

Genome-Wide Identification and Expression Analysis of WRKY Gene Family in *Nelumbo nucifera*

Yidan Liu

Nanjing Agricultural University

Yingchun Xu

Nanjing Agricultural University

Huan Liu

Nanjing Agricultural University

Bei Wang

Nanjing Agricultural University

Xiao Zhang

Nanjing Agricultural University

Yanjie Wang

Nanjing Agricultural University

Qijiang Jin

jqj@njau.edu.cn

Nanjing Agricultural University

Research Article

Keywords: Phylogenetic tree, Gene structure, evolution, expression pattern, stress response

Posted Date: November 10th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3327207/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

WRKY transcription factors play important roles in various stress responses in diverse plant species. We identified a total of 54 WRKY genes (NnWRKY) in the lotus genome. All NnWRKY genes were found to be scattered in 14 of 50 scaffolds. A phylogenetic tree was constructed and 54 NnWRKY genes were classified into 5 separate monophyletic groups. Gene structure analysis indicates that WRKY IIa might be the most ancestral in the lotus WRKY gene family. dN/dS ratios were always less than 1 for both WRKY domains and higher than 1 in amino acid sequence outside the domain, implying the relaxed purifying or positive selection on this region. Six amino acid residues were identified in the NnWRKY domain, which presumably have contributed to the functional divergence of lotus WRKY genes. We tested two pairs of codon substitution models to test positive selection using the likelihood ratio test and one positive selection site (625G, corresponding to NNU_013102) was identified. Moreover, gene conversion events might have played a significant role during the evolutionary history of the lotus WRKY family. Expression profiling revealed that NnWRKY genes might primarily have functions in sink tissues. Several NnWRKY genes were found to be changed obviously and might have roles in rhizome development, energy metabolism and stress responses. The present study provides a foundation for further functional studies of NnWRKYs in lotus.

Introduction

Transcription factors (TFs) are important regulatory proteins that regulate gene expression temporally and dynamically via binding to specific DNA sequences of genes^[1,2]. Many of these genes are found to be associated with numerous plant signaling and regulatory networks related stress responses^[3,4]. The WRKY protein family is one of the largest families of transcription factors in plants^[5~7]. They are first identified in plants and named due to their highly conserved WRKYGQK sequence at the N-terminus of the proteins, as well as a conserved metal chelating zinc finger motif^[8,9]. The conserved elements of the domain are directly involved in binding to the consensus sequence, TGAC(C/T), of the W-box cis-acting element^[1]. In addition, the WRKY domain also binds to a sugar responsive cis-element, SURE. Previous work often divided them into three main groups according to the number of WRKY domains and the pattern of the zinc-finger motif. Proteins with two WRKY domains and a zinc finger motif pattern of C2-H2 (C-X4-5-C-X22-23-H-X1-H) belong to group I. The motif also existed in Group II WRKY genes^[10]. In contrast, group II members containing a single WRKY domain and are often subdivided into 3 to 5 subgroups on the variation of conserved structural motifs. Group 3 proteins also have a single WRKY domain, but have a different zinc finger motif pattern of C2-HC (C-X7-C-X23-H-X1-C).

Since the first WRKY genes were cloned from sweet potato^[11], more WRKY genes have been identified from other plants^[10~13]. In recent years, interest in WRKY transcription factors has increased, because there is considerable evidence showing that WRKY proteins play central roles in various aspects of biological processes and adaptation to the environment^[5,6]. Expression and genetic analysis showed that they were key regulators in plant response to abiotic stresses^[14~17], i.e. salt^[18], wounding^[19], heat and

drought^[5,10,20~22], and biotic stresses, i.e. infection by bacterial^[23], fungal^[24], and viral pathogens^[25]. Research has also focused on additional roles of WRKY genes in plant development processes such as germination, senescence, secondary wall formation and trichome development^[6]. A large number of loss-of function knockout/down mutants or overexpression lines of WRKY genes have been produced and showed the positive or negative roles of WRKY TFs in plant defense responses and development processes. Transgenic tomato overexpressing WRKY39 exhibit enhanced resistance to salt and drought stresses and PstDC3000 infection, possibly via activating the expression of pathogenesis- and stress-related genes^[26]. Arabidopsis, lacking WRKY41, showed reduced primary seed dormancy and thermoinhibition, while transgenically overexpressing WRKY41 increases ABI3 expression which play an essential role in the establishment of seed dormancy^[27]. In rice, overexpression of OsWRKY89 resulted in reduction in the growth and internode length^[8,22,28]. All these results suggested that numerous members in WRKY family facilitate the evolution and adaption of plant to complex environmental and developmental systems.

Lotus is an economically important aquatic plant for food, medicine, and flower. It is planted world-wide and have special meaning in Asia culture. As lotus shows fast growth and can improve water quality, it was intensively studied for the purpose of biological cleaning. The genome of lotus has also been sequenced years ago. In this study, we performed a comprehensive characterization and analyses of WRKY gene family from lotus, including gene sequence, structure, conserved motifs and evolution. Using available transcriptome datasets, we demonstrated their expression patterns in various tissues, developmental stages and stress conditions. The current results could provide valuable information for the future work on the function of WRKYs in lotus.

Materials and Methods

Identification of lotus WRKY genes.

To identify lotus WRKYs at the genome level, nucleotide acid and amino acid sequence of all protein-coding genes were obtained from the latest version of the lotus genome (Genome ID: 16884) deposited in COGE (<https://genomeevolution.org/coge/GenomeInfo.pl?gid=16884>). Hidden Markov model (HMM) of the WRKY domain (PF03106) downloaded from the PFAM database (<http://pfam.xfam.org/>) was used to identify putative *WRKY* genes using HMMER v. 3.0^[29] with an E-value cut-off of $< 1e-10$. Only the longest variant of each gene was selected and assigned as lotus *WRKY* genes (NnWRKY). The gene structures were visualized using gene model obtained from lotus genome data (Genome ID: 16884) in gene structure display server (GSDS, <http://gsds.cbi.pku.edu.cn/>)^[30].

Multiple sequence alignment and phylogenetic analysis.

To investigate the evolutionary relationships among NnWRKYs, the amino acid or nucleotide sequences of *WRKY* genes were aligned using MEGA program (v7), and sequence identity and similarity were also calculated. Non-synonymous (d_N) and synonymous (d_S) substitution of different region of WRKY gene

were calculated using PAML version 4. The alignment of 54 NnWRKY nucleotide sequences was used to reconstruct the phylogeny tree by the Bayesian method with MrBayes software package version 3.1.2^[31]. Four independent runs were analyzed, each using the general time reversible (GTR) model plus gamma distributed rates (GTR + I + G), as determined by Modeltest version 3.7^[32]. One millions generations of Markov Chain Monte Carlo simulations were run with four separate chains. Trees were summarized and compared to check for concordant topologies. The consensus tree of all compatible groupings among all runs was used in the following analyses.

Protein property and conserved motif analysis.

The molecular weight (MW), and isoelectric point (*pI*) of NnWRKYs were calculated using ProParam (<http://web.expasy.org/protparam>). Analysis and visualization of the conserved motifs in WRKY proteins was performed using MEME (<http://meme.sdsc.edu/meme/cgi-bin/meme.cgo>).

Molecular Evolution Analysis.

Analysis for the functional divergence between groups were carried out using the program DIVERGE^[33]. The coefficient of type I (θ_I) and type II (θ_{II}) functional divergence, and the likelihood ratio test statistics between any two groups were calculated. The likelihood test can be approximated to a χ^2 distribution with 1 degree of freedom. A significant θ_I means altered functional constraints between duplicated genes. A significant θ_{II} indicates radical change in amino acid property, but no altered functional constraints^[34~36]. The sites contributed to the functional divergence were predicted according to their posterior probabilities (Q_k) and a cut-off value, $Q(k) \geq 0.95$, was used empirically.

Evaluation of positive selection was performed with the program codeml from PAML version 4^[37], using a variety of models of codon sequence evolution. Two pairs of models were explored: the one-ratation model (M0) and the discrete model (M3), and the beta model (M7) and beta + ω model (M8). We detected variation in ω values among sites by employing a likelihood ratio test (LRT) between M0 vs. M3 and M7 vs. M8 according to Yang et al.^[38]. For χ^2 statistics, the freedom was 3 for M0/M3 tests and 2 for M7/M8 tests^[39]. If the LRT was statistically significant, we identified the positively selected sites under model M8 using the Bayesian method.

Analysis for conserved motifs in NnWRKY proteins was carried out using the program GeneConv^[40] with default setting. GeneConv finds and ranks the highest scoring fragments which have potentially undergone gene conversion by *P* values globally and pairwise^[41]. Global and pairwise fragments were all included in the results.

In Silicon Expression Analysis.

Take advantage of available transcriptome data, the expression patterns of NnWKRY genes were analyzed. The transcriptome data used in this analysis including rhizome apical meristem (SRS413495),

rhizome elongation zone (SRS413496), rhizome internode (SRS414456, SRS414455, SRS414457, SRS414458, and SRS414459), leaf (SRS413489, SRS413488, SRS413490, SRS413492, and SRS413491), petiole (SRS413501, SRS413498, SRS413500, SRS413499, and SRS413502), root (SRS412280, SRS412272, SRS412279, SRS412273, and SRS412278), different rhizome develop stage (stolon: SRR2052538, SRR2052542; middle swelling stage: SRR2052568, SRR2052569; later swelling stage: SRR2052570, SRR2052573) and heat treatment (SRP070871). The library preparation methods and statistics of sequencing data were detailed in Kim et al.^[42], Liu et al.^[43], and Yang et al.^[44]. Our in house transcriptome data of lotus under copper and submergence stresses were used. The sequencing libraries were constructed from 3-month-old lotus seedlings with or without complete submergence treatment (24 h) or 5 mM copper treatment (24 h). Library construction and sequencing (without replicates) were performed at the Beijing Genomics Institute (BGI, Shenzhen, China), following the manufacture's protocols.

After converted into FASTQ format using the SRA toolkit, all raw read were filtered with NGS QC Toolkit (v2.3.3) to get clean reads. We indexed the genome of lotus with Bowtie2 and mapped the clean reads to the reference genome using TopHat. The expression levels were shown as fragments per kb per million mapped read (FPKM) determined by cufflinks program (cufflinks v2.2.142) with default settings.

Results

Characterization of 54 WRKY-encoding sequences in lotus.

Accumulating evidence suggested the important roles of *WRKY* genes in modulating the expression of genes related to plant's stimuli response and development processes^[45]. For instance, in soybean and *Arabidopsis*, studies suggested that GmWRKY21, 31, and 54, and AtWKRY25, and 33 confer differential tolerance to stressful environment respectively^[46]. OsWRKY78 are reported to regulates stem elongation and seed development in rice^[47]. However, *WRKY* gene family in lotus has so far been poorly described. Completion of the lotus genome by Ming et al.^[48] makes it possible to carry out genome-wide study on *WRKY* gene family of lotus. WRKY proteins displayed highly conservation in thier WRKY domain^[49] across different plant species. Thus, HMMER searches were performed to search local lotus protein data with HMM profile of conserved WRKY domains as queries^[50]. A total of 54 lotus WRKYs (NnWRKYs) genes which contained complete WKRY domain were identified in lotus genome (Table 1). Using identified genes as queries, we further performed a homology search from amino acid sequences of lotus to identify new potentially WKRY encoding genes, which however resulted in no new WKRY genes. Domain analysis confirmed the presence of one or two WKRY domains in the predicted WKRY proteins. The properties of identified NnWRKY proteins were characterized and listed in Table 1, including protein length, molecular weight, isoelectric point, and motifs. The identified NnWRKYs contained a diverse number of amino acid from 106 (NNU_023323) to 897 (NNU_008419) in length with an average length of 407. The protein weight ranged from 12.21 (NNU_023323) to 98.71 (NNU_008419) kDa. The isoelectric point (PI) varied from 4.57 (NNU_021524) to 9.80 (NNU_020936), with 32 numbers

have a $pI < 7$ and others have a $pI > 7$. All *NnWRKY* genes were found to scattered in 14 of 50 scaffolds. We found that most *NnWRKY* genes were located in the scaffold 1. This uneven distribution of *NnWRKY* genes on lotus scaffolds might be related to the scaffold length. The number of *WKRY* gene family were relative variable among different species, i.e. there were 102 *WRKY* genes in *Oryza sativa* L. ssp. *Japonica* ^[51], 116 in *Gossypium raimondii* ^[52], 37 in *Physcomitrella Patens* ^[49], which indicating that duplicated copies of *WKRY* genes are frequently retained after a whole-genome duplication event.

Table 1
WRKY transcript factor family in lotus.

NnWRKY (Genebank accession)	Deduced		WRKY domain			Megascaffold
	polypeptide		pI	Zinc- Finger Type	Domain Number	
	Length (aa)	MW (kDa)				
NNU_005293	743	80.25	5.6	C ₂ H ₂	2	1
NNU_005576	461	50.53	9.1	C ₂ H ₂	2	3
NNU_009557	473	52.36	8.99	C ₂ H ₂	2	3
NNU_013081	565	61.64	6.74	C ₂ H ₂	2	3
NNU_019405	608	66.79	5.64	C ₂ H ₂	2	2
NNU_000906	538	58.50	7.41	C ₂ H ₂	2	4
NNU_008419	897	98.71	5.95	C ₂ H ₂	2	5
NNU_013102	629	69.51	7.66	C ₂ H ₂	2	3
NNU_014662	785	84.71	5.82	C ₂ H ₂	2	18
NNU_007725	588	64.27	6	C ₂ H ₂	1	3
NNU_012242	693	75.89	8.6	C ₂ H ₂	1	3
NNU_024662	496	55.04	5.42	C ₂ H ₂	1	10
NNU_005172	307	34.02	6.68	C ₂ H ₂	1	1
NNU_015799	493	55.50	6.42	C ₂ H ₂	1	1
NNU_002747	303	32.89	4.98	C ₂ H ₂	1	2
NNU_022759	296	32.11	6.06	C ₂ H ₂	1	9
NNU_019387	313	34.83	7.19	C ₂ H ₂	1	2
NNU_000764	576	62.37	6.22	C ₂ H ₂	1	4
NNU_022872	622	66.90	5.9	C ₂ H ₂	1	5
NNU_005007	613	66.19	6.35	C ₂ H ₂	1	1

NnWRKY (Genebank accession)	Deduced polypeptide			WRKY domain		Megascaffold
	Length (aa)	MW (kDa)	pI	Zinc- Finger Type	Domain Number	
NNU_024102	685	74.89	6.26	C ₂ H ₂	1	1
NNU_021907	497	54.44	7.31	C ₂ H ₂	1	33
NNU_025395	503	56.43	7.34	C ₂ H ₂	1	1
NNU_024663	526	57.63	5.86	C ₂ H ₂	1	10
NNU_000937	311	34.28	6.39	C ₂ H ₂	1	4
NNU_008364	308	33.93	6.78	C ₂ H ₂	1	5
NNU_001380	321	35.45	8.74	C ₂ H ₂	1	13
NNU_023785	327	36.04	7.35	C ₂ H ₂	1	8
NNU_009014	299	33.30	5.84	C ₂ H ₂	1	9
NNU_014445	310	34.51	6.21	C ₂ H ₂	1	2
NNU_010128	128	14.55	9.4	C ₂ H ₂	1	6
NNU_023323	106	12.21	9.22	C ₂ H ₂	1	1
NNU_003142	184	20.96	9.37	C ₂ H ₂	1	5
NNU_004050	183	20.82	9.49	C ₂ H ₂	1	4
NNU_010326	196	22.17	9.4	C ₂ H ₂	1	1
NNU_007649	303	33.67	6.21	C ₂ H ₂	1	3
NNU_011033	302	34.01	5.72	C ₂ H ₂	1	3
NNU_008704	216	24.62	8.19	C ₂ H ₂	1	2
NNU_025262	222	25.41	7.77	C ₂ H ₂	1	8
NNU_010097	223	25.88	9.19	C ₂ H ₂	1	6
NNU_014805	158	18.97	9.65	C ₂ H ₂	1	1

NnWRKY (Genebank accession)	Deduced polypeptide			WRKY domain		Megascaffold
	Length (aa)	MW (kDa)	pI	Zinc- Finger Type	Domain Number	
NNU_009041	282	32.07	4.94	C ₂ H ₂	1	9
NNU_014473	281	32.00	5.67	C ₂ H ₂	1	2
NNU_010321	365	40.11	5.07	C ₂ H ₂	1	1
NNU_021524	288	31.65	4.57	C ₂ H ₂	1	15
NNU_011577	358	39.41	9.51	C ₂ H ₂	1	1
NNU_020936	360	40.55	9.8	C ₂ H ₂	1	1
NNU_004768	326	36.50	6.14	C ₂ HC	1	15
NNU_005350	329	37.06	5.46	C ₂ HC	1	3
NNU_004769	355	38.99	5.37	C ₂ HC	1	15
NNU_007278	686	78.85	8.72	C ₂ HC	1	16
NNU_010416	362	40.55	6.24	C ₂ HC	1	1
NNU_013507	363	40.37	5.27	C ₂ HC	1	4
NNU_013509	362	40.56	5.51	C ₂ HC	1	4

Phylogenetic analysis of lotus WKRY genes.

To study the evolutionary relationships between lotus *WRKY* genes, a phylogenetic tree of 54 *NnWRKY* genes was created based on multiple sequence alignment of nucleotide sequence data ^[53](Huelsenbeck et al., 2001). Using Modeltest 3.7 ^[32], we first test the best-fitting nucleotide substitution model for phylogenetic tree construction using Bayesian method, and GTR + G + I were selected. Based on the statistical support values and previous works ^[7, 10], 54 *NnWRKY* genes were classified into 5 separate monophyletic groups, termed WKRY I (14), WKRY IIa (10), WKRY IIb (6), WKRY IIc (17) and WKRY III (7) (Fig. 1). This classification was also consistent with the previous described classification based on the presence of WKRY and zinc finger domains ^[51, 52, 54]. For example, the members that contain two WKRY domains and feature the C₂H₂-type zinc finger motif (N: Cx₄Cx₂₂₋₂₃ HxH; C: Cx₄Cx₂₃HxH) were categorized into the group I. However, there were also some exceptions, i.e. some *NnWRKY* proteins in

group I which carries one WRKY domain appeared to be more related to the group. The same phenomenon was also found in *Arabidopsis*^[7], which might attribute to the loss of the WRKY domain, as domain gain or loss appear to have shaped the WKRY family^[55]. These findings revealed the existence of five variants in NnWKRY genes recovery as five distinct clades in the Bayesian reconstruction.

Gene structure analysis of NnWRKYs.

As gene structures were an important part in assessing the evolution of *NnWRKY* gene family, we then examined the intron/exon distribution of *NnWRKY* genes^[17, 19, 56]. In spite of the highly conserved WKRY domains, the length of these full sequences is highly variable and even more distinct in their gene sequences (Fig. 2). All *NnWRKY* genes contain at least one intron in their CDSs: 6 have one intron; 23 have two introns; 5 have three introns; 11 have four introns; 6 have five introns. Besides, NNU_013102, NNU_024102, and NNU_008419 were more complicated, which have 6, 8 and 10 introns respectively (Fig. 2). Introns were particularly abundant in members of WKRY IIa, which has an average intron number of 4.9, followed by it was WKRY I which have 4. In contrast, WKRY IIb, IIc and III only have an average number of introns approximate to 2 (Fig. 3). During the evolution of intron, intron loss was more frequently than intron gain following segmental duplication^[57~59], which indicated that WKRY IIa might be the most ancestral that other *NnWRKY* genes might derive from it.

Conserved sequences in WKRY proteins.

A multiple sequence alignment of 54 full-length NnWKRY sequences showed that nucleotide identity among them varied from 23.53% (between NNU_016707 and NNU_022788) to 91.04% (Fig. 4). Their full length amino acid sequence was even more variable, ranged from 4.35 to 91.35%. However, despite their full gene or protein sequences were highly distinct, several residues in the WKRY domain were conserved in all *NnWRKY* genes, implied that corresponding genes encoding active enzymes. In contrast, amino acid sequences outside of the WKRY domain were the most divergence sequence.

We then calculated ratio (ω) of the synonymous substitution rate (d_S) versus the non-synonymous substitution rate (d_N) among different regions of WKRY protein, including full-length sequence, C-terminal WKRY domain, N-terminal WKRY domain and sequence outside WKRY domain. It was found that d_N/d_S ratios were always less than 1 for both WKRY domains, suggesting purifying selection on these regions which could maintain the function conservation of *WRKY* genes^[60, 61]. In contrast, the ratio was much higher in regions outside WKRY domain with most d_N/d_S ratios higher than 1 (Fig. 5), implying the relaxed purifying or positive selection on this region^[60, 61].

Further analysis of the sequence feature of NnWRKY proteins using software MEME^[62] with default settings, predicted 20 conserved motifs. We listed the sequence of these motifs (Fig. S1) and visualized their distribution in Fig. 6. As shown in Fig. 6, motifs 1, 5, and 14 contained the characterized sequence (WRKYGQK) of WKRY domain, and at least one of them are presented in every sequence. Besides, one or more conservation motifs outside of the WKRY domain can be detected in a WKRY protein: Motifs 4 and

15 were found in most members of WRKY I; motif 9 was found in the WRKY IIa; motif 15 was identified in most members of WRKY IIc; motif 3 was unique in WRKY III. Generally, WRKY motif compositions were more similar between members in the same group, indicating functional similarity among them.

Analysis of functional divergence.

Functional divergence after gene duplication has been known as an important source of evolutionary novelty^[63]. After gene duplication, one gene copy is free to change as it will not cause function deficit of the gene, and accumulating amino acid changes will contribute to functional divergence. To investigate whether amino acid substitutions in the NnWKRY sequences could have caused adaptive functional diversification, type I and type II functional divergence between groups of the *WRKY* gene family were estimated by posterior analysis^[33]. Type I functional divergence represents altered functional constraints (e.g. different evolutionary rates) between duplicate genes, while type II indicate no altered functional constraints but radical change in amino acid (i.e., charge, hydrophobic)^[33]. Our results showed that type I coefficients (θ_I) of functional divergence in WRKY I/WRKY IIb and WRKY IIa/WRKY IIb were significantly greater than zero ($P < 0.05$). Meanwhile, we did not find any evidence for type II functional divergence between groups (Table 2), indicated that type I functional divergence was the dominant pattern for the evolution of WRKY family in lotus. Further, we predicted amino acid residues which are potentially responsible for the functional divergence on the base of sites specific profiles. To greatly reduce false positive, a cut-off value, $Q(k) \geq 0.95$, was used empirically as cutoffs for estimated key amino acid residues. Six amino acid residues (875, 878, 879, 881, 912 and 914) in all comparisons were identified as being most important for the functional divergence (Fig. 7). These amino acids were all identified in the WRKY domain and presumably have contributed to the functional divergence of lotus WKRY genes.

Table 2
Functional divergence between groups of the *NnWRKY* genes.

Type I	Type II						
	$\theta_1 \pm \text{SE}^z$	LRT	P	$Q_k \geq 0.80^y$	$Q_k \geq 0.95^x$	$\theta_{II} \pm \text{SE}$	P
WRKY I/WRKY IIa	0.354 ± 0.248	2.038	0.153	1	0	-0.939 ± 1.240	0.449
WRKY I/WRKY III	0.388 ± 0.379	1.051	0.305	0	0	-1.103 ± 2.005	0.582
WRKY I/WRKY IIb	0.55 ± 0.277	3.936	0.047	1	1	-2.390 ± 3.233	0.46
WRKY I/WRKY IIc	0.001 ± 0.022	0	0.964	0	0	-1.028 ± 1.297	0.428
WRKY IIa/WRKY III	0.762 ± 0.47	2.628	0.105	3	1	-0.423 ± 1.357	0.755
WRKY IIa/WRKY IIb	0.942 ± 0.463	4.148	0.042	11	5	-1.613 ± 2.491	0.483
WRKY IIa/WRKY IIc	0.289 ± 0.225	1.654	0.199	1	0	-0.845 ± 0.908	0.352
WRKY III/WRKY IIb	0.001 ± 0.022	0	0.964	0	0	nan	nan
WRKY III/WRKY IIc	0.001 ± 0.022	0	0.964	0	0	-0.297 ± 1.23	0.81
WRKY IIb/WRKY IIc	0.685 ± 0.319	4.6	0.968	3	0	-1.466 ± 2.351	0.533
^z The coefficient of functional divergence between two subgroups and its standard error.							
^y The number of critical amino acid residues with posterior prob-ability (Q_k) > 0.80.							
^x The number of critical amino acid residues with posterior probability (Q_k) > 0.95.							

Detection of positive selection and gene conversion events.

ML methods and two pairs of codon substitution models in PAML were chosen to test positive selection using likelihood ratio test (LRT) [38, 63, 64]. First, we compared one ratio model (M0) with the discrete model (M3) to evaluate the variations in d_N/d_S ratio among codon positions below each group. Overall, testing the selective pressure with the basic model M0 showed that the ML estimates of d_N/d_S values for all groups were close to zero, with a mean of 0.281, indicating that each group of *NnWRKY* genes have been under purifying selection (Table 3). Interestingly, the log likelihood differences between model M3 and M0 were statistically significant for all groups, revealing that the evolutionary forces acting on

individual codon site are fluctuated and some genes might have been under positive selection. Then, the presence of codons that evolved under positive selection was further tested by comparing the M8 and M7 models. According to the criterion described previously^[54] group IIa was believed to have undergone positive selection. We identified the positively selected sites under model M8 using the Bayesian method, and one positive selection sites (625G, corresponding to NNU_013102) were identified ^[65, 64].

Table 3
Detection of positive selection among codons of *NnWKRY* under site-specific model.

Group	N ^z	$d_N/d_S(\omega)$ under M0 ^y	2 Δ /M3 vs. M0 ^x	2 Δ /M8 vs. M7 ^x	M8 Estimates ^w	Positively Selected Sites ^v
WRKY I	14	0.356	477.1361**	11.39814**	p1 = 0.05110, ω = 2.92059 p = 0.87912, q = 1.09104	-
WRKY IIa	10	0.31614	1269.635**	15.37348**	p1 = 0.03044, ω = 44.74768, p = 0.56713, q = 0.65863	625 G (NNU_013102)
WRKY IIb	6	0.16165	300.6672**	0.00018	p1 = 0.00001, ω = 2.03772, p = 0.71318, q = 2.67776	-
WRKY IIc	17	0.27169	1358.758**	177.983**	p1 = 0.00001, ω = 1.00000, p = 0.42827, q = 0.55990	-
WRKY III	7	0.30279	228.188**	2.541058	p1 = 0.23920, ω = 1.00000, p = 0.77790, q = 1.87365	--

^zNumber of the sequences in the group.

^y $d_N/d_S(\omega)$ is the average ratio over sites under a codon model with one-ratio.

^xDouble asterisks represent $P < 0.01$ for chi-square (χ^2) test, while single asterisk represents $P < 0.05$.

^w ω is estimated under model M8; p1 is the inferred proportion of positively selected sites; p and q are the parameters of the beta distribution.

^vThe amino acid, which is listed in positively selected position, had a posterior probability > 0.95 under M8, df = 1.

Gene conversion is an important population genetic process which also acted as an important force in evolution [66]. Possible gene conversion events during evolution of *NnWKRY* genes were analyzed using the program GeneConv [40]. Possible gene conversion event between ancestors of two sequences in the

alignment and originated from outside of the alignment were all included in this analysis. The results demonstrated that WRKY I, IIa, IIc and III possessed at least one gene conversion event (Table 4). No gene conversion event was identified from the genes in WRKY IIb group. Most of the gene conversion events (47%) were identified within WRKY I group and they can be detected in 9 members of this group. All these results suggested that gene conversion events might have played significant role during the evolutionary history of lotus WRKY family, which is also found in WKRYs of other species as showed previously^[10].

Gene expression analysis.

Previous work demonstrated that *WRKY* genes serve important functions in development and stress responses of plants^[49]. To gain more information on the role of lotus *WRKY* gene family, in this study, available transcriptome data of lotus, related to different tissues, stresses, and developmental stages, were used to analyze expression patterns of *WRKY* genes^[43, 44, 67, 68].

NnWRKY genes showed differential expression level in various lotus tissues (Fig. 8). Except NNU_004769 which was not detected in the petiole, no more tissue-specific *WRKY* genes were found. Among them, NNU_013509 showed the highest expression level while NNU_004769, NNU_025395, NNU_024662, and NNU_011033 showed the lowest transcript level (with average FPKM < 1) in most examined tissues. Genes from the same group frequently showed similar expression patterns within specific tissues. For example, NNU_005293, NNU_005576, and NNU_009557, which belonging to WRKY Ia, had relative lower expression level in old leaf and petiole (average FPKM < 10) and high expression level in root (average FPKM = 39.60) and rhizome (average FPKM = 72.79). These results indicated that *NnWRKY* genes might primarily have functions in sink tissues.

In response to heat, submergence and copper stresses, the total transcript abundance of *NnWRKY*s did not show apparent difference among different stresses, while Cu treatment induced significantly higher abundance of *NnWRKY* genes than other treatments ($p < 0.05$) (Fig. 9). In response to heat and submergence stresses, *WRKY* genes in lotus are tend to decrease, as evidenced by slightly decreased abundance of *NnWRKY* genes. In response to heat treatment, most *NnWRKY* genes were down-regulated. Some genes, i.e. NNU_019387, NNU_010416, and NNU_013507 were down-regulated by heat for over 2 folds. Only 14 of 54 *NnWRKY* genes were up-regulated by heat stress, with NNU_010326 being the highest upregulated gene ($\log_2(\text{heat/ck}) = 1.72$). Two *NnWRKY* genes (NNU_000937 and NNU_005007) were mostly induced ($\log_2(\text{heat/ck}) > 2$) by submergence treatment, while most *NnWRKY* genes (i.e. NNU_005172 and NNU_008364) were down-regulated. Transcript level of NNU_005172 was reduced to less of 14% of normal condition. It is interesting to note that the homologous gene of NNU_005172 in soybean, WKRY 20, plays important role in drought stress^[69, 70]. Overexpression of soybean WKRY 20 in alfalfa and soybean could enhance drought tolerance of the plant, possibly by regulating ABA signaling^[69, 70]. These results indicate that WKRY 20 might played as a switch to modulate plant response to different kinds of water stress via regulating ABA signaling. Some metabolic processes of plants use copper as an essential micronutrient, which acting as a constituent of enzymes

and other proteins ^[71~75]. Excessive copper has been found to cause oxidation of biomolecules and damage cell membrane ^[76]. Under copper stress, the transcript abundance of most *NnWRKY* genes was increased. Numerous works reported the roles of WKRY gene family members in enhancing plant oxidative stress tolerance ^[77]. WKRY proteins domain have W-box binding site to bind W-box, which is the key regulatory element involving in stress response ^[1]. Therefore, these genes may be the important factors in the copper stress metabolic network of lotus.

Transcriptional expression of *NnWRKY* during lotus rhizome and flower bud development were also investigated (Fig. 10). We did not find significant difference among the overall abundance of *NnWRKY* genes in lotus during various development stages. During different development stage of rhizome, 24 *NnWRKY* genes showed obvious difference (fold change > 2 or specific expressed) from stage 1 to stage 3. Among them, 3 and 13 genes were down-regulated during the stage 2 and 3 in comparison with stage 1 respectively. One and 6 *NnWRKY* genes showed specific expression pattern, that 1 and 5 genes were not detected in stage 2 and 3 respectively. Only two genes in stage 3 were up-regulated for over 2 fold and 3 genes were expressed weakly in stages 2 and 3 (FPKM < 1). During development of lotus flower bud, 4 genes were obviously changed ($\log_2$ fold > 2 or specifically expressed) in compared with stage when the flower bud was not started to form. The expression of NNU_008419 and NNU_024102 decreased for over 2 fold during flower bud development. NNU_013081 was not detected during flower develop stage. Only one gene (NNU_014445) was found to be significantly induced during flower bud development, which although expressed at very low abundance (FPKM < 1)

The gene expression patterns above pointed to potential co-regulated relationships between various stresses and development processes. Thus, we constructed the co-regulated network of *NnWKRY* to various stimuli using the transcriptome datasets described above (Fig. 11). The results showed that stresses up-regulated WKRY genes were commonly down-regulated by various development processes, i.e. NNU_005007, NNU_024102 and NNU_013509. The contrasting expression pattern under stresses and development processes indicated that these WKRY genes might play important roles in lotus energy metabolism and participate in stress response. GO analysis confirmed this conclusion, showed that these genes were mainly involved in plant stresses responses (Fig. S2). It is interesting to note that heat stresses downregulated a series of *WRKY* genes as development did. We speculated that heat stress at 37°C might have promoting effect on lotus development in short time. Some of these obviously changed genes in the network were then validated by semi-quantitative RT-PCR. As shown in Fig. S3, RT-PCR results of genes under different stresses were similar in magnitude to those obtained by deep sequencing, suggesting that the gene expression data generated from the RNA-Seq data applied in this study were reliable. The gene expression patterns of *NnWKRY* genes provides important clues for investigating gene function of lotus *WRKY* gene family.

Conclusion

Collectively, our results characterize the structure, evolutionary and some functional features of the lotus *WRKY* gene family. Based on systematic analysis of lotus genomes, we identified 54 *NnWRKY* genes. According to statistical support values of phylogenetic analysis and previous works, *NnWRKY* genes were further divided into five separate monophyletic groups. Gene structure analysis showed that NNU_008419 might be the most ancestral. *NnWRKY* domains evolved faster than regions outside *WRKY* domain. Six amino acid residues (875, 878, 879, 881, 912 and 914) in all comparisons were identified as being most important for the functional divergence. Of the 5 groups, group IIa was believed to have undergone positive selection. Gene conversion events had been an important force in evolutionary history of lotus *WRKY* family. Expression patterns showed that *NnWRKY* genes might mainly have function in sink tissues, and several genes might play important roles in lotus stress responses and various development processes.

Declarations

Ethics approval and consent to participate

This study does not involve any experiments involving humans or animals, and does not involve any endangered protected species. The laboratory conducting the experiment in this study has been engaged in the collection, evaluation, and research of lotus germplasm resources for a long time, providing material support for this study. In addition, the experiments conducted in this study were conducted in accordance with relevant institutional, national, and international laws and regulations.

Consent for publication

Not applicable

Availability of data and materials

Nucleotide acid and amino acid sequence of all protein-coding genes were obtained from the latest version of the lotus genome (Genome ID: 16884) deposited in COGE (<https://genomeevolution.org/coge/GenomeInfo.pl?gid=16884>).

Competing interests

The authors declare no competing interests.

Funding

This research was funded by the National Natural Science Foundation of China (no. 31971710; 31772346; 32071829); Funding of agricultural science and technology innovation of Jiangsu Province, China (CX(21)3019); A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

Authors' contributions

Y.C.X. and Q.J.J. developed this research. Y.D.L., H.L., B.W., X.Z., Y.J.W. conducted experiments and conducted data analysis. Y.D.L. and H.L. wrote the manuscript. Y.C.X. and Q.J.J. revised the manuscript and provided guidance for this study. All authors read and agree with the final manuscript.

References

1. Brivanlou, A.H., J.E. Darnell, and Jr. 2002. Signal transduction and the control of gene expression. *Science*. 295:813-818.
2. Yao, Q.Y., E.H. Xia, F.H. Liu, and L.Z. Gao. 2015. Genome-wide identification and comparative expression analysis reveal a rapid expansion and functional divergence of duplicated genes in the WRKY gene family of cabbage, *Brassica oleracea* var. *capitata*. *Gene*. 557:35-42.
3. Han, X., D. Kumar, H. Chen, S. Wu, and J.Y. Kim. 2014. Transcription factor-mediated cell-to-cell signaling in plants. *J. Exp. Bot.* 65:1737-1749.
4. Priest, H.D., S.A. Filichkin, and T.C. Mockler. 2009. Cis-regulatory elements in plant cell signaling. *Curr. Opin. Plant Biol.* 12:643-649.
5. Okay, S., E. Derelli, and T. Unver. 2014. Transcriptome-wide identification of bread wheat WRKY transcription factors in response to drought stress. *Molecular Genetics & Genomics*. 289:765-781
6. Rushton, P.J., I.E. Somssich, P. Ringler, and Q.J. Shen. 2010. WRKY transcription factors. *Trends Plant Sci.* 15:247-258.
7. Wang, Q.S., M.H. Wang, X.Z. Zhang, B.J. Hao, S.K. Kaushik, and Y.C. Pan. 2011. WRKY gene family evolution in *Arabidopsis thaliana*. *Genetica*. 139:973-983.
8. Goel, R., A. Pandey, P.K. Trivedi, and M.H. Asif. 2016. Genome-Wide Analysis of the *Musa* WRKY Gene Family: Evolution and Differential Expression during Development and Stress. *Front Plant Sci.* 7:299.
9. Wang, Y., L. Feng, Y. Zhu, Y. Li, H. Yan, and Y. Xiang. 2015. Comparative genomic analysis of the WRKY III gene family in populus, grape, arabidopsis and rice. *Biol. Direct.* 10:1-27.
10. Eulgem, T., P.J. Rushton, S. Robatzek, and I.E. Somssich. 2000. The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* 5:199-206.
11. Ishiguro, S. and K. Nakamura. 1994. 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 β -amylase from sweet potato. *Mol. Genet. Genomics*. 244:563-571.
12. Mangelsen, E., J. Kilian, K.W. Berendzen, U.H. Kolukisaoglu, K. Harter, C. Jansson, and D. Wanke. 2008. Phylogenetic and comparative gene expression analysis of barley (*Hordeum vulgare*) WRKY transcription factor family reveals putatively retained functions between monocots and dicots. *BMC Genomics*. 9:1-17.
13. Wu, K.L., Z.J. Guo, H.H. Wang, and J. Li. 2005. The WRKY Family of Transcription Factors in Rice and *Arabidopsis* and Their Origins. *DNA Res.* 12:9-26.

14. Birkenbihl, R.P. and I.E. Somssich. 2011. Transcriptional Plant Responses Critical for Resistance Toward Necrotrophic Pathogens. *Front. Plant Sci.* 2:76.
15. Ishihama, N. and H. Yoshioka. 2012. Posttranslational regulation of WRKY transcription factors in plant immunity. *Curr. Opin. Plant Biol.* 15:431-437.
16. Pandey, S.P. and I.E. Somssich. 2009. The role of WRKY transcription factors in plant immunity. *Plant Physiol.* 150:1648-1655.
17. Song, Y. and J. Gao. 2014. Genome-wide analysis of WRKY gene family in *Arabidopsis lyrata* and comparison with *Arabidopsis thaliana* and *Populus trichocarpa*. *Science Bulletin.* 59:754-765.
18. Shi, W., L. Hao, J. Li, D. Liu, X. Guo, and H. Li. 2014. The *Gossypium hirsutum* WRKY gene GhWRKY39-1 promotes pathogen infection defense responses and mediates salt stress tolerance in transgenic *Nicotiana benthamiana*. *Plant Cell Rep.* 33:483-498.
19. Wang, X., Y. Yan, Y. Li, X. Chu, C. Wu, and X. Guo. 2014. GhWRKY40 , a Multiple Stress-Responsive Cotton WRKY Gene, Plays an Important Role in the Wounding Response and Enhances Susceptibility to *Ralstonia solanacearum* Infection in Transgenic *Nicotiana benthamiana*. *PLoS One.* 9:e93577.
20. Baloglu, M.C., B. Inal, M. Kavas, and T. Unver. 2014. Diverse expression pattern of wheat transcription factors against abiotic stresses in wheat species. *Gene.* 550:117.
21. Mustafa, E., T. Mine, A. Guray, G. Tugba, I. Behcet, I. Emre, I. Emre, and U. Turgay. 2015. Transcriptome analysis of wheat inoculated with *Fusarium graminearum*. *Front. Plant Sci.* 6.
22. Wu, X.L., Y. Shiroto, S. Kishitani, Y. Ito, and K. Toriyama. 2009. Enhanced heat and drought tolerance in transgenic rice seedlings overexpressing OsWRKY11 under the control of HSP101 promoter. *Plant Cell Rep.* 28:21-30.
23. Liu, B., Y.B. Hong, Y.F. Zhang, X.H. Li, L. Huang, H.J. Zhang, D.Y. Li, and F.M. Song. 2014. Tomato WRKY transcription factor SlDRW1 is required for disease resistance against *Botrytis cinerea* and tolerance to oxidative stress. *Plant Sci.* 227:145-156.
24. Kayum, M.A., H.J. Jung, and J.I. Park. 2015. Identification and expression analysis of WRKY family genes under biotic and abiotic stresses in *Brassica rapa*. *Mol. Genet. Genomics.* 290:79-95.
25. Liu, Y., M. Schiff, and S.P. Dineshkumar. 2004. Involvement of MEK1 MAPKK, NTF6 MAPK, WRKY/MYB transcription factors, COI1 and CTR1 in N-mediated resistance to tobacco mosaic virus. *Plant J. Cell Mol. Biol.* 38:800-809.
26. Sun, X.C., Y.F. Gao, H.R. Li, S.Z. Yang, and Y.S. Liu. 2015. Overexpression of SlWRKY39 leads to enhanced resistance to multiple stress factors in tomato. *J. Plant Biol.* 58:52-60.
27. Ding, Z.J., J.Y. Yan, G.X. Li, Z.C. Wu, S.Q. Zhang, and S.J. Zheng. 2014. WRKY41 controls *Arabidopsis* seed dormancy via direct regulation of ABI3 transcript levels not downstream of ABA. *Plant J.* 79:810–823.
28. Wang, H., J. Hao, X. Chen, Z. Hao, X. Wang, Y. Lou, Y. Peng, and Z. Guo. 2007. Overexpression of rice WRKY89 enhances ultraviolet B tolerance and disease resistance in rice plants. *Plant Mol. Biol.* 65:799-815.

29. Eddy, S.R. 1998. HMMER: profile HMMs for protein sequence analysis.
30. Hu, B., J. Jin, A.Y. Guo, H. Zhang, J. Luo, and G. Gao. 2015. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*. 31:1296.
31. Ronquist, F. and J.P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 19:1572-1574.
32. Posada, D. and K.A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics*. 14:817-818.
33. Gu, X. and K.V. Velden. 2002. DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family. *Bioinformatics*. 18:500-501.
34. Gu, X. 2006. A simple statistical method for estimating type-II (cluster-specific) functional divergence of protein sequences. *Mol Biol Evol*. 23:1937-1945.
35. Xun, G. 2003. Functional divergence in protein (family) sequence evolution. *Genetica*. 118:133-141.
36. Yang, Z., Q. Gao, C. Sun, W. Li, S. Gu, and C. Xu. 2009. Molecular evolution and functional divergence of HAK potassium transporter gene family in rice (*Oryza sativa* L.). *Journal of Genetics & Genomics*. 36:161-172.
37. Yang, Z. 2015. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol*. 24:1586-1591.
38. Yang, Z., W.S. Wong, and R. Nielsen. 2005. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol*. 22:1107-1118.
39. Mondragon-Palomino, M. and B.S. Gaut. 2005. Gene conversion and the evolution of three leucine-rich repeat gene families in *Arabidopsis thaliana*. *Mol. Biol. Evol*. 22:2444-2456.
40. Sawyer, S. 1999. GENECONV: a computer package for the statistical detection of gene conversion.
41. Posada, D. 2002. Evaluation of methods for detecting recombination from DNA sequences: empirical data¹⁹.
42. Kim, M.-J., W. Nelson, C.A. Soderlund, and D.R. Gang. 2013. Next-Generation Sequencing-Based Transcriptional Profiling of Sacred Lotus "China Antique". *Trop. Plant Biol*. 6:161-179.
43. Liu, X., F. Du, N. Li, Y. Chang, and D. Yao. 2016. Gene Expression Profile in the Long-Living Lotus: Insights into the Heat Stress Response Mechanism. *PLoS One*. 11.
44. Yang, M., L. Zhu, C. Pan, L. Xu, Y. Liu, W. Ke, and P. Yang. 2015. Transcriptomic Analysis of the Regulation of Rhizome Formation in Temperate and Tropical Lotus (*Nelumbo nucifera*). *Sci. Rep*. 5:180-181.
45. Chen, L., Y. Song, S. Li, L. Zhang, C. Zou, and D. Yu. 2012. The role of WRKY transcription factors in plant abiotic stresses. *Biochim Biophys Acta*. 1819:120-128.
46. Zhou, Q.Y., A.G. Tian, H.F. Zou, Z.M. Xie, G. Lei, J. Huang, C.M. Wang, H.W. Wang, J.S. Zhang, and S.Y. Chen. 2008. Soybean WRKY-type transcription factor genes, GmWRKY13, GmWRKY21, and GmWRKY54, confer differential tolerance to abiotic stresses in transgenic *Arabidopsis* plants. *Plant Biotechnol. J*. 6:486-503.

47. Zhang, C.Q., Y. Xu, Y. Lu, H.X. Yu, M.H. Gu, and Q.Q. Liu. 2011. The WRKY transcription factor OsWRKY78 regulates stem elongation and seed development in rice. *Planta*. 234:541-554.
48. Ming, R., R. VanBuren, Y. Liu, M. Yang, Y. Han, L.T. Li, Q. Zhang, M.J. Kim, M.C. Schatz, M. Campbell, J. Li, J.E. Bowers, H. Tang, E. Lyons, A.A. Ferguson, G. Narzisi, D.R. Nelson, C.E. Blaby-Haas, A.R. Gschwend, Y. Jiao, J.P. Der, F. Zeng, J. Han, X.J. Min, K.A. Hudson, R. Singh, A.K. Grennan, S.J. Karpowicz, J.R. Watling, K. Ito, S.A. Robinson, M.E. Hudson, Q. Yu, T.C. Mockler, A. Carroll, Y. Zheng, R. Sunkar, R. Jia, N. Chen, J. Arro, C.M. Wai, E. Wafula, A. Spence, Y. Han, L. Xu, J. Zhang, R. Peery, M.J. Haus, W. Xiong, J.A. Walsh, J. Wu, M.L. Wang, Y.J. Zhu, R.E. Paull, A.B. Britt, C. Du, S.R. Downie, M.A. Schuler, T.P. Michael, S.P. Long, D.R. Ort, J.W. Schopf, D.R. Gang, N. Jiang, M. Yandell, C.W. dePamphilis, S.S. Merchant, A.H. Paterson, B.B. Buchanan, S. Li, and J. Shen-Miller. 2013. Genome of the long-living sacred lotus (*Nelumbo nucifera* Gaertn.). *Genome Biol.* 14:R41.
49. Zhang, J. 2009. Review on the function of WRKY transcription factor family. *Acta Bot. Boreali-Occidentia Sinica*. 29:2137-2145.
50. Eddy, S.R. 2001. HMMER: Profile hidden Markov models for biological sequence analysis. *Comput Chem.* 24:43-55.
51. Ross, C.A., Y. Liu, and Q.X.J. Shen. 2007. The WRKY gene family in rice (*Oryza sativa*). *J. Integr. Plant Biol.* 49:827-842.
52. Dou, L.L., X.H. Zhang, C.Y. Pang, M.Z. Song, H.L. Wei, S.L. Fan, and S.X. Yu. 2014. Genome-wide analysis of the WRKY gene family in cotton. *Mol Genet Genomics*. 289:1103-1121.
53. Huelsenbeck, J.P., F. Ronquist, R. Nielsen, and J.P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*. 294:2310-2314.
54. Ling, J., W. Jiang, Y. Zhang, H. Yu, Z. Mao, X. Gu, S. Huang, and B. Xie. 2011. Genome-wide analysis of WRKY gene family in *Cucumis sativus*. *BMC Genomics*. 12:471.
55. Rossberg, M., K. Theres, A. Acarkan, R. Herrero, T. Schmitt, K. Schumacher, G. Schmitz, and R. Schmidt. 2001. Comparative sequence analysis reveals extensive microcolinearity in the lateral suppressor regions of the tomato, Arabidopsis, and Capsella genomes. *Plant Cell*. 13:979-988.
56. Boudet, N., S. Aubourg, C. Toffanionioche, M. Kreis, and A. Lecharny. 2001. Evolution of intron/exon structure of DEAD helicase family genes in Arabidopsis, Caenorhabditis, and Drosophila. *Genome Res*. 11:2101-2114.
57. Hu, W., H.B. Yang, Y. Yan, Y.X. Wei, W.W. Tie, Z.H. Ding, J. Zuo, M. Peng, and K.M. Li. 2016. Genome-wide characterization and analysis of bZIP transcription factor gene family related to abiotic stress in cassava. *Sci. Rep.* 6.
58. Lin, H., W. Zhu, J.C. Silva, X. Gu, and C.R. Buell. 2006. Intron gain and loss in segmentally duplicated genes in rice. *Genome Biol.* 7:R41.
59. Nuruzzaman, M., R. Manimekalai, A.M. Sharoni, K. Satoh, H. Kondoh, H. Ooka, and S. Kikuchi. 2010. Genome-wide analysis of NAC transcription factor family in rice. *Gene*. 465:30-44.
60. Song, J., Z.H. Gao, X.M. Huo, H.L. Sun, Y.S. Xu, T. Shi, and Z.J. Ni. 2015. Genome-wide identification of the auxin response factor (ARF) gene family and expression analysis of its role associated with

- pistil development in Japanese apricot (*Prunus mume* Sieb. et Zucc). *Acta Physiologiae Plantarum* 37.
61. Wang, X., X. Shi, B. Hao, S. Ge, and J. Luo. 2005. Duplication and DNA segmental loss in the rice genome: implications for diploidization. *New Phytol* 165:937-946.
 62. Bailey, T.L. and C. Elkan. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers.
 63. Yang, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology & Evolution* 15:568-573.
 64. Yang, Z. and J.P. Bielawski. 2000. Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution*. 15:496-502.
 65. Nielsen, R. and Z. Yang. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics*. 148:929-936.
 66. Sawyer, S. 1989. Statistical tests for detecting gene conversion. *Mol. Biol. Evol.* 6:526-538.
 67. Cheng, L., S. Li, J. Yin, L. Li, and X. Chen. 2013. Genome-Wide Analysis of Differentially Expressed Genes Relevant to Rhizome Formation in Lotus Root (*Nelumbo nucifera* Gaertn). *PLoS One*. 8:e67116.
 68. Zhang, Y., S.H. Yang, Y.J. Song, and J.H. Wang. 2014. Genome-wide characterization, expression and functional analysis of CLV3/ESR gene family in tomato. *BMC Genomics*. 15.
 69. Luo, X., X. Bai, X. Sun, D. Zhu, B. Liu, W. Ji, H. Cai, L. Cao, J. Wu, and M. Hu. 2013. Expression of wild soybean WRKY20 in *Arabidopsis* enhances drought tolerance and regulates ABA signaling. *J. Exp. Bot.* 64:2155.
 70. Tang, L., H. Cai, H. Zhai, X. Luo, Z. Wang, L. Cui, and X. Bai. 2014. Overexpression of Glycine soja WRKY20 enhances both drought and salt tolerance in transgenic alfalfa (*Medicago sativa* L.). *Plant Cell Tiss. Org* 118:77-86.
 71. Hall, J.L. 2002. Cellular Mechanisms for Heavy Metal Detoxification and Tolerance. *Journal of Experimental Botany*. 53:1-11.
 72. Li, H.L., D. Guo, Z.P. Yang, X. Tang, and S.Q. Peng. 2014. Genome-wide identification and characterization of WRKY gene family in *Hevea brasiliensis*. *Genomics*. 104:14-23.
 73. Maksymiec, W. 1998. Effect of copper on cellular processes in higher plants. *Photosynthetica*. 34:321-342.
 74. Martins, L.L., M.P. Mourato, S. Baptista, R. Reis, F. Carneiro, A.M. Almeida, P. Fevereiro, and A. Cuypers. 2014. Response to oxidative stress induced by cadmium and copper in tobacco plants (*Nicotiana tabacum*) engineered with the *trehalose-6-phosphate synthase* gene (*AtTPS1*). *Acta Physiol Plant*. 36:755-765.
 75. Yruea, I. 2009. Copper in plants: acquisition, transport and interactions. *Funct. Plant Biol.* 36:409-430.

76. Mithöfer, A., B. Schulze, and W. Boland. 2004. Biotic and heavy metal stress response in plants: evidence for common signals. *FEBS Lett. Febs Letters*. 566:1-5.
77. Gong, X.Q., J.B. Hu, and J.H. Liu. 2014. Cloning and characterization of FcWRKY40 , A WRKY transcription factor from *Fortunella crassifolia* linked to oxidative stress tolerance. *Plant Cell Tiss. Org.* 119:1-14.

Figures

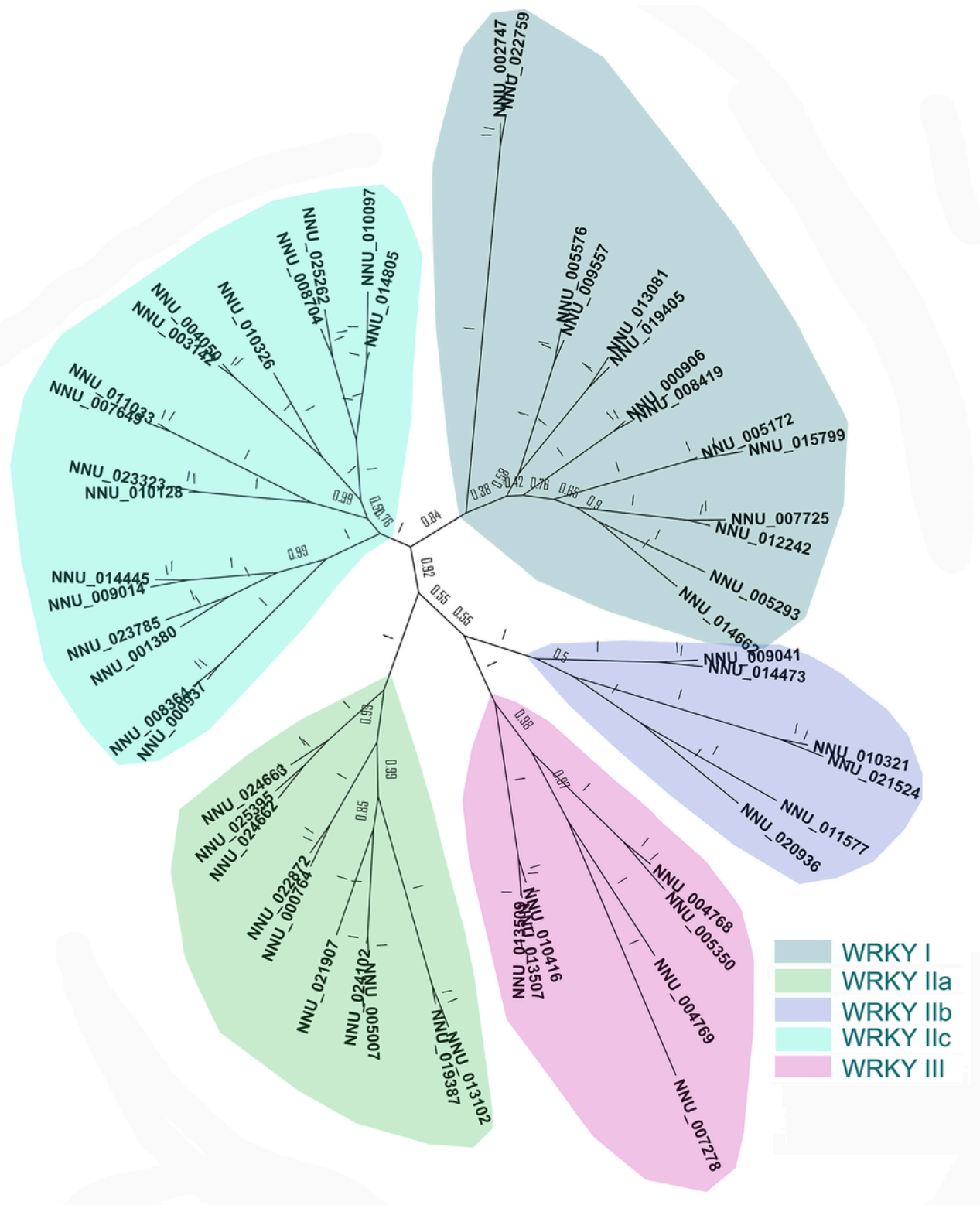


Figure 1

Phylogenetic relationships of DNA sequences within the NnWRKY family. Phylogenetic tree based on the nucleotide sequence data. The number indicate the Bayesian probabilities for each phylogenetic clade.

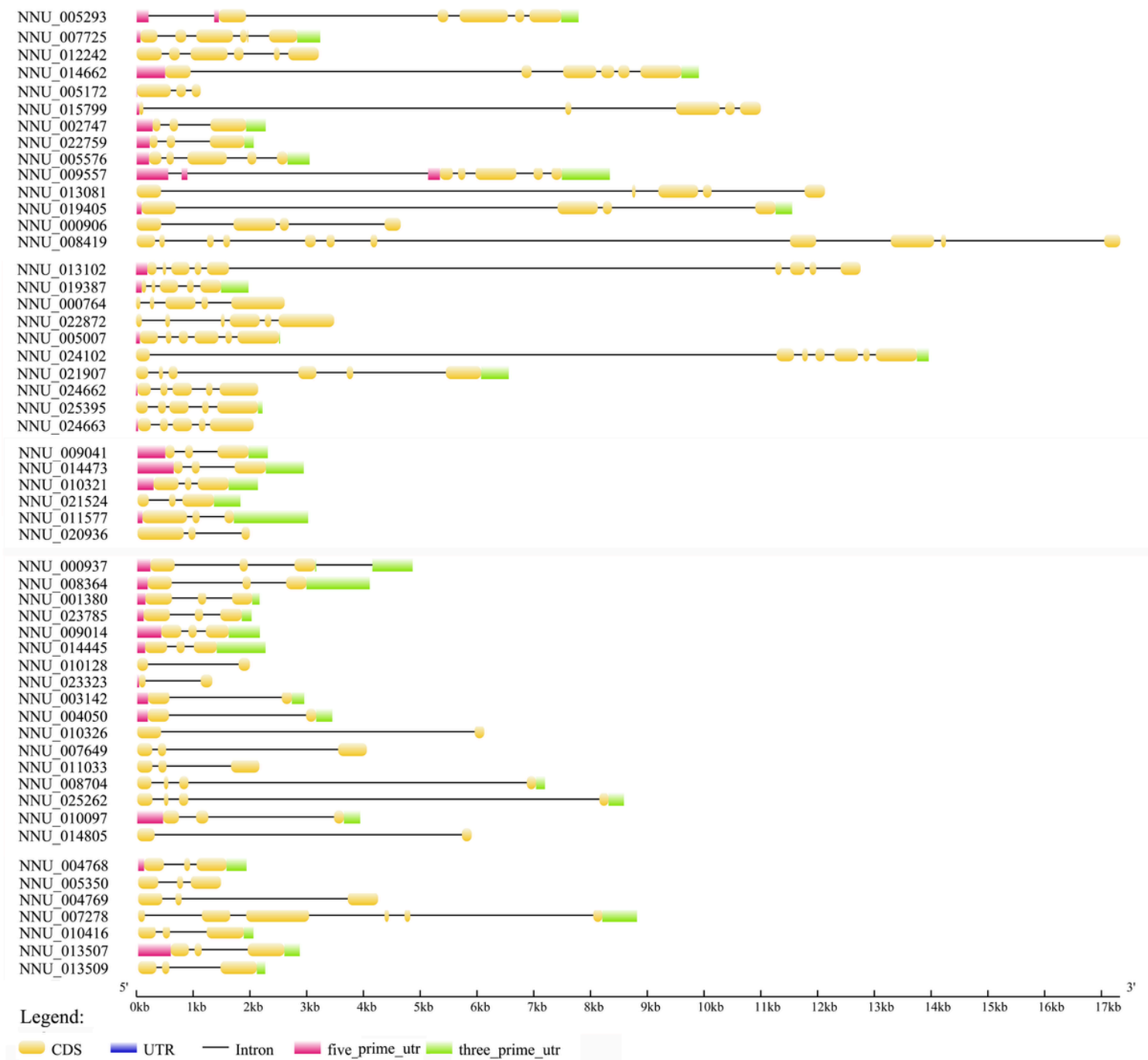


Figure 2

Structure analysis of *NnWRKY* genes. The graphic representation of the optimized gene model displayed using GSDS.

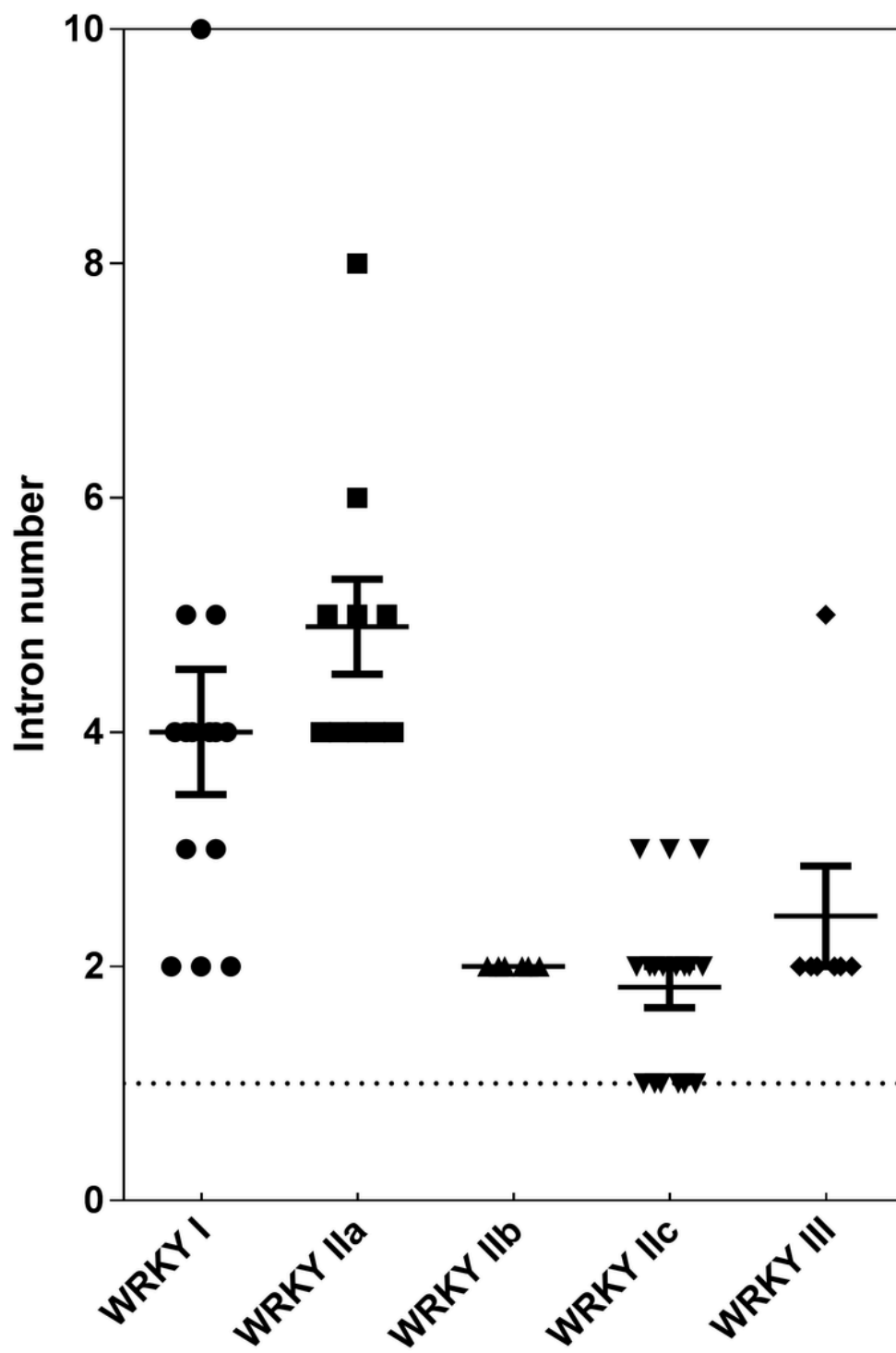


Figure 3

Intron numbers in members of different groups. The dot represent the intron number of each gene within groups, the error bars represent means $\pm$ SE. (□□□□)

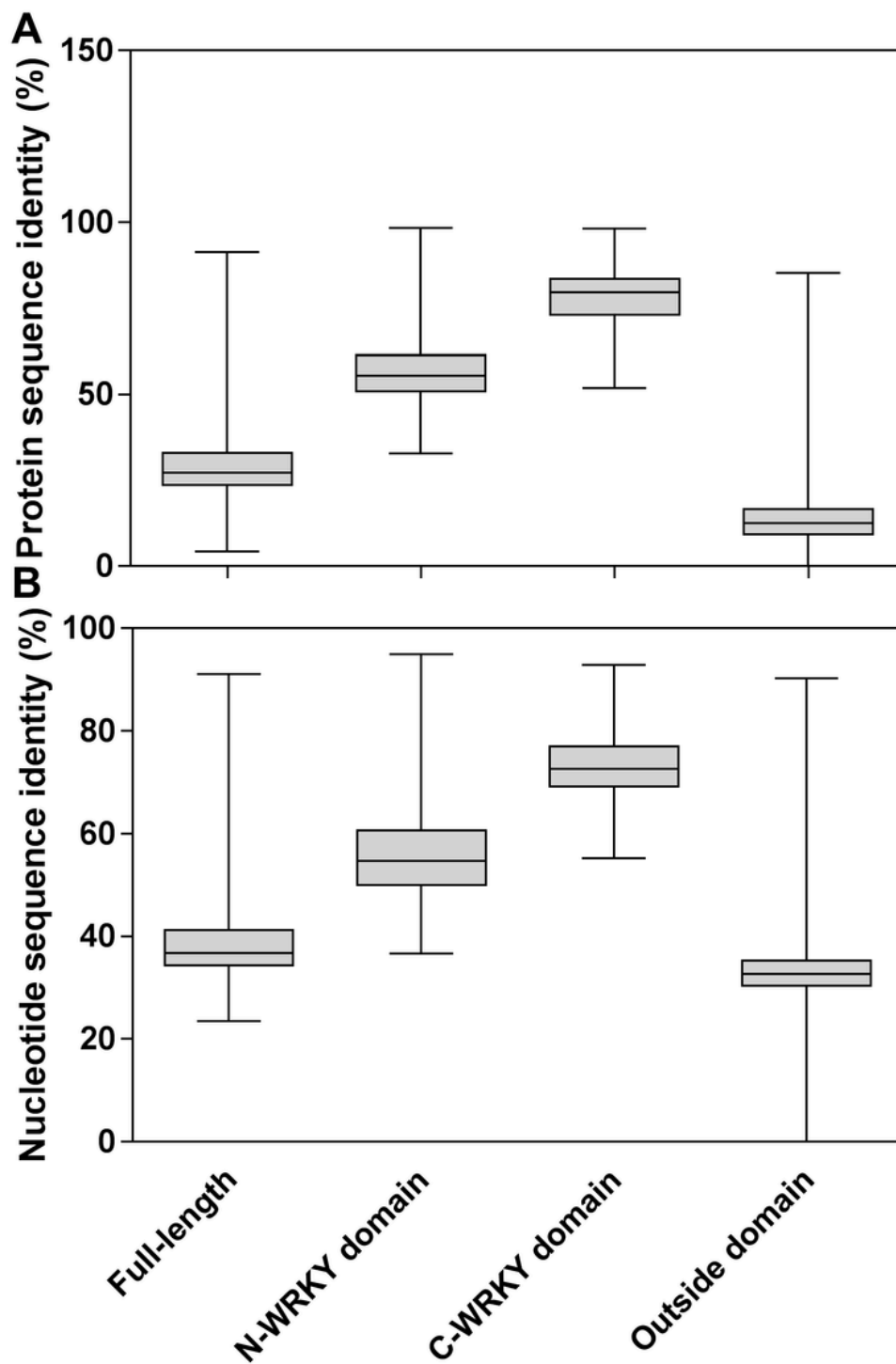


Figure 4

Pairwise sequence identities for different regions of lotus *WRKY* genes. Pairwise sequence identities between N-WRKY and C-WRKY domain, full length sequence and sequence outside domain were calculated. The box extends from the 25th to 75th percentiles which giving the relative standing of the data. The line in the middle of the box is plotted at the median, and the whiskers go down to the smallest value and up to the largest.

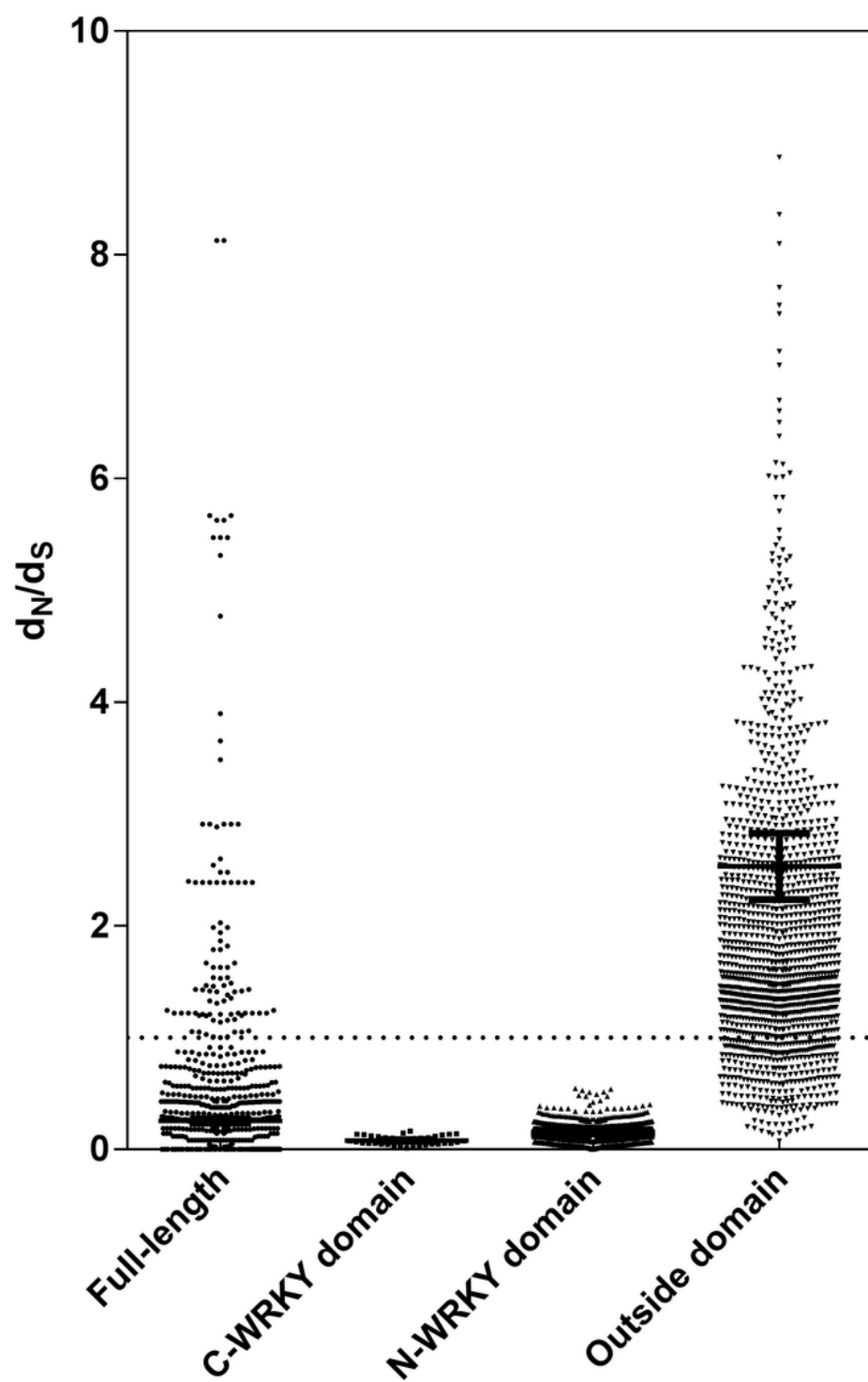


Figure 5

d_N/d_S calculation of different regions of NnWKRYgenes. The dots represent the d_N/d_S value of gene pairs, and the error bars represent means $\pm$ SE.

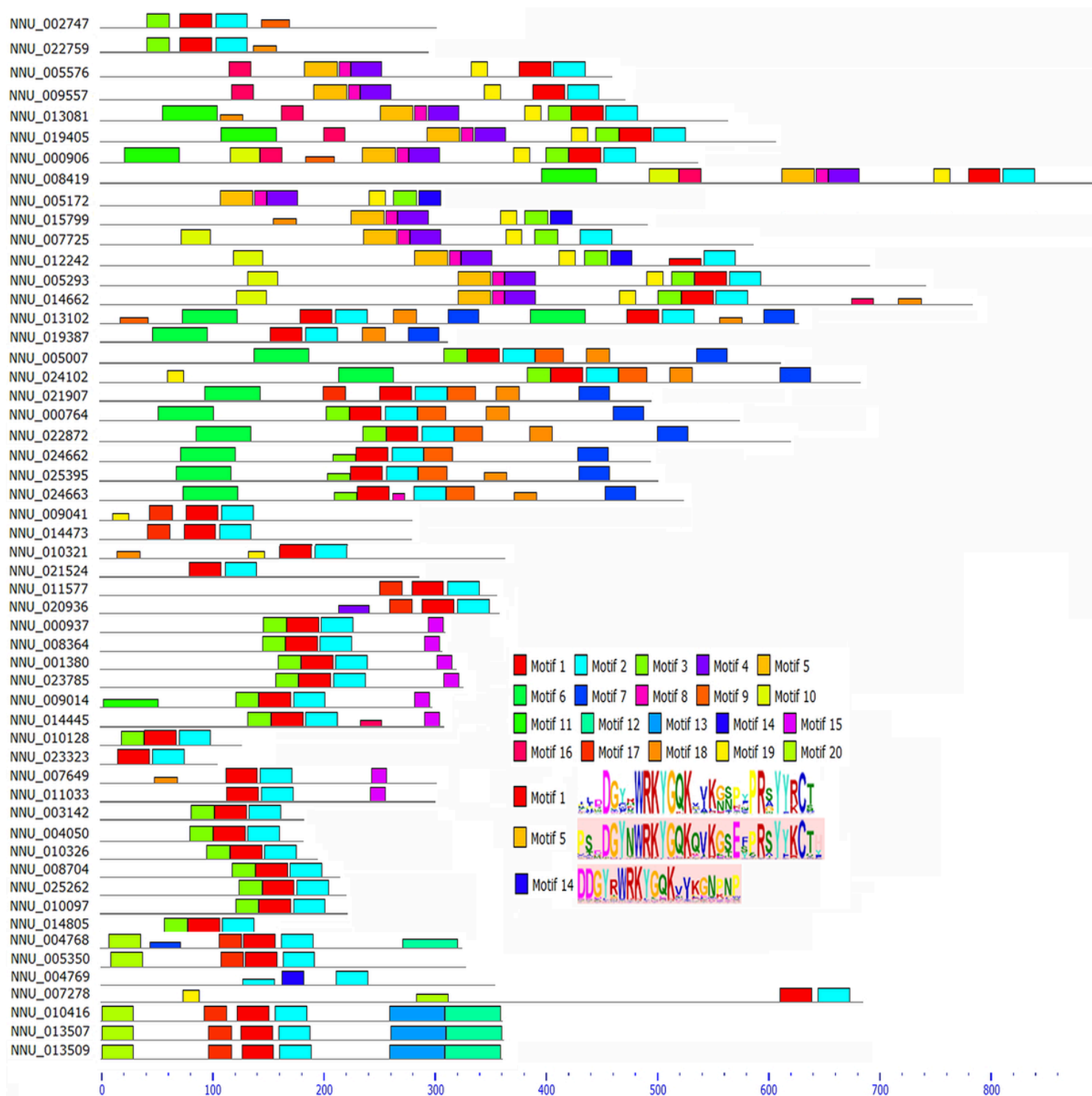


Figure 6

Schematic diagram of amino acid motifs of NnWRKY protein. The different-colored boxes represent different motifs and their position in each WRKY sequence.

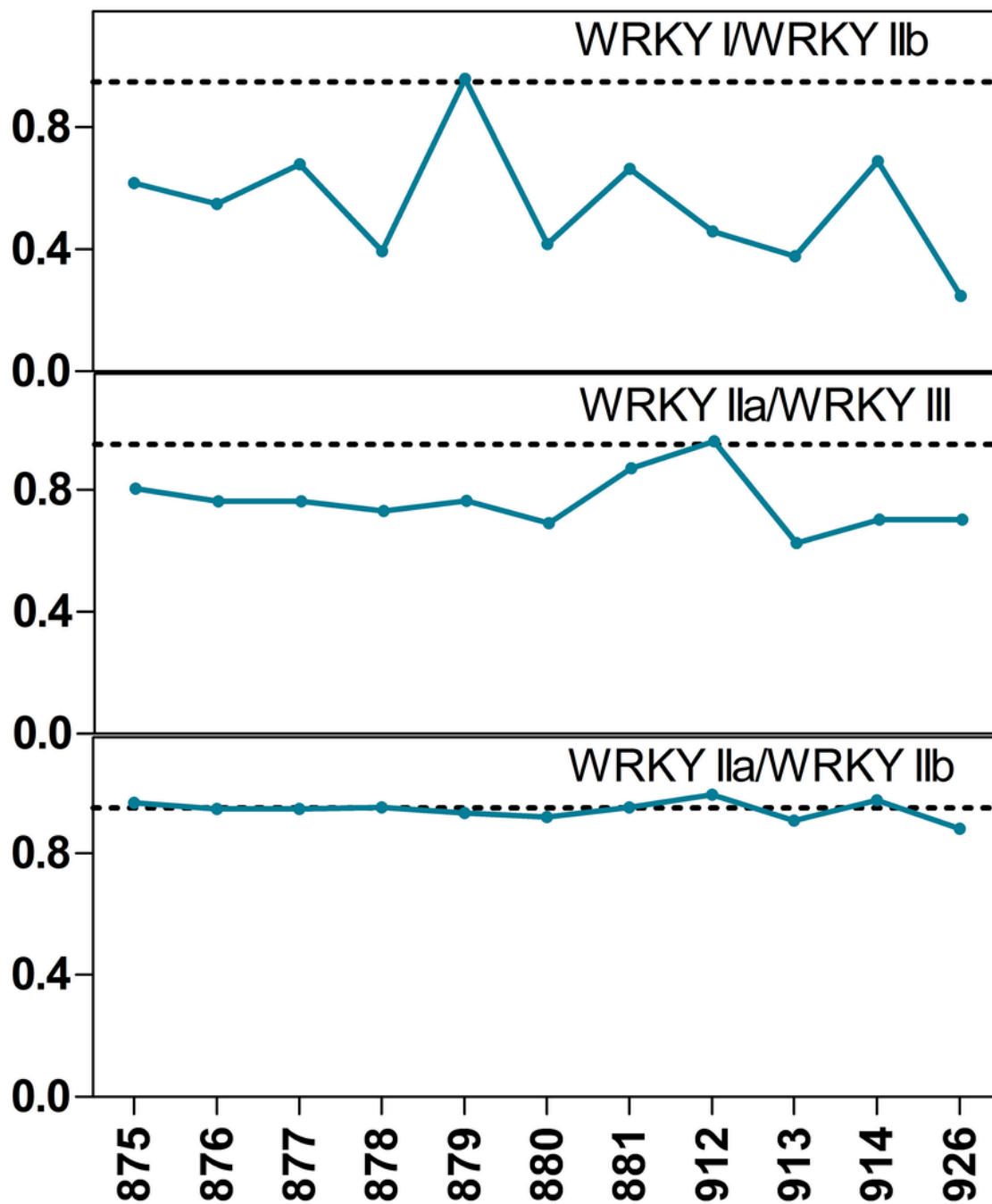


Figure 7

Type I functional divergence among the *NnWRKY* gene members. Posterior probability (PP) profiles of the site-specific type I functional divergence. The line indicates cutoff=0.95.

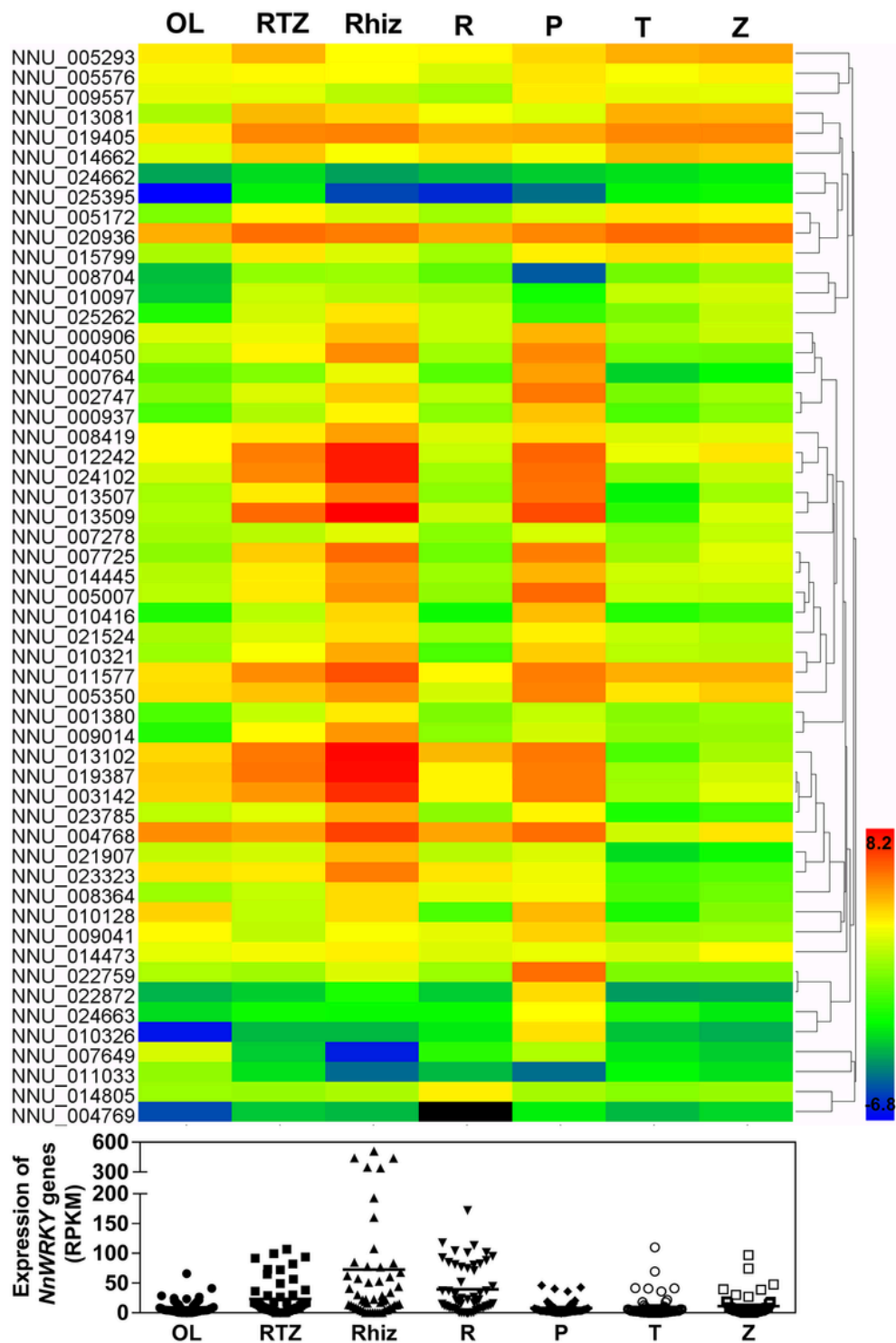


Figure 8

Expression profiles of *NnWRKY* genes in different tissues. Tissues were collected from mature lotus (*Nelumbo nucifera* Gaertn,'Chinese Antique') and used for library preparation and sequencing. Heatmap showing hierarchical clustering of *NnWRKY* genes using FPKM normalized $\log_2$ transformed data. The color bar shows the range of expression values from highest expression level (red) to lowest expression level (blue). The bottom scatter plot shows the FPKM value of each genes and average FPKM. OL, leaf;

RTZ, Combined Rhiz Tip Zone; Rhiz, Rhizome internode; R, Root; P, Petiole; T, Rhizome apical tip; Z, Rhizome elongation zone.

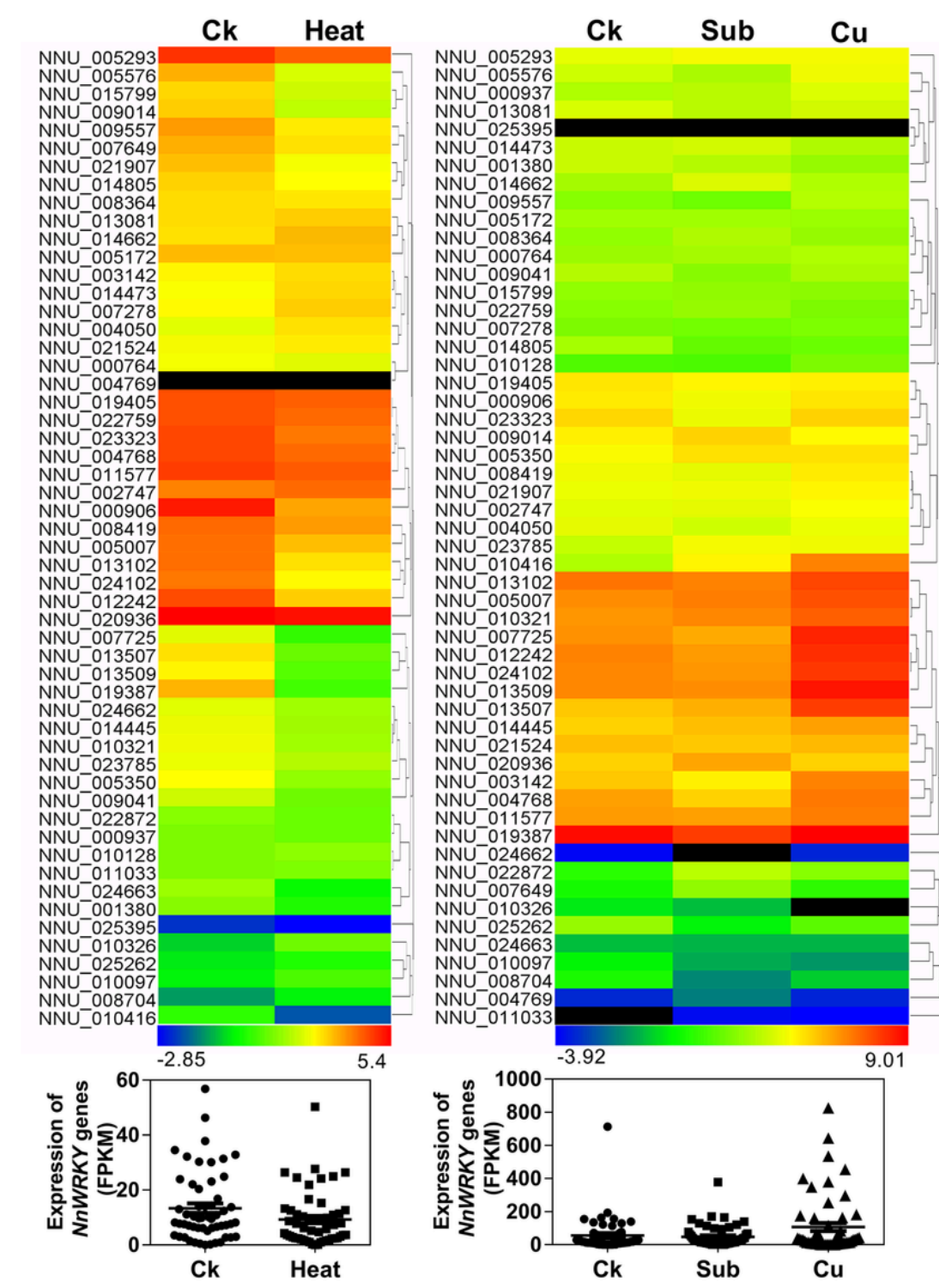


Figure 9

Expression profiles of NnWKRY genes upon heat, submergence and copper stresses. For heat treatment, one-month-old seedlings (*Nelumbo nucifera* Gaertn, 'Chinese Antique') were transfer to 37 °C for 1 h. Three-month-old lotus seedlings (*Nelumbo nucifera* Gaertn,'Honghe') were used for submergence

treatment (complete submergence for 24 h) and copper treatment (5 mM copper treatment for 24 h). Heatmap showing hierarchical clustering of *NnWRKY* genes using FPKM normalized $\log_2$ transformed data. The color bar shows the range of expression values from highest expression level (red) to lowest expression level (blue). The bottom scatter plot figure shows the FPKM value of each genes and average FPKM. Sub, submergence treatment. Cu, copper treatment.

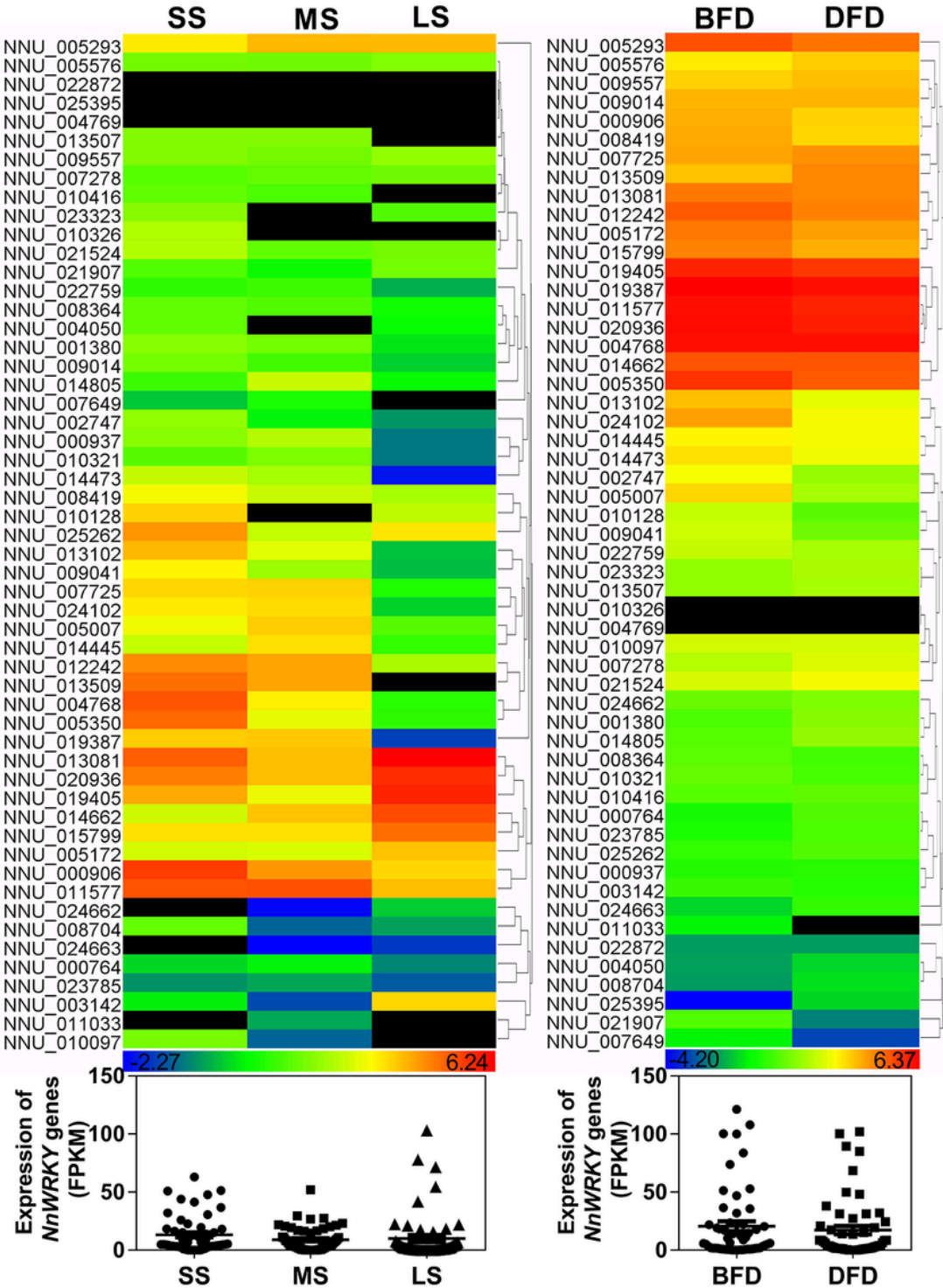


Figure 10

Expression profiles of *NnWRKY* genes in different development stage. Rhizome of Lotus (*Nelumbo nucifera* Gaertn, 'RL') were sampled at three rhizome developmental stages including SS (stolon), MS (middle swelling stage), LS (later swelling stage). Three-month-old lotus (*Nelumbo nucifera* Gaertn) were sample for its bud at two different developmental stages including BFD (before flower bud development) and DFD (during flower bud development). Heatmap showing hierarchical clustering of *NnWRKY* genes using FPKM normalized $\log_2$ transformed data. The color bar shows the range of expression values from highest expression level (red) to lowest expression level (blue). The bottom scatter plot figure shows the FPKM value of each genes and average FPKM.

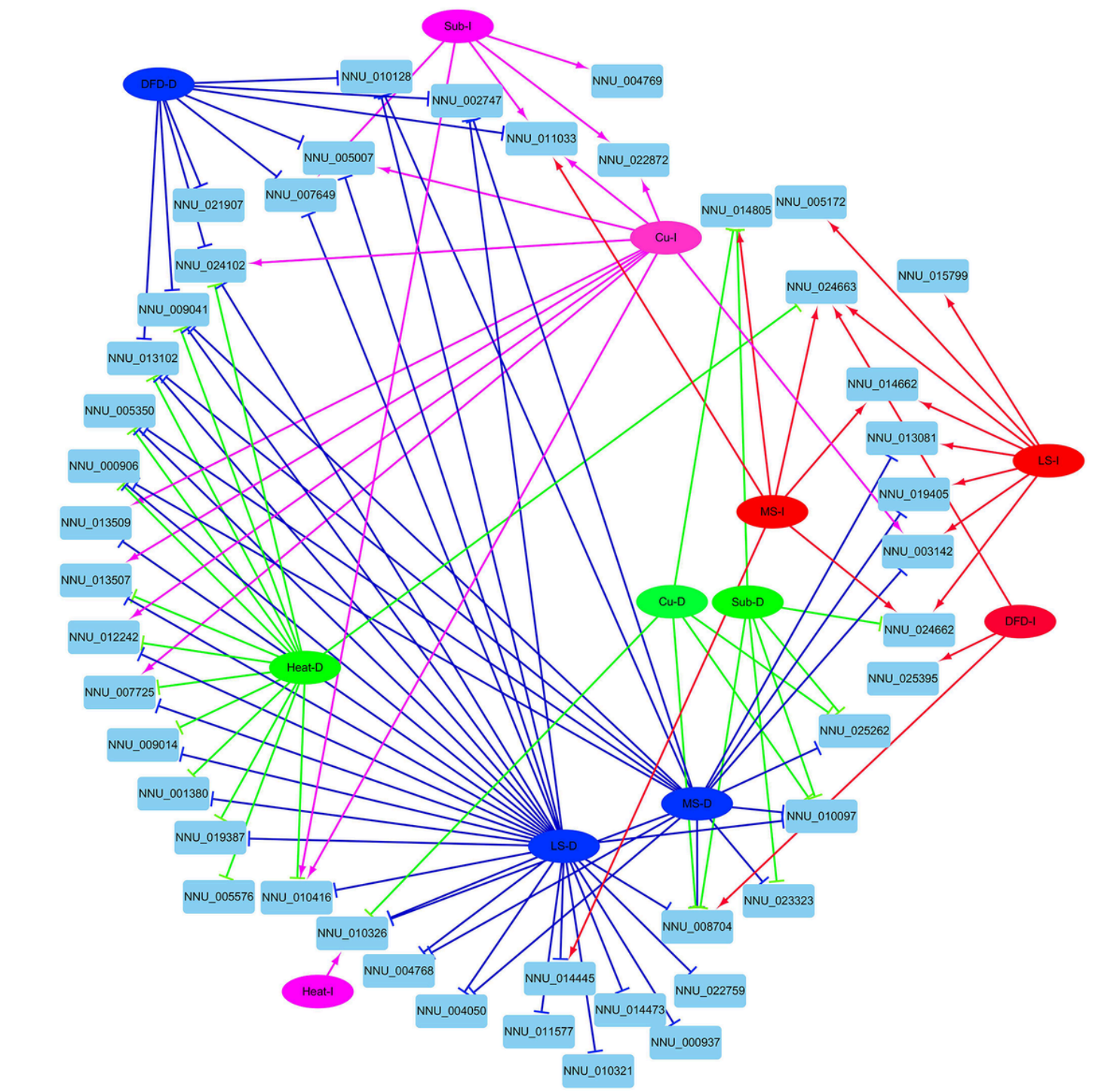


Figure 11

Co-regulated network of NnWKRY to various stimuli. –I, induce gene expression. –D, inhibit gene expression. SS, stolon. MS, middle swelling stage. LS, later swelling stage. BFD, before flower bud development. DFD, during flower bud development. Sub, submergence treatment. Cu, copper treatment. OL, leaf; RTZ, Combined Rhiz Tip Zone; Rhiz, Rhizome internode; R, Root; P, Petiole; T, Rhizome apical tip; Z, Rhizome elongation zone.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementary.docx](#)