	3	474,491	LjSGA_066036.1	LjWRKY62	nd	nd	nd
CM0004.1290.r2.m	LjWRKY27	III	4	40,133,355	<i>LjSGA_069579.1</i>	<i>LjWRKY63</i>	<i>Iic</i>	<i>nd</i>	<i>nd</i>
CM0004.880.r2.d	LjWRKY28	Iie	4	39,837,881	LjSGA_074792.1	LjWRKY64	Iie	nd	nd
CM0007.1120.r2.a	LjWRKY29	Iic	4	3,358,814	LjSGA_084767.1	LjWRKY65	nd	nd	nd
CM0046.1360.r2.d	LjWRKY30	Iie	4	36,250,655	LjSGA_086809.1	LjWRKY66	nd	nd	nd
CM0179.130.r2.m	LjWRKY31	I	4	28,951,362	LjSGA_091481.1	LjWRKY67	nd	nd	nd
CM0227.580.r2.m	LjWRKY32	I	4	10,519,027	LjSGA_107504.1	LjWRKY68	nd	nd	nd
CM0244.860.r2.m	LjWRKY33	Iid	4	30,871,468	LjSGA_116672.1	LjWRKY69	nd	nd	nd
CM0333.340.r2.d	LjWRKY34	Iic	4	32,856,509	LjT10P12.70.r2.d	LjWRKY70	Iib	nd	nd
CM0333.510.r2.a	LjWRKY35	Iic	4	32,985,615	LjT45D24.80.r2.d	LjWRKY71	Iib	nd	nd
CM0337.220.r2.d	LjWRKY36	Iib	4	3,852,684					

<sup>a</sup> Annotation IDs are from miyakogusa website <http://www.kazusa.or.jp/lotus/>.

Italics indicated pseudogenes.

Chr: chromosome.

nd: nondetermined.

LjWRKY, WRKY genes of *Lotus japonicas*.

of the WRKY domain, generally 2/3 of the normal WRKY domain length. Although the LjWRKY51 protein, including two WRKY domains, lacked a complete WRKY domain at the C-terminal region, this protein was retained for subsequent analyses.

Among these 61 WRKY genes, there were 12 group I WRKY genes, 42 group II WRKY genes, and 7 group III WRKY genes, based on the number of WRKY domains and

the type of zinc finger motifs (Tables 1 and 2). To obtain a better classification within group II WRKY proteins, we constructed a phylogenetic tree with 42 group II LjWRKY using the *Arabidopsis* WRKY sequence as a reference [2] (Figure 1). It was found that LjWRKY genes in group II could be divided into six subgroups, including 5 members in subgroup Iia, 8 in subgroup Iib, 13 in subgroup Iic, 5 in subgroup Iid, 9 in subgroup Iie, and 2 in subgroup Iix

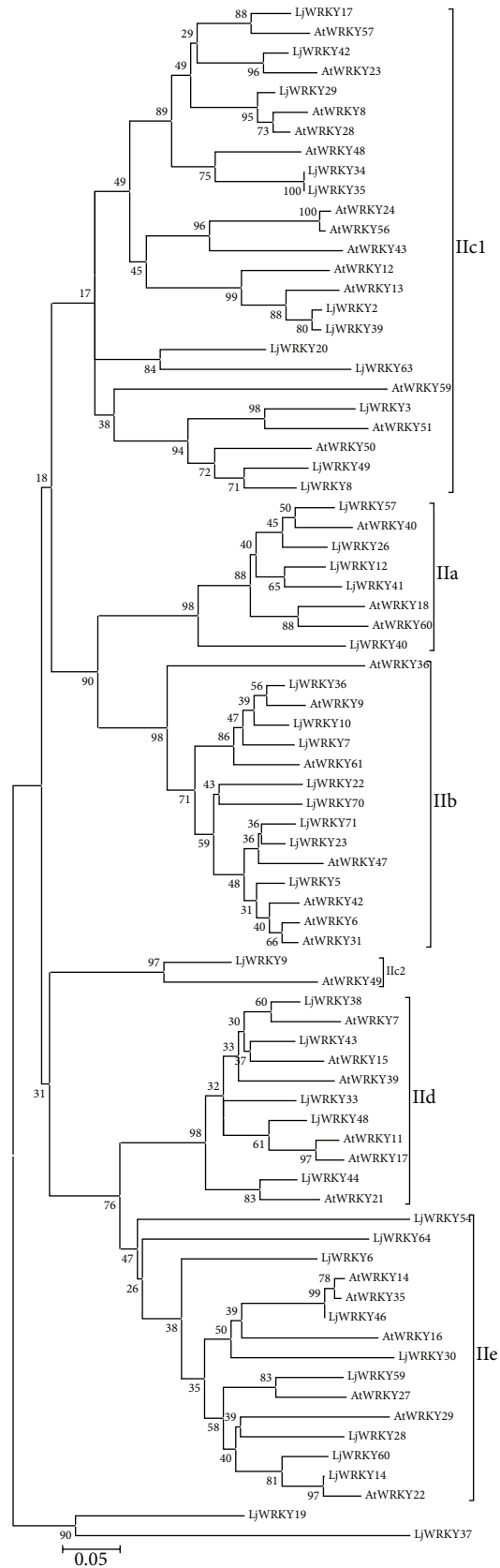


FIGURE 1: Phylogenetic tree of *Arabidopsis* WRLY and *Lotus japonicus* WRKY domain. The phylogenetic tree was constructed using MEGA 4.0 by the Neighbor-Joining (NJ) method with 1,000 bootstrap replicates. The percentage bootstrap scores higher than 50% are indicated on the nodes. The AtWRKY domains from Eulgem et al. (2000).

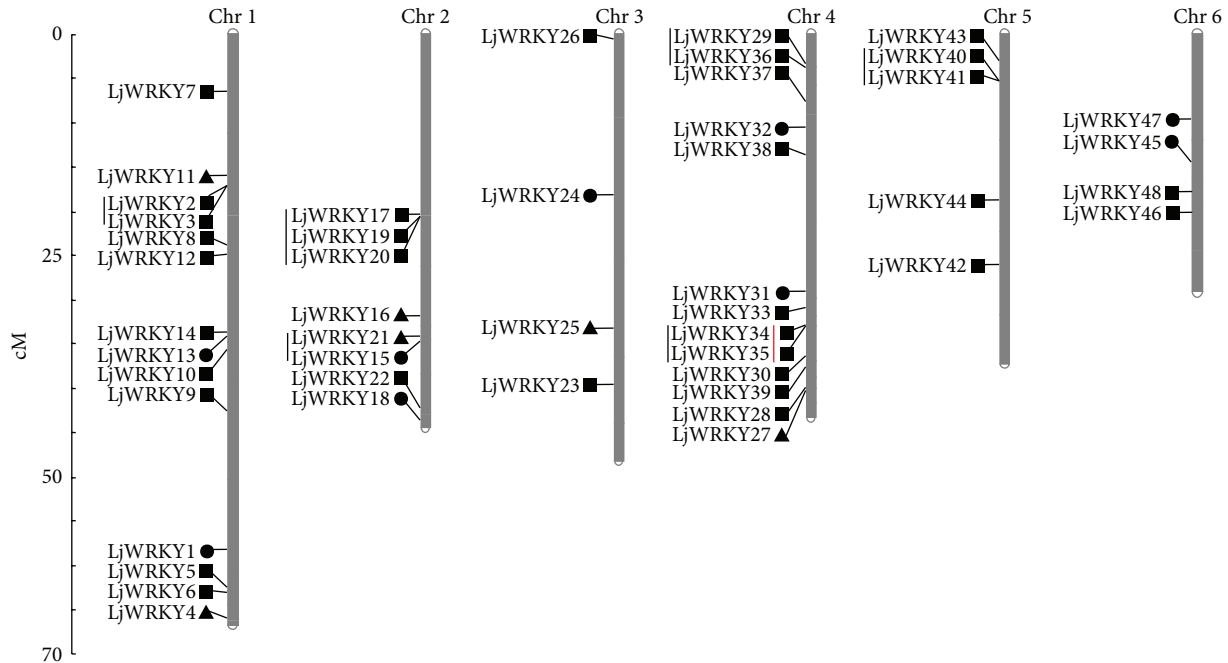


FIGURE 2: Chromosomal locations of *Lotus japonicus* WRKY genes. The chromosome numbers were shown at the top of each chromosome (chromosome; gray bars). The names on the left side of each chromosome correspond to the approximate location of each WRKY gene. The markers next to the gene names represented the groups to which each WRKY gene belongs (●: group I; ■: group II; ▲: group III). The black lines on the left side of the names of the WRKY genes indicated the clusters of gene on each chromosome. The red line on the left side of the markers indicated the duplicated genes. Unmapped WRKY genes were not shown.

(Tables 1 and 2 and Figure 1). Among them, LjWRKY19 and LjWRKY37 genes did not cluster with *Arabidopsis*, and we temporarily named this group IIx (Figure 1).

We found 9 potential pseudogenes among 61 WRKY genes, with either a premature stop codon or a frame shift mutation. Among these pseudogenes, 3 genes were found in group I, 2 in group IIa, 2 in group IIc, and 2 in groups IIx and IIe (Table 1).

**3.2. Gene Duplication and Chromosomal Locations of WRKY Genes.** Gene duplication, including tandem and segmental duplication events, plays a crucial role in genomic expansions. Two or more duplicated genes located in the same chromosome are defined as tandem duplication, while other types of gene duplication are defined as segmental duplication events. We detected only one tandem duplication (LjWRKY34 and LjWRKY35) on chromosome 4 (Figure 2), suggesting that WRKY genes in *L. japonicus* are not recently generated by gene duplication.

A total of 48 WRKY genes could be mapped to chromosomes 1–6, and the others could not be conclusively mapped to any chromosomes, because some WRKY genes have no precise location information. Fourteen WRKY genes including two group I, 10 group II, and two group III genes were located on chromosome 1. Thirteen genes were mapped to chromosome 4. Four genes were mapped to chromosome 3 (one group I, two group II, and one group III genes) and chromosome 6 (two group I and two group II genes), respectively. Eight (two group I, four group II, and two group

III genes) and five WRKY genes (five group II genes) were found on chromosome 2 and chromosome 5, respectively (Figure 2).

A gene cluster is defined as a chromosome region with two or more genes located within 200 kb sequence [60]. Using this criterion, we found 13 WRKY genes forming six gene clusters. Chromosomes 2 and 4 each contain two gene clusters, while only one gene cluster was found on chromosomes 1 and 5, respectively (Figure 2). No clusters were found on chromosomes 3 and 6. Through chromosomal location and gene cluster analysis, we found that the number of genes on chromosomes is disproportionate to the number of gene clusters. For example, 14 WRKY genes were found on chromosome 1, which contains only one gene cluster.

**3.3. Phylogenetic Analysis and Gene Structure.** Amino acid residues of WRKYGQK are the distinguishing regions of the WRKY transcription factor [2, 27]. Multiple alignment analysis of LjWRKY domains found that mutations occurred at R, Y, and Q in the conserved WRKYGQK sequence (Figure 3). Further study showed that the variation arose from amino acid substitutions of R to K or L or from K to C and from Q to Y, E, L, or K (Figure 3). On the other hand, we identified a CX<sub>4</sub>CX<sub>22</sub>HXX motif in subgroup In and IIx genes, a CX<sub>4</sub>CX<sub>23</sub>HXX motif in subgroup Ic and IIe genes, and a CX<sub>5</sub>CX<sub>23</sub>HXX motif in subgroup IIa, IIc, IId, and IIe genes (Figure 3). We found that the WRKY domain was replaced by WKKY in subgroup In and IIx genes, suggesting that subgroup IIx genes originated from

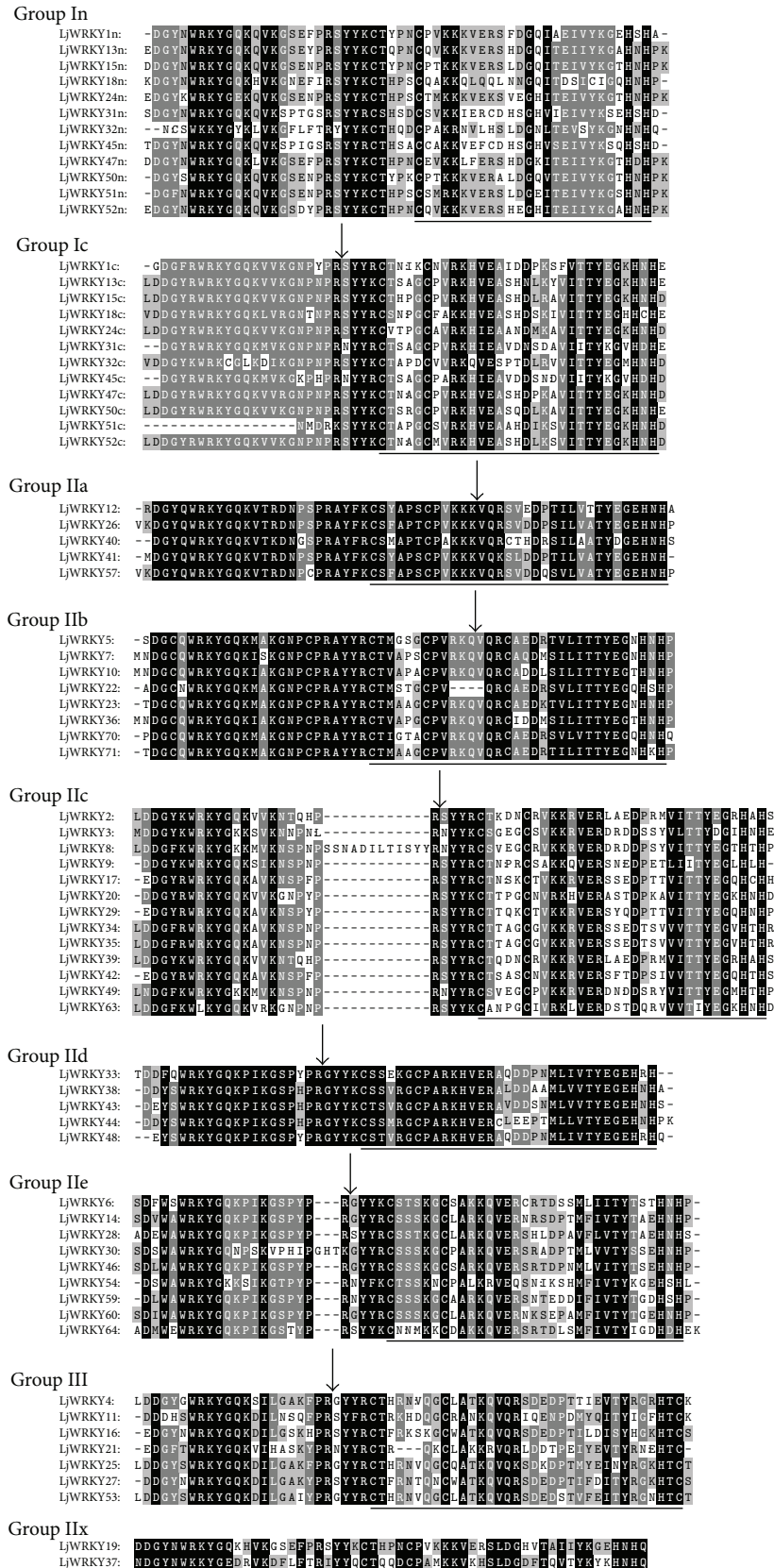


FIGURE 3: Amino acid residue alignment of *Lotus japonicus* WRKY domain. Alignment was performed using the Clustal W program and is displayed with the software Gendoc. Residues that were highly conserved within each of the major groups are in black. The position of a conserved intron was indicated by an arrowhead. The black lines indicated the conserved zinc finger motifs.

TABLE 2: The number of WRKY gene in eight plants.

	Group I	Group IIa	Group IIb	Group IIc	Group IId	Group IIe	Group IIx	Group III	Total	Genome size (Mb)
LjWRKY	12	5	8	13	5	9	2	7	61	472
MtWRKY	19	6	8	19	8	8	0	13	81	375
AtWRKY <sup>a</sup>	13	4	7	18	7	9	0	14	72	125
OsWRKY <sup>a</sup>	15	4	8	15	7	11	0	36	96	480
CsWRKY <sup>a</sup>	10	4	4	16	8	7	0	6	55	367
VvWRKY <sup>a</sup>	12	4	7	14	6	7	0	5	55	487
BdWRKY <sup>b</sup>	17	3	6	21	6	10	0	23	86	272
PtWRKY <sup>c</sup>	50	5	9	13	13	4	0	10	104	485

<sup>a</sup>Data from Ling et al. (2011) [59]; <sup>b</sup>data from Tripathi et al. (2012) [64]; <sup>c</sup>data from He et al. (2012) [61].

N-terminal WRKY domain of group I genes. The subgroup IIa and IIe genes seem to have formed after the group I genes lost alternative WRKY domains.

The WRKY domain phylogenetic tree can be subdivided into eight clades: In, Ic, IIa, IIb, IIc, IId, IIe, and III (IIx nested in In; Figure 4(a)). The group I proteins contain two WRKY domains located at the N-terminal domain (In) or the C-terminal domain (Ic). Although In and Ic belong to group I, members of In and Ic were clustered in different clades, representing the sequence divergence of WRKY domain in In and Ic.

Clade Ic contained 14 members in *L. japonicus*, including 12 members with WRKY domains at the C-terminal region and two group II members (LjWRKY20 and LjWRKY63) (Figure 4(a)). These two group II members were clustered with LjWRKY24c and LjWRKY32c, respectively, indicating a common origin of their domains. Moreover, two group II members (LjWRKY19 and LjWRKY37) were also found in the In clade.

Group II can be divided into five clades with high bootstrap values (Figure 4(a)). Further analysis revealed that IIa and IIb form a single clade and IId and IIe form a clade, while IIc contained 11 LjWRKY members form a clade (Figure 4(a)). This suggests that the domains had a recent gene ancestor or formed under similar selective pressures. Clade III contained seven members which are more similar to clade IId and clade IIe than any other members (Figure 4(a)), suggesting that they may have shared an ancestor before divergence of group II and group III.

During analysis of the cDNA and DNA sequences, we found that most of the LjWRKY genes contained two types of introns in their WRKY domains. The phase-2 intron is spliced exactly after the R position, similar to the splicing position observed in *Arabidopsis* [2]. We designate the phase-2 intron as an R-type intron. A phase-0 intron is located before the V position, at the sixth amino acid after the second C residue in the C<sub>2</sub>H<sub>2</sub> zinc finger motif [22]. We designate the phase-0 intron as a V-type intron. Interestingly, in subgroups Ic, IIc, IId, IIe, and III, the R-type intron is located before the zinc finger motif region in the WRKY domain of genes, while in subgroups IIa and IIb, the V-type intron is found within the zinc finger motif region in the WRKY domain (Figure 4(b)). Furthermore, introns have been lost from LjWRKY28 (subgroup IIe), LjWRKY51c

(subgroup Ic), and subgroups In and IIx (Figure 4(b)). Intron loss can be considered as the result of intron turnover, the result of homologous recombination between an intron-containing allele and a mature mRNA [22].

**3.4. Conserved Motifs in LjWRKY Proteins.** With the exception of the conserved 60 amino acid residues, no functional or structural homologies were previously known from the remainder of the WRKY protein sequences [2]. Analysis of the 20 motifs revealed that LjWRKY motif lengths ranged from 11 to 113 and the distribution of 20 motifs in each amino acid sequence varied greatly (Table 3 and Figure 5). In addition, the function of the majority of LjWRKY motifs could not be predicted. Unexpectedly, we observed a herpes virus glycoprotein motif (motif14) in LjWRKY34 and LjWRKY35, which has not been reported in previous studies of WRKY genes. It will be interesting to analyze the function of this motif in LjWRKY genes in the future.

Compared with motifs of the WRKY protein from *Arabidopsis* [2], *Populus trichocarpa* [61], and *Oryza sativa* [22], we detected three conserved motifs in LjWRKY genes, including Leu zipper, HARF, and NLS motifs. Subgroup IIa (LjWRKY12, LjWRKY26, LjWRKY41, and LjWRKY51), IIb (LjWRKY5, LjWRKY10, LjWRKY22, LjWRKY23, LjWRKY-36, LjWRKY70, and LjWRKY71), and IIe (LjWRKY28) genes contained a Leu zipper motif (motif7). This motif is a hypothetical structure common to a new class of DNA binding proteins. A HARF motif (motif16) was distributed in subgroup IId (LjWRKY33, LjWRKY38, LjWRKY43, LjWRKY44, and LjWRKY48) genes and an NLS motif (motif13) was observed in group I, subgroup IId, subgroup IIe, and group III (LjWRKY13, LjWRKY47, LjWRKY33, LjWRKY38, LjWRKY43, LjWRKY44, LjWRKY48, LjWRKY6, LjWRKY14, LjWRKY30, LjWRKY46, LjWRKY60, LjWRKY4, LjWRKY-21, LjWRKY25, LjWRKY27, and LjWRKY53) genes, but their functions are not clear. Previous studies have shown that WRKY proteins contain a coactivator motif (LXXLL or LXLXLX. L, leucine; X, any amino acid), suggesting the role of these motifs in plant immune responses [22, 61]. In this study, we found a probable coactivator motif in group III genes (LjWRKY4, LjWRKY16, LjWRKY25, LjWRKY27, and LjWRKY53), suggesting the involvement of group III genes in response to pathogens.

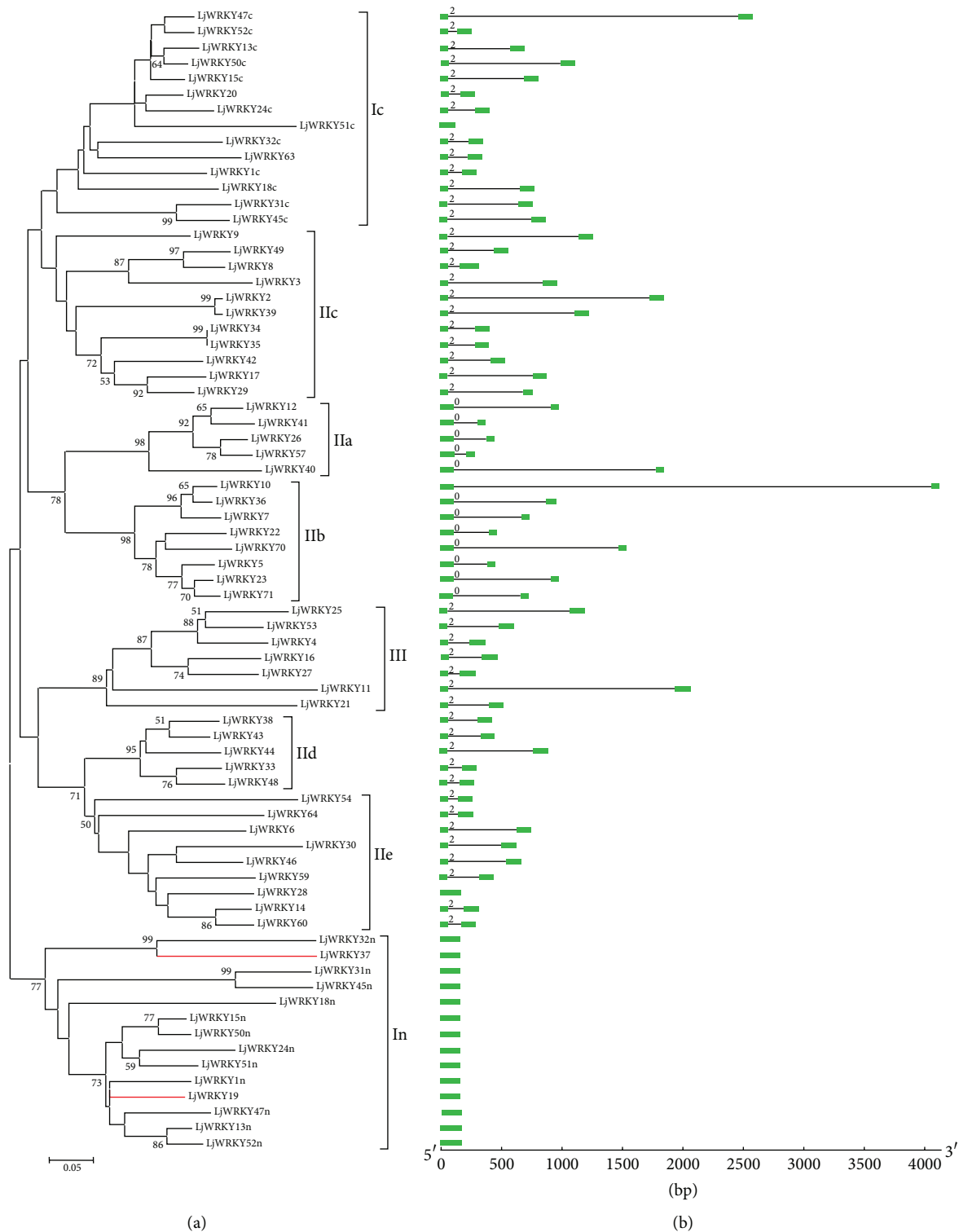


FIGURE 4: Phylogenetic relationships and gene structure of *Lotus japonicus* WRKY domains. (a) Multiple alignments of WRKY domain amino acids executed by Clustal W and the phylogenetic tree constructed using MEGA 4.0 by the Neighbor-Joining (NJ) method with 1,000 bootstrap replicates. The red line indicated group IIx member. The percentage bootstrap scores higher than 50% were indicated on the nodes. (b) Exon-intron structures of WRKY domain genes from *Lotus japonicus*. Exons and introns were represented by green boxes and black lines, respectively. The number indicated introns phases. The sizes of exons and introns can be estimated using the scale at the bottom.



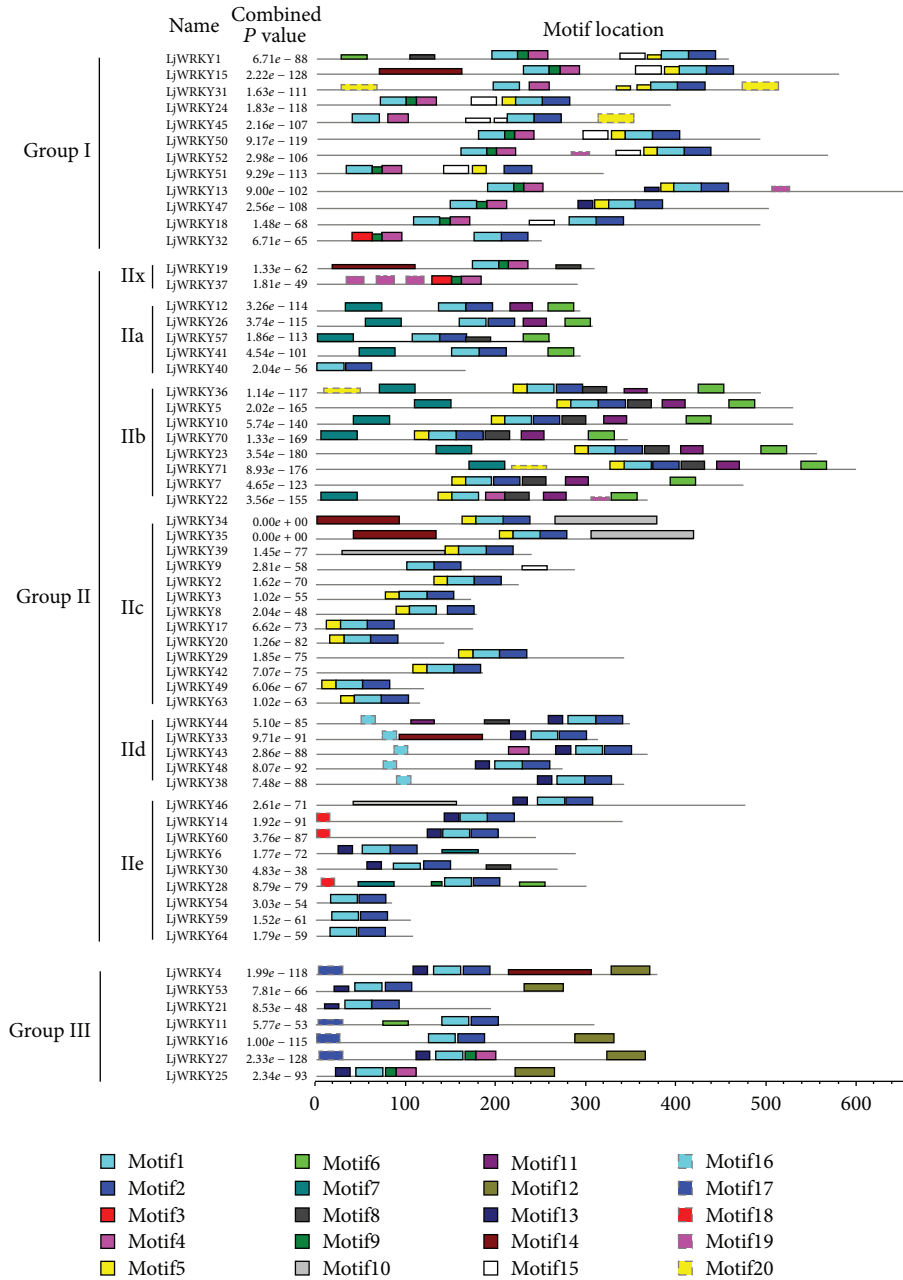


FIGURE 5: Distribution of 20 predicted conserved motifs in *Lotus japonicus* WRKY proteins. Motifs of *Lotus japonicus* WRKY proteins were identified by MEME program. The conserved amino acid sequences and length of each motif are shown in Table 3.

3.5. Evolutionary Analysis of Group III Genes in Plants to Determine Selection Pressure in *L. japonicus* and *M. truncatula*. In order to study the phylogenetic relationships of group III genes, a phylogenetic tree was estimated using the WRKY domain of seven species, including monocots and dicots. Group III LjWRKY genes did not form a clade; on the contrary, group III LjWRKY genes formed a clade with MtWRKY genes (Figure 6). This suggests that we did not find any paralogs of group III LjWRKY genes, but it suggests that group III LjWRKY genes are orthologous to MtWRKY genes. Paralogous relationships were observed among WRKY

genes in other species (i.e., MtWRKY, AtWRKY, PtWRKY, OsWRKY, and BdWRKY). Gene duplication events are considered as the most likely process to result in paralogous copies of genes [59].

To detect whether selection pressure affected group III LjWRKY genes,  $\omega$  ( $d_N/d_S$ ) was calculated for phylogenetic nodes in PAML (Table 4 and Figure 7). In *L. japonicus* and *M. truncatula*, the ML estimations of  $\omega$  values for all nodes under the model M0 were <1 (Table 4), suggesting that group III LjWRKY and MtWRKY genes have been under purifying selection during evolution. Nevertheless, the log

TABLE 3: Consensus sequences of WRKY motif in *Lotus japonicus* proteins.

Motif	Width	Consensus sequences <sup>a</sup>
Motif1	29	ILDDGYRWRKYGQKVIKGNPYPRSYRCT
Motif2	29	PGCPVKKHVERSAEDPSMVITTYEGEHNH
Motif3	21	WRKYGQKVVKGSPFPRSYK
Motif4	21	RSLDGHVTEITYKGxHNC PKP
Motif5	15	VREPRVAVQTKSEVD
Motif6	28	LVEAMAAALTADPNFTAALAAAISSIG
Motif7	39	EELERLNAENKKLREMLDQMNENYNALQMHLVELMQKQK
Motif8	26	PAAMAMASTTSAASMLLSGSMPSAD
Motif9	11	HPNCPAKKKVE
Motif10	113	QNLQVQNFQNNVQQAAPSFSTTPSFNAATPVPPISYRNAIIPFPNVANPAPNTH FNTSSFSGFLQGINGVDFVQSSRFMANNNQGLLRNGLLQDMFVPAHMEFGGGGRR
Motif11	24	CSASMATLSASAPPTITLDTQS
Motif12	42	DITEIISAPTSVTNSPILDLLLDNVELDSNFPFNTPELFS
Motif13	15	KKRKMRVKRTVRVPA
Motif14	91	MPPPPSPPLPPLSPLPLPPFHQQTSSFGSKDLLTIEDFDPALFD WNPTTTTNTAAADVTSPPLFTTQISHVPSPATSNILPEDSFDV
Motif15	27	SGGGELDDDEPDAKRWKGEENDGYSA
Motif16	15	LLNRTGHARFRAP
Motif17	26	MEHLHKWEQKALINELIQGMELARKL
Motif18	14	MDDDWDLHAVVRGC
Motif19	19	MSVMPEFLQQGLNSSPVNV
Motif20	39	QCSGDIEGELTGEAMELGGEKAIESARALLSIGFEIKPC

<sup>a</sup>Consensus amino acid sequence was derived from MEME program.

likelihood ratio differences between models M3 and M0 were statistically significant for all nodes tested, except nodes 1 and 2 in MtWRKY (Table 4). This indicates that some genes may be under positive selection. Interestingly, we further analyzed the positive selection in group III genes with models M8 and M7. The  $\omega$  values for all nodes were  $\geq 1$  under M8. However, only one node identified one positive selection site in group III MtWRKY genes under model M8 (Table 4). This result shows that group III WRKY genes in *L. japonicus* and *M. truncatula* have not undergone positive selection.

#### 4. Discussion

WRKY genes are commonly found in land plants and many WRKY genes have been identified and classified in *Arabidopsis* [2], *Oryza sativa* [6, 22, 62], *Hordeum vulgare* [63], *Cucumis sativus* [59], *Brachypodium distachyon* [64], and *Populus trichocarpa* [61]. However, little information has been reported on WRKY genes in leguminous forage crops. In 2008, approximately 67% of the *L. japonicus* genome (472 Mb) sequences were available on public databases, representing 91.3% coverage of the gene space [49]. In our current work, we conducted an analysis of 61 WRKY genes in the *L. japonicus* genome.

One hundred and four WRKY genes were identified in *P. trichocarpa*, while only 55 WRKY genes were discovered in *C. sativus* and *Vitis vinifera* genomes, respectively (Table 2). The relative number of WRKY genes was not associated with

genome size. For example, the number of WRKY genes was about 2x greater in *P. trichocarpa* (104) than in *C. sativus* (55), while these two plants have an approximately equal genome size (458 Mb and 487 Mb; Table 2). By analyzing the number of WRKY gene groups or subgroups (except for subgroup IIx) in each species where they have been characterized, we discovered an uneven distribution of the number of WRKY genes in each group. In rice, the largest number of WRKY genes (36) was found in group III, but only four genes were found to belong to subgroup IIa (Table 2). In *C. sativus*, however, more WRKY genes (16) are classified as subgroup IIc and fewer WRKY genes (4) are categorized in subgroup IIa (Table 2).

Although gene duplication events seem to have led to the expansion of WRKY genes in the *Arabidopsis* [2] and *Oryza* [22] genomes, duplicated WRKY genes were not detected in *C. sativus* [59]. It is not yet clear whether gene duplication typically occurs during LjWRKY gene evolution. Among the 61 WRKY genes, we found that, in *L. japonicus*, only 2 of them were involved in duplication events. In contrast, 11 gene duplication events were identified in the model plant *M. truncatula*. In addition, there were 42 PtWRKY gene duplication events identified in the *P. trichocarpa* genome [61], with 29 out of 42 PtWRKY genes arising from segmental duplication. This comparison suggests that the expansion of LjWRKY genes is not necessarily dependent on gene duplication events.

TABLE 4: Likelihood ratio test results of group III *Lotus japonicus* WEKY and *Medicago truncatula* WRKY genes.

Node <sup>a</sup>	$d_N/d_S$ under M0 <sup>b</sup>	$2\Delta \ln L$ M3 versus M0	$2\Delta \ln L$ M8 versus M7	M8 estimates <sup>c</sup>	Number of positive selection site <sup>d</sup>
Group III LjWRKY					
1	0.29349	67.57**	0.000074	$\omega = 1$ ( $P = 0.72, q = 1.2$ )	0
2	0.29349	67.57**	0.000074	$\omega = 1$ ( $P = 0.72, q = 1.22$ )	0
3	0.25749	74.47**	0.095776	$\omega = 1.13$ ( $P = 0.84, q = 2$ )	0
4	0.23855	112.17**	0.000198	$\omega = 1$ ( $P = 0.8, q = 1.8$ )	0
Group III MtWRKY					
1	0.31282	3.45	0.000016	$\omega = 1$ ( $P = 0.77, q = 1.44$ )	0
2	0.28717	4.31	0.000042	$\omega = 1$ ( $P = 1.9, q = 4.18$ )	0
3	0.27266	27.39**	0.178716	$\omega = 2.19$ ( $P = 0.52, q = 0.93$ )	1
4	0.26004	70.18**	0.000078	$\omega = 1$ ( $P = 0.76, q = 1.51$ )	0
5	0.24433	81.88**	0.000204	$\omega = 1$ ( $P = 0.74, q = 1.48$ )	0
6	0.00517	43.43**	0.029352	$\omega = 34.48$ ( $P = 1.09, q = 61.42$ )	0
7	0.20188	209.83**	0.000302	$\omega = 1$ ( $P = 0.64, q = 1.68$ )	0
8	0.18564	233.88**	0.000154	$\omega = 2.5$ ( $P = 0.6, q = 1.8$ )	0

<sup>\*</sup> $P < 0.05$  and <sup>\*\*</sup> $P < 0.01$  ( $\chi^2$  test).

<sup>a</sup>Node number from the phylogenetic tree.

<sup>b</sup> $d_N/d_S$  is the average ratio over sites under a codon model with one ratio.

<sup>c</sup> $\omega$  was estimated under model M8;  $P$  and  $q$  are the parameters of the beta distribution.

<sup>d</sup>The number of amino acid sites estimated to have undergone positive selection under M8.

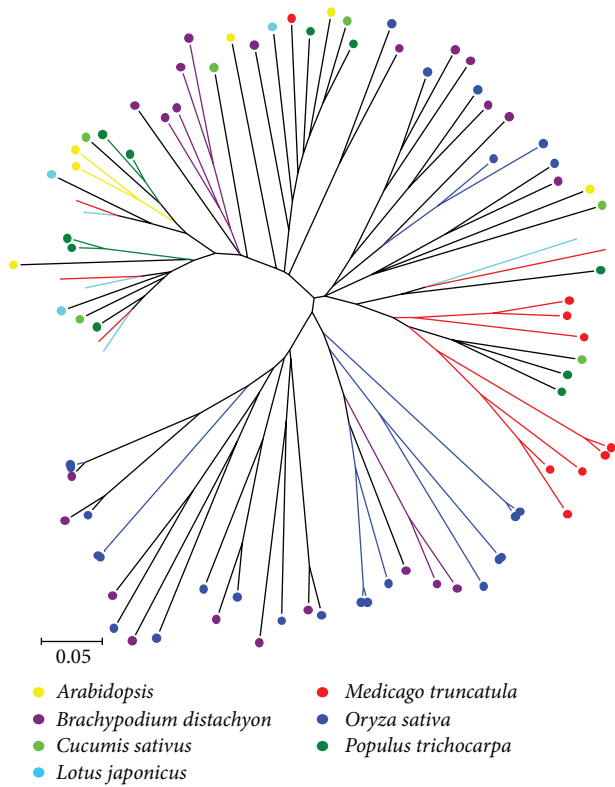


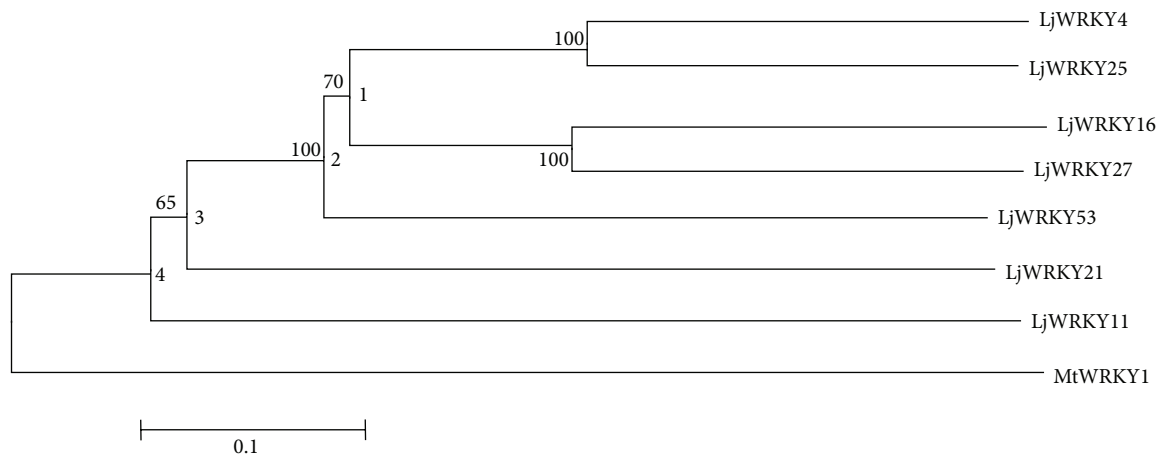
FIGURE 6: Phylogenetic tree of group III WRKY domains from *Arabidopsis*, *Brachypodium distachyon*, *Cucumis sativus*, *Lotus japonicus*, *Medicago truncatula*, *Oryza sativa*, and *Populus trichocarpa*. The phylogenetic tree was constructed using MEGA 4.0 by the Neighbor-Joining (NJ) method with 1,000 bootstrap replicates. Each WRKY predicted orthologous gene was indicated in a specific color.

We found few duplicated LjWRKY genes in the *L. japonicus* genome and there are at least three possible explanations: (1) most duplicated genes have been lost after segmental duplication events in LjWRKY genes [65]; (2) LjWRKY genes that are nonfunctional or duplicate the function of other copies are inclined to disappear to avoid fitness cost; and (3) the draft sequenced genome in *L. japonicus* may not yet contain all WRKY genes present in the genome.

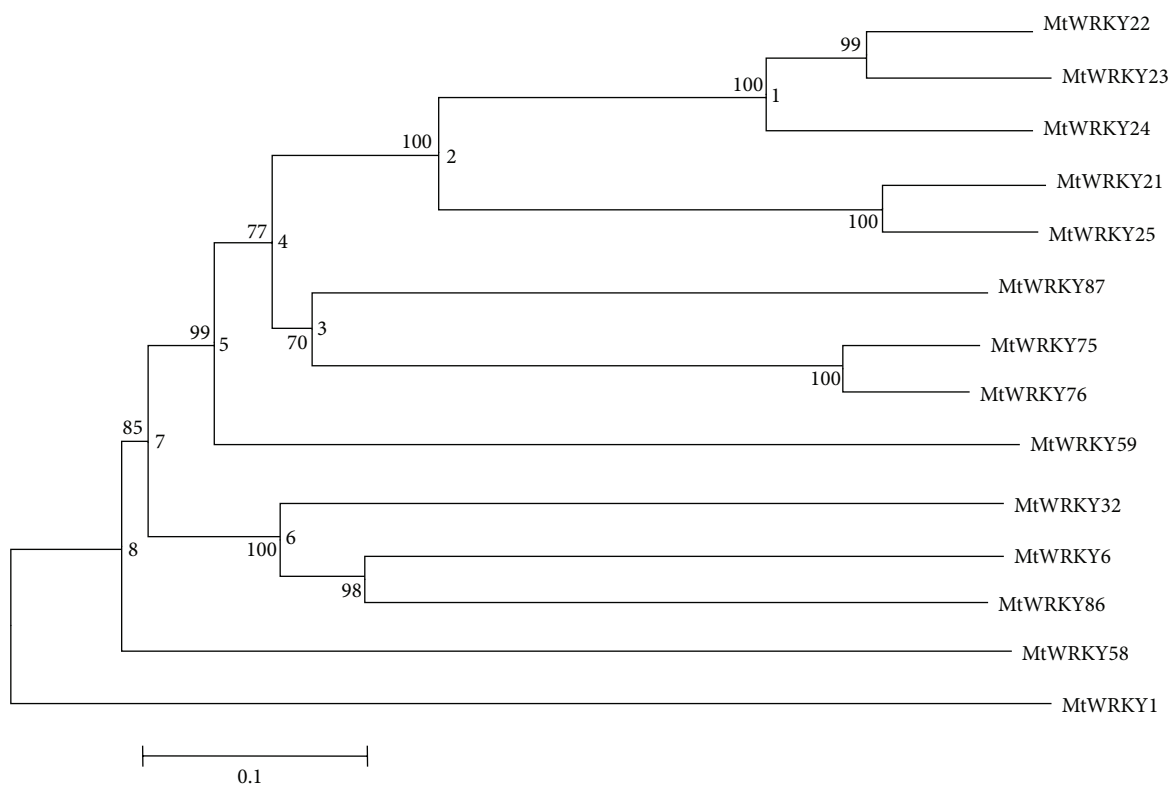
The WRKY gene sequences contain two types of conserved introns (R-type and V-type) [22], and we found them in this study. In *Arabidopsis* [2] and *C. sativus* [59] subgroup IIa and IIb WRKY genes, R-type introns are inserted in domains located at the fourth amino acid (K residue) after the second C residue in the zinc finger region and there are no V-type introns. These results suggest that introns in subgroup IIa and IIb WRKY genes may have different origins.

The average length of the conserved introns (597 bp) in *L. japonicus* WRKY domains was longer than that in *Arabidopsis* (241 bp) but shorter than that in *M. truncatula* (705 bp). In nematodes and mammals, there is a dramatic decline in average intron size when there is increased gene expression [66]. However, the correlation between expression of WRKY genes in various plants and the intron length needs to be tested.

We found an interesting phenomenon which may provide context to interpret group I, II, and III WRKY gene origin and evolutionary relationships. Two conserved motifs (motif4 and/or motif9) were observed after motif1 (contains the conserved sequence WRKYGQK) in the N-terminal region, while, in the C-terminal region, other conserved motifs (motif5 or/and motif13) occur before motif1 (Figure 5). In addition, analysis of group II and III LjWRKY gene motifs



(a)



(b)

FIGURE 7: Phylogenetic tree of group III WRKY genes of *Lotus japonicus* and *Medicago truncatula*. The phylogenetic trees were constructed using MEGA 4.0 by the Neighbor-Joining (NJ) method with 1,000 bootstrap replicates. Numbers on the left of each internal node represented bootstrap support values and numbers on the right of each node represented the nodes that were used for positive selection analysis. MtWRKY1 gene was used as outgroup. The tree represented phylogenetic relationships among LjWRKY genes (a) and MtWRKY genes (b).

revealed that most LjWRKY genes contain motif1, motif5, and/or motif13. However, motif1, motif4, and/or motif9 in the N-terminal region of group I genes are distributed in LjWRKY19, LjWRKY37, (group II) and LjWRKY25 and LjWRKY27 (group III), respectively (Figure 5). Previous research showed that group II and III WRKY genes evolved from group I through the loss of the N-terminal WRKY domain [2, 5]. However, our results indicate that some WRKY

genes in groups II and III may have originated from the N-terminal region of group I. From the phylogenetic relationship, we found that LjWRK19 and LjWRKY37 (group II) clustered with group I<sub>n</sub> with well-supported bootstrap values, suggesting that they have a common origin or ancestry. Moreover, we consider that some group III WRKY genes could have resulted from a mutation in the zinc finger motif in the N-terminal of group I genes. To confirm our inference,

we analyzed the conserved motifs in MtWRKY genes. The same phenomenon was detected in MtWRKY genes based on analysis of location of motifs (data not shown) and 13 MtWRKY genes might have evolved from N-terminal of group I genes.

The natural selection pressure imposed by pathogens is expected to be diverse in different plants [67] and WRKY genes and NBS-LRR genes forming a fused gene may effectively resist a wide variety of pathogens. Fusion genes contain the C-terminal WRKY motif and a NBS-LRR (nucleotide-binding site-leucine-rich repeat) motif in the R gene was identified in AtWRKY [68, 69] and OsWRKY [22] genes. The R gene is mainly involved in a pathogen response pathway in plants [70]. Fusion gene, one gene included WRKY and NBS-LRR domains and/or motifs, was not detected in *L. japonicus*.

The majority of group III AtWRKY genes under positive selection are expressed in response to various abiotic stresses [59]. In contrast, the expression of group III CsWRKY genes shows that these genes are under purifying selection and are specialized to respond as single type of stress [59]. Ling et al. showed that positive selection may have resulted in the functional divergence of duplicated genes during the expansion of group III WRKY genes in *Arabidopsis* [59]. Similarly, our selection pressure study of group III LjWRKY and MtWRKY genes found that expansion of these genes may be under purifying selection, although gene duplication events occurred within these genes. Purifying selection may generate genes with conserved functions or pseudogenization [71] in duplicated group III MtWRKY genes. Therefore, we speculate that group III LjWRKY and MtWRKY genes may be more conservative in their response to stress.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This research was supported by grants from the National Basic Research Program of China (2014CB138702). The authors are grateful to the members of the State Key Laboratory of Grassland Agro-ecosystems for their assistance in this study. They thank associate editor Professor John Parkinson and an anonymous reviewer for taking extensive time and care for providing thoughtful comments on the manuscript.

## References

- [1] E. Martinez, "Multi-protein complexes in eukaryotic gene transcription," *Plant Molecular Biology*, vol. 50, no. 6, pp. 925–947, 2002.
- [2] T. Eulgem, P. J. Rushton, S. Robatzek, and I. E. Somssich, "The WRKY superfamily of plant transcription factors," *Trends in Plant Science*, vol. 5, no. 5, pp. 199–206, 2000.
- [3] J.-J. Liu and A. K. M. Ekramoddoullah, "Identification and characterization of the WRKY transcription factor family in *Pinus monticola*," *Genome*, vol. 52, no. 1, pp. 77–88, 2009.
- [4] P. J. Rushton, J. T. Torres, M. Parniske, P. Wernert, K. Hahlbrock, and I. E. Somssich, "Interaction of elicitor-induced DNA-binding proteins with elicitor response elements in the promoters of parsley PR1 genes," *EMBO Journal*, vol. 15, no. 20, pp. 5690–5700, 1996.
- [5] Y. Zhang and L. Wang, "The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants," *BMC Evolutionary Biology*, vol. 5, p. 1, 2005.
- [6] Z. Xie, Z.-L. Zhang, X. Zou et al., "Annotations and functional analyses of the rice WRKY gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells," *Plant Physiology*, vol. 137, no. 1, pp. 176–189, 2005.
- [7] I. Ciolkowski, D. Wanke, R. P. Birkenbihl, and I. E. Somssich, "Studies on DNA-binding selectivity of WRKY transcription factors lend structural clues into WRKY-domain function," *Plant Molecular Biology*, vol. 68, no. 1-2, pp. 81–92, 2008.
- [8] T. Eulgem, P. J. Rushton, E. Schmelzer, K. Hahlbrock, and I. E. Somssich, "Early nuclear events in plant defence signalling: rapid gene activation by WRKY transcription factors," *EMBO Journal*, vol. 18, no. 17, pp. 4689–4699, 1999.
- [9] C. Sun, S. Palmqvist, H. Olsson, M. Borén, S. Ahlandsberg, and C. Jansson, "A novel WRKY transcription factor, SUSIBA2, participates in sugar signaling in barley by binding to the sugar-responsive elements of the *iso1* promoter," *Plant Cell*, vol. 15, no. 9, pp. 2076–2092, 2003.
- [10] S. de Pater, V. Greco, K. Pham, J. Memelink, and J. Kijne, "Characterization of a zinc-dependent transcriptional activator from *Arabidopsis*," *Nucleic Acids Research*, vol. 24, no. 23, pp. 4624–4631, 1996.
- [11] S. Ishiguro and K. Nakamura, "Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5' upstream regions of genes coding for sporamin and  $\beta$ -amylase from sweet potato," *Molecular and General Genetics*, vol. 244, no. 6, pp. 563–571, 1994.
- [12] S. A. Goff, D. Ricke, T.-H. Lan et al., "A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*)," *Science*, vol. 296, no. 5565, pp. 92–100, 2002.
- [13] T. Huang and J. G. Duman, "Cloning and characterization of a thermal hysteresis (antifreeze) protein with DNA-binding activity from winter bittersweet nightshade, *Solanum dulcamara*," *Plant Molecular Biology*, vol. 48, no. 4, pp. 339–350, 2002.
- [14] N. Kato, E. Dubouzet, Y. Kokabu et al., "Identification of a WRKY protein as a transcriptional regulator of benzylisoquinoline alkaloid biosynthesis in *Coptis japonica*," *Plant and Cell Physiology*, vol. 48, no. 1, pp. 8–18, 2007.
- [15] D.-J. Kim, S. M. Smith, and C. J. Leaver, "A cDNA encoding a putative SPF1-type DNA-binding protein from cucumber," *Gene*, vol. 185, no. 2, pp. 265–269, 1997.
- [16] V. Levée, I. Major, C. Levasseur, L. Tremblay, J. MacKay, and A. Séguin, "Expression profiling and functional analysis of *Populus WRKY23* reveals a regulatory role in defense," *New Phytologist*, vol. 184, no. 1, pp. 48–70, 2009.
- [17] J. Li, G. Brader, and E. T. Palva, "The WRKY70 transcription factor: a node of convergence for jasmonate-mediated and salicylate-mediated signals in plant defense," *Plant Cell*, vol. 16, no. 2, pp. 319–331, 2004.
- [18] N. L. Mantri, R. Ford, T. E. Coram, and E. C. K. Pang, "Transcriptional profiling of chickpea genes differentially regulated in response to high-salinity, cold and drought," *BMC Genomics*, vol. 8, article 303, 2007.
- [19] C. Marchive, R. Mzid, L. Deluc et al., "Isolation and characterization of a *Vitis vinifera* transcription factor, VvWRKY1, and its

- effect on responses to fungal pathogens in transgenic tobacco plants," *Journal of Experimental Botany*, vol. 58, no. 8, pp. 1999–2010, 2007.
- [20] L. Pnueli, E. Hallak-Herr, M. Rozenberg et al., "Molecular and biochemical mechanisms associated with dormancy and drought tolerance in the desert legume *Retama raetam*," *Plant Journal*, vol. 31, no. 3, pp. 319–330, 2002.
- [21] B. Ülker and I. E. Somssich, "WRKY transcription factors: from DNA binding towards biological function," *Current Opinion in Plant Biology*, vol. 7, no. 5, pp. 491–498, 2004.
- [22] K.-L. Wu, Z.-J. Guo, H.-H. Wang, and J. Li, "The WRKY family of transcription factors in rice and *Arabidopsis* and their origins," *DNA Research*, vol. 12, no. 1, pp. 9–26, 2005.
- [23] N. D. Young, F. Debelle, G. E. D. Oldroyd et al., "The *Medicago* genome provides insight into the evolution of rhizobial symbioses," *Nature*, vol. 480, no. 7378, pp. 520–524, 2011.
- [24] Q.-Y. Zhou, A.-G. Tian, H.-F. Zou et al., "Soybean WRKY-type transcription factor genes, GmWRKY13, GmWRKY21, and GmWRKY54, confer differential tolerance to abiotic stresses in transgenic *Arabidopsis* plants," *Plant Biotechnology Journal*, vol. 6, no. 5, pp. 486–503, 2008.
- [25] G. Glöckner, L. Eichinger, K. Szafranski et al., "Sequence and analysis of chromosome 2 of *Dictyostelium discoideum*," *Nature*, vol. 418, no. 6893, pp. 79–85, 2002.
- [26] R. Ramamoorthy, S.-Y. Jiang, N. Kumar, P. N. Venkatesh, and S. Ramachandran, "A comprehensive transcriptional profiling of the WRKY gene family in rice under various abiotic and phytohormone treatments," *Plant and Cell Physiology*, vol. 49, no. 6, pp. 865–879, 2008.
- [27] P. J. Rushton, I. E. Somssich, P. Ringler, and Q. J. Shen, "WRKY transcription factors," *Trends in Plant Science*, vol. 15, no. 5, pp. 247–258, 2010.
- [28] T. Asai, G. Tena, J. Plotnikova et al., "Map kinase signalling cascade in *Arabidopsis* innate immunity," *Nature*, vol. 415, no. 6875, pp. 977–983, 2002.
- [29] D. Yu, C. Chen, and Z. Chen, "Evidence for an important role of WRKY DNA binding proteins in the regulation of NPR1 gene expression," *Plant Cell*, vol. 13, no. 7, pp. 1527–1539, 2001.
- [30] Z.-J. Guo, Y.-C. Kan, X.-J. Chen, D.-B. Li, and D.-W. Wang, "Characterization of a rice WRKY gene whose expression is induced upon pathogen attack and mechanical wounding," *Acta Botanica Sinica*, vol. 46, no. 8, pp. 955–964, 2004.
- [31] C. Chen and Z. Chen, "Isolation and characterization of two pathogen- and salicylic acid-induced genes encoding WRKY DNA-binding proteins from tobacco," *Plant Molecular Biology*, vol. 42, no. 2, pp. 387–396, 2000.
- [32] Y. Liu, M. Schiff, and S. P. Dinesh-Kumar, "Involvement of MEK1 MAPKK, NTF6 MAPK, WRKY/MYB transcription factors, COI1 and CTR1 in N-mediated resistance to tobacco mosaic virus," *Plant Journal*, vol. 38, no. 5, pp. 800–809, 2004.
- [33] R. S. Cormack, T. Eulgem, P. J. Rushton, P. Köchner, K. Hahlbrock, and I. E. Somssich, "Leucine zipper-containing WRKY proteins widen the spectrum of immediate early elicitor-induced WRKY transcription factors in parsley," *Biochimica et Biophysica Acta*, vol. 1576, no. 1–2, pp. 92–100, 2002.
- [34] T. Eulgem and I. E. Somssich, "Networks of WRKY transcription factors in defense signaling," *Current Opinion in Plant Biology*, vol. 10, no. 4, pp. 366–371, 2007.
- [35] S. Li, Q. Fu, W. Huang, and D. Yu, "Functional analysis of an *Arabidopsis* transcription factor WRKY25 in heat stress," *Plant Cell Reports*, vol. 28, no. 4, pp. 683–693, 2009.
- [36] S. Vandenamee, K. van der Kelen, J. Dat et al., "A comprehensive analysis of hydrogen peroxide-induced gene expression in tobacco," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 26, pp. 16113–16118, 2003.
- [37] M. M. Izaguirre, A. L. Scopel, I. T. Baldwin, and C. L. Ballaré, "Convergent responses to stress. Solar ultraviolet-B radiation and *Manduca sexta* herbivory elicit overlapping transcriptional responses in field-grown plants of *Nicotiana longiflora*," *Plant Physiology*, vol. 132, no. 4, pp. 1755–1767, 2003.
- [38] M. Skibbe, N. Qu, I. Galis, and I. T. Baldwin, "Induced plant defenses in the natural environment: *Nicotiana attenuata* WRKY3 and WRKY6 coordinate responses to herbivory," *Plant Cell*, vol. 20, no. 7, pp. 1984–2000, 2008.
- [39] W. Grunewald, M. Karimi, K. Wiczorek et al., "A role for AtWRKY23 in feeding site establishment of plant-parasitic nematodes," *Plant Physiology*, vol. 148, no. 1, pp. 358–368, 2008.
- [40] H. C. Yong, H.-S. Chang, R. Gupta, X. Wang, T. Zhu, and S. Luan, "Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in *Arabidopsis*," *Plant Physiology*, vol. 129, no. 2, pp. 661–677, 2002.
- [41] M. Luo, E. S. Dennis, F. Berger, W. J. Peacock, and A. Chaudhury, "MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in *Arabidopsis*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 48, pp. 17531–17536, 2005.
- [42] R. Zentella, Z.-L. Zhang, M. Park et al., "Global analysis of DELLA direct targets in early gibberellin signaling in *Arabidopsis*," *Plant Cell*, vol. 19, no. 10, pp. 3037–3057, 2007.
- [43] Z.-L. Zhang, Z. Xie, X. Zou, J. Casaretto, T.-H. D. Ho, and Q. J. Shen, "A rice WRKY gene encodes a transcriptional repressor of the gibberellin signaling pathway in aleurone cells," *Plant Physiology*, vol. 134, no. 4, pp. 1500–1513, 2004.
- [44] X. Zou, D. Neuman, and Q. J. Shen, "Interactions of two transcriptional repressors and two transcriptional activators in modulating gibberellin signaling in aleurone cells," *Plant Physiology*, vol. 148, no. 1, pp. 176–186, 2008.
- [45] K. Hinderhofer and U. Zentgraf, "Identification of a transcription factor specifically expressed at the onset of leaf senescence," *Planta*, vol. 213, no. 3, pp. 469–473, 2001.
- [46] S. Robatzek and I. E. Somssich, "Targets of AtWRKY6 regulation during plant senescence and pathogen defense," *Genes and Development*, vol. 16, no. 9, pp. 1139–1149, 2002.
- [47] C. S. Johnson, B. Kolevski, and D. R. Smyth, "Transparent Testa Glabra2, a trichome and seed coat development gene of *Arabidopsis*, encodes a WRKY transcription factor," *Plant Cell*, vol. 14, no. 6, pp. 1359–1375, 2002.
- [48] C. Chen and Z. Chen, "Potentiation of developmentally regulated plant defense response by AtWRKY18, a pathogen-induced *Arabidopsis* transcription factor," *Plant Physiology*, vol. 129, no. 2, pp. 706–716, 2002.
- [49] S. Sato, Y. Nakamura, T. Kaneko et al., "Genome structure of the legume, *Lotus japonicus*," *DNA Research*, vol. 15, no. 4, pp. 227–239, 2008.
- [50] R. D. Finn, J. Mistry, B. Schuster-Böckler et al., "Pfam: clans, web tools and services," *Nucleic acids research*, vol. 34, pp. 247–251, 2006.
- [51] J. D. Thompson, D. G. Higgins, and T. J. Gibson, "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap

- penalties and weight matrix choice," *Nucleic Acids Research*, vol. 22, no. 22, pp. 4673–4680, 1994.
- [52] T. Zhou, Y. Wang, J.-Q. Chen et al., "Genome-wide identification of NBS genes in *japonica* rice reveals significant expansion of divergent non-TIR NBS-LRR genes," *Molecular Genetics and Genomics*, vol. 271, no. 4, pp. 402–415, 2004.
- [53] K. Tamura, J. Dudley, M. Nei, and S. Kumar, "MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0," *Molecular Biology and Evolution*, vol. 24, no. 8, pp. 1596–1599, 2007.
- [54] Z. Yang, S. Gu, X. Wang, W. Li, Z. Tang, and C. Xu, "Molecular evolution of the CPP-like gene family in plants: insights from comparative genomics of *Arabidopsis* and rice," *Journal of Molecular Evolution*, vol. 67, no. 3, pp. 266–277, 2008.
- [55] A.-Y. Guo, Q.-H. Zhu, X. Chen, and J.-C. Luo, "GSDS: a gene structure display server," *Yi Chuan*, vol. 29, no. 8, pp. 1023–1026, 2007.
- [56] T. L. Bailey and C. Elkan, "The value of prior knowledge in discovering motifs with MEME," in *Proceedings of the 3rd International Conference on Intelligent Systems for Molecular Biology*, vol. 3, pp. 21–29, 1995.
- [57] M. Suyama, D. Torrents, and P. Bork, "PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments," *Nucleic Acids Research*, vol. 34, pp. W609–W612, 2006.
- [58] Z. Yang, "PAML 4: phylogenetic analysis by maximum likelihood," *Molecular Biology and Evolution*, vol. 24, no. 8, pp. 1586–1591, 2007.
- [59] J. Ling, W. Jiang, Y. Zhang et al., "Genome-wide analysis of WRKY gene family in *Cucumis sativus*," *BMC Genomics*, vol. 12, article 471, 2011.
- [60] E. B. Houb, "The arms race is ancient history in *Arabidopsis*, the wildflower," *Nature Reviews Genetics*, vol. 2, no. 7, pp. 516–527, 2001.
- [61] H. He, Q. Dong, Y. Shao et al., "Genome-wide survey and characterization of the WRKY gene family in *Populus trichocarpa*," *Plant Cell Reports*, vol. 31, no. 7, pp. 1199–1217, 2012.
- [62] C. A. Ross, Y. Liu, and Q. J. Shen, "The WRKY gene family in rice (*Oryza sativa*)," *Journal of Integrative Plant Biology*, vol. 49, no. 6, pp. 827–842, 2007.
- [63] E. Mangelsen, J. Kilian, K. W. Berendzen et al., "Phylogenetic and comparative gene expression analysis of barley (*Hordeum vulgare*) WRKY transcription factor family reveals putatively retained functions between monocots and dicots," *BMC Genomics*, vol. 9, article 194, 2008.
- [64] P. Tripathi, R. C. Rabara, T. J. Langum et al., "The WRKY transcription factor family in *Brachypodium distachyon*," *BMC Genomics*, vol. 13, article 270, 2012.
- [65] X. Wang, X. Shi, B. Hao, S. Ge, and J. Luo, "Duplication and DNA segmental loss in the rice genome: implications for diploidization," *New Phytologist*, vol. 165, no. 3, pp. 937–946, 2005.
- [66] C. I. Castillo-Davis, S. L. Mekhedov, D. L. Hartl, E. V. Koonin, and F. A. Kondrashov, "Selection for short introns in highly expressed genes," *Nature Genetics*, vol. 31, no. 4, pp. 415–418, 2002.
- [67] J. Li, J. Ding, W. Zhang et al., "Unique evolutionary pattern of numbers of gramineous NBS-LRR genes," *Molecular Genetics and Genomics*, vol. 283, no. 5, pp. 427–438, 2010.
- [68] L. Deslandes, J. Olivier, N. Peeters et al., "Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 13, pp. 8024–8029, 2003.
- [69] L. Deslandes, J. Olivier, F. Theuilières et al., "Resistance to *Ralstonia solanacearum* in *Arabidopsis thaliana* is conferred by the recessive RRS1-R gene, a member of a novel family of resistance genes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 4, pp. 2404–2409, 2002.
- [70] J. L. Dangl and J. D. G. Jones, "Plant pathogens and integrated defence responses to infection," *Nature*, vol. 411, no. 6839, pp. 826–833, 2001.
- [71] J. Zhang, "Evolution by gene duplication: an update," *Trends in Ecology and Evolution*, vol. 18, no. 6, pp. 292–298, 2003.