Molecular Plant Breeding 2023, Vol.14, No.11, 1-14
http://genbreedpublisher.com/index.php/mpb

Research Article Open Access

Identification and Expression Profiles of the WRKY Gene Family in Pecan (*Carya illinoinensis*)

Min Wang [1,2], Jinxiu Chen [1], Xiang Tai [1], Xiaowei Zhu [1], Cancan Zhu [2] ✉, Tianyue Bo [1] ✉

1 Horticulture Research Institute, Shanghai Academy of Agricultural Sciences, Shanghai, 201403, China

2 Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, 210014, China

✉ Corresponding author email: tybo@saas.sh.cn; zhucancan858@163.com

Molecular Plant Breeding, 2023, Vol.14, No.11 doi: 10.5376/mpb.2023.14.0011

Received: 21 Apr., 2023

Accepted: 25 May, 2023

Published: 05 Jun., 2023

Copyright © 2023 Wang et al., This is an open access article published under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preferred citation for this article:

Wang M., Chen J.X., Tai X., Zhu X.W., Zhu C.C., and Bo T.Y., 2023, Identification and expression profiles of the WRKY gene family in pecan (*Carya illinoinensis*), Molecular Plant Breeding, 14(11): 1-14 (doi: 10.5376/mpb.2023.14.0011)

Abstract WRKY gene family encodes a large transcription factor family that play critical roles in various physiological processes. However, a systematic analysis of the WRKY transcription factor family has not been reported in pecan (*Carya illinoinensis*). In this study, a total of 89 putative pecan *WRKY* genes (*CiWRKYs*), named *CiWRKY1-89*, were identified from the whole genome of pecan. Most of WRKY domain sequences in CiWRKY proteins were "WRKYGQK", but there were "WRKYGKK" in CiWRKY19, CiWRKY27 and CiWRKY88, and "WKKYGQK" in CiWRKY85. All *CiWRKYs* were unevenly distributed on 16 chromosomes, the largest number of genes on chromosome 3. Based on phylogenetic analysis, the 89 putative *CiWRKYs* could be classified into three major groups. The *CiWRKY* genes shared similar exon-intron distribution, and conserved motifs within the same subgroups. Expression profiles indicated that *CiWRKY16*, *CiWRKY30*, *CiWRKY42*, *CiWRKY62*, and *CiWRKY69* genes were involved in flower differentiation and development, and majority of *CiWRKYs* genes were differentially expressed during embryo development. The present study provides reference for further comparative genomics and functional studies of this important class of transcriptional regulators in pecan.

Keywords Pecan (*Carya illinoensis*); WRKY transcription factor; Phylogenetics; Gene expression

WRKY transcription factors constitute one of the largest and important gene families in higher plants (Ulker et al., 2004). WRKY proteins contain a highly conserved WRKY domain, which consists of about 60 amino acids at the N-terminus, either C_2H_2 or C_2HC zinc-finger motifs at the C-terminus (Rushton et al., 2010). The WRKY family could be divided into three major groups (Group I, II, and III) according to the number of WRKY domains and the type of zinc finger structures (Eulgem et al., 2000). Group I contains two WRKY domains, and the zinc finger structure type is C_2H_2. *WRKY* genes in Group II have only one WRKY domain, with a C_2H_2 zinc-finger motif. This group can also be further divided into five subgroups (II-a, b, c, d, and e). Group III contains one WRKY domain, characterized by a C_2HC zinc-finger motif.

Some *WRKY* genes have been proved to participate in the regulation of flowering time (Wang et al., 2023). *AtWRKY12* positively regulates flowering. However, *AtWRKY13* presents a negative function in flowering. Further results showed that both *AtWRKY12* and *AtWRKY13* can directly bind to the *FUL* gene's promoter and produce distinct regulatory effects on the downstream target genes (Li et al., 2016). Overexpression of *Mangifera indica MlWRKY12* in *Arabidopsis* exhibited early flowering (Yu et al., 2013). Similarly, *Chimonanthus praecox CpWRKY71* overexpression in *Arabisopsis* exhibited early flowering and leaf senescence phenotype (Huang et al., 2019). Strawberry *FvWRKY71* accelerated flowering in transgenic strawberry plants by directly regulating the expression of *FvFUL*, *FvSEP1*, *FvAGL42*, *FvLFY*, and *FvFPF1* (Lei et al., 2020). A number of studies have reported that several *WRKY* genes participate in the process of seed germination, postgermination growth, and dormancy (Chen et al., 2017). Five *LrWRKYs* were significantly expressed during the whole fruit development stages (Tiika et al., 2020). 50% of the *MaWRKYs* were highly expressed in fruit ripening (Goel et al., 2016). *FvWRKY48* can bind to the promoter of *FvPLA* and increase its expression, accelerating fruit softening (Zhang et al., 2022). The homozygous *mini3-1* (also known as *AtWRKY10*) mutants produced significantly smaller seed weight and size (Luo et al., 2005). The *OsWRKY78*-RNAi plants showed a semi-dwarf and small kernel

phenotype, indicating that *OsWRKY78* may play a major role in stem elongation regulation and seed development (Zhang et al., 2011).

With the growing number of fully sequenced plant genomes, WRKY TFs have been surveyed in many plant species, such as *Arabidopsis thaliana* (Rushton et al., 2010), *Oryza sativa* (Ramamoorthy et al., 2008), *Vitis vinifera* (Wang et al., 2014), *Populus trichocarpa* (He et al., 2012), and *Glycine max* (Yang et al., 2017). However, the WRKY gene family has not been systematically studied after the completion of pecan genome sequencing. According to the published pecan genome data, a total of 89 *CiWRKY* genes were identified. The evolutionary relationship, chromosome location, gene structure, and conserved motif were analyzed comprehensively. At the same time, the expression characteristics of *CiWRKYs* during different stages of flower and embryo development were studied. These results provided a theoretical basis for the study of the molecular mechanism of pecan *CiWRKY* genes.

1 Results

1.1 Identification of WRKY gene family in the pecan genome

In order to identify *CiWRKYs* comprehensively, the HMM profile of WRKY domain (PF03106) and the Arabidopsis WRKY protein sequences were as queries to search for putative *CiWRKY* genes. Finally, a total of 89 *CiWRKY* genes were obtained from the pecan genome and renamed from *CiWRKY1~CiWRKY89* based on their chromosome positions. The detailed information of each *CiWRKY* gene was listed (Table 1), including gene ID, group, gene length, Molecular weight (MW), isoelectric point (pI), and subcellular localization. The deduced length of CiWRKY proteins ranged from 152 aa (CiWRKY72) to 747 aa (CiWRKY77). The predicted molecular weight ranged from 17.655 kD (CiWRKY72) to 80.303 kD (CiWRKY77) and pI value varied from 4.89 (CiWRKY69) to 9.94 (CiWRKY47). The majority of CiWRKY proteins (95.5%) were predicted to be located in the nucleus. Whereas CiWRKY5 and CiWRKY85 were located in chloroplasts, CiWRKY12 and CiWRKY51 were located in peroxisome.

1.2 Phylogenetic analysis and chromosomal distribution of *CiWRKYs*

The conserved domain of WRKY proteins in pecan were evaluated. Out of 89 CiWRKY members, 85 were properly conservative in the 'WRKYGQK' domain (Figure 1). Based on the phylogenetic tree of WRKY proteins from pecan and Arabidopsis, all the 89 CiWRKY proteins could be divided into three major groups (Figure 2). There were 16 CiWRKY proteins in Group I, each of which contained two WRKY domains and the C_2H_2-type zinc-finger motifs. 60 CiWRKY proteins assigned to Group II, which harbored one WRKY domain and C_2H_2-type zinc-finger motifs. The members of Group II were further classified into five subgroups and comprised of Group II-a, -b, -c, -d, and -e with 6, 10, 27, 15, and 16 members, respectively. Finally, 10 CiWRKY proteins, each with a single WRKY domain and C_2HC zinc-finger structure, were assigned to Group III. CiWRKY19, CiWRKY27, and CiWRKY88 exhibited sequence divergence in the WRKY domain. Therefore, three CiWRKY proteins (CiWRKY62, CiWRKY69, and CiWRKY85) were not classified into any group. Totally 89 candidate *CiWRKY*s were unevenly distributed on sixteen pecan chromosomes (Figure 3). Chromosome 1 had the largest number (9, 10.11%) of *BoWRKYs*, chromosome 14 and 16 had the least number of *CiWRKYs*, only *CiWRKY86* and *CiWRKY89* respectively. Chromosome 7 contained seven *CiWRKYs*, which all belonged to Group II.

1.3 Motif analysis and exon-intron organization of *CiWRKY* genes

Fifteen conserved motifs in full length CiWRKY proteins were identified by using the MEME online tool (http://meme.sdsc.edu/meme/intro.html) (Figure 4). It can be observed that the motif 1 and 2, which are the WRKY domains, widely distributed in 89 members. Some motifs are shared by specific group such as motif 9 present in Group IIb. Group I contained the largest number of motifs, and motif 5, 13, 15, and 61 only existed in Group I. As expected, members in the same family shared similar motif compositions, suggesting functional similarities. The exon-intron structure of all *CiWRKY* genes was analyzed to gain more insight into the evolution of the WRKY family in pecan (Figure 4). As a result, 39 *CiWRKY* genes (39/89) contained two introns, 22 *CiWRKY* genes were found to possess four introns, 12 *CiWRKYs* had three introns and ten *CiWRKYs* had only one intron. *CiWRKY76* contained the largest number of introns. All the Group III *CiWRKYs* contained two introns. Members in the same subgroups shared similar gene structures.

Table 1 Identification, classification and physicochemical properties of *CiWRKY* genes

Gene name	Gene ID	Group	Number of amino acid	Molecular weight	Theoretical pI	Instability index	Aliphatic index	Grand average of hydropathicity	Subcellular localization
CiWRKY1	CiPaw.01G014300	IId	340	37190.96	9.50	57.00	64.56	-0.628	nucleus
CiWRKY2	CiPaw.01G066400	IIe	335	36720.32	5.27	57.21	50.72	-0.787	nucleus
CiWRKY3	CiPaw.01G067800	IIc	205	23309.64	9.30	49.00	67.51	-0.679	nucleus
CiWRKY4	CiPaw.01G073900	III	352	39323.53	5.15	53.91	60.65	-0.691	nucleus
CiWRKY5	CiPaw.01G127700	IIb	405	44206.63	7.70	57.79	65.41	-0.551	chloroplast
CiWRKY6	CiPaw.01G136600	IIc	334	37583.59	5.88	66.90	46.44	-0.987	nucleus
CiWRKY7	CiPaw.01G254500	IIa	312	34585.67	7.61	47.08	63.75	-0.758	nucleus
CiWRKY8	CiPaw.01G306900	III	379	42069.11	5.97	61.05	69.21	-0.634	nucleus
CiWRKY9	CiPaw.01G307000	III	321	36128.22	5.40	60.87	65.36	-0.735	nucleus
CiWRKY10	CiPaw.02G007000	IId	334	36957.96	9.38	42.40	66.26	-0.590	nucleus
CiWRKY11	CiPaw.02G034700	IIe	329	35772.31	5.05	62.39	54.59	-0.719	nucleus
CiWRKY12	CiPaw.02G035600	IIc	213	24307.73	9.23	49.56	64.04	-0.700	peroxisome
CiWRKY13	CiPaw.02G073500	IIc	326	36519.61	7.66	64.56	49.94	-0.918	nucleus
CiWRKY14	CiPaw.02G162100	IIa	316	35093.18	8.81	44.35	59.59	-0.870	nucleus
CiWRKY15	CiPaw.02G199000	III	363	40112.92	6.33	59.67	58.07	-0.701	nucleus
CiWRKY16	CiPaw.03G021200	I	517	56487.63	6.96	59.56	56.83	-0.771	nucleus
CiWRKY17	CiPaw.03G057700	IIc	186	21592.31	9.55	40.21	51.29	-0.965	nucleus
CiWRKY18	CiPaw.03G132400	IIc	210	23726.84	9.15	41.99	67.33	-0.605	nucleus
CiWRKY19	CiPaw.03G133500	IIc	190	21170.42	5.72	39.39	57.47	-0.750	nucleus
CiWRKY20	CiPaw.03G158300	IIb	632	68461.67	5.79	43.21	59.83	-0.703	nucleus
CiWRKY21	CiPaw.03G191000	IIc	305	34106.80	6.01	57.33	57.87	-0.836	nucleus
CiWRKY22	CiPaw.03G258800	IId	345	38228.54	9.49	55.15	71.77	-0.631	nucleus
CiWRKY23	CiPaw.04G014200	I	411	45185.20	8.17	59.63	57.69	-0.905	nucleus
CiWRKY24	CiPaw.04G019000	IIb	584	64364.93	7.57	49.74	60.67	-0.682	nucleus
CiWRKY25	CiPaw.04G039100	IIc	192	21792.63	9.42	45.92	54.79	-0.808	nucleus
CiWRKY26	CiPaw.04G089600	IIc	240	27202.55	9.22	52.87	55.62	-0.814	nucleus
CiWRKY27	CiPaw.04G090900	IIc	168	18918.78	7.12	59.53	49.29	-1.023	nucleus
CiWRKY28	CiPaw.04G114200	IIb	628	68098.75	6.48	47.06	60.67	-0.661	nucleus
CiWRKY29	CiPaw.04G182800	IId	344	38678.81	9.52	53.54	63.49	-0.801	nucleus
CiWRKY30	CiPaw.04G195800	I	464	51206.31	5.87	50.55	56.85	-0.924	nucleus
CiWRKY31	CiPaw.05G042300	IIb	528	58129.76	6.06	42.26	69.47	-0.688	nucleus
CiWRKY32	CiPaw.05G107300	IIc	181	20599.07	9.47	44.70	53.31	-0.839	nucleus
CiWRKY33	CiPaw.05G126900	IIc	230	25523.74	6.18	57.82	40.22	-1.108	nucleus
CiWRKY34	CiPaw.05G129000	I	539	58344.52	8.36	54.85	55.77	-0.829	nucleus

Continued Table 1

Gene name	Gene ID	Group	Number of amino acid	Molecular weight	Theoretical pI	Instability index	Aliphatic index	Grand average of hydropathicity	Subcellular localization
CiWRKY36	CiPaw.05G257700	III	376	42114.93	6.6	65.43	59.44	-0.697	nucleus
CiWRKY37	CiPaw.05G261200	IIe	407	44301.32	5.57	63.24	51.38	-0.876	nucleus
CiWRKY38	CiPaw.06G007700	III	380	42296.65	5.65	59.07	57.76	-0.710	nucleus
CiWRKY39	CiPaw.06G081500	IIb	337	37105.99	8.96	54.34	57.89	-0.718	nucleus
CiWRKY40	CiPaw.06G089200	I	527	57123.12	6.58	55.53	58.69	-0.739	nucleus
CiWRKY41	CiPaw.06G091000	IIc	331	36543.33	6.45	57.17	58.28	-0.751	nucleus
CiWRKY42	CiPaw.06G104900	IIc	180	20383.81	9.30	47.26	50.89	-0.869	nucleus
CiWRKY43	CiPaw.06G149300	IIb	530	58783.37	5.53	43.87	69.75	-0.721	nucleus
CiWRKY44	CiPaw.07G118300	IIe	265	30046.68	5.40	49.74	62.19	-0.869	nucleus
CiWRKY45	CiPaw.07G119500	IIc	308	34065.14	6.72	66.56	61.46	-0.639	nucleus
CiWRKY46	CiPaw.07G125900	IIb	544	59762.15	8.13	50.69	64.60	-0.619	nucleus
CiWRKY47	CiPaw.07G199600	IId	315	34426.04	9.94	52.59	63.78	-0.625	nucleus
CiWRKY48	CiPaw.07G211500	IId	344	36959.98	9.67	45.42	65.78	-0.480	nucleus
CiWRKY49	CiPaw.07G229500	IIa	306	34486.7	7.65	49.13	65.69	-0.751	nucleus
CiWRKY50	CiPaw.07G229600	IIa	264	29905.44	8.51	59.45	69.13	-0.755	nucleus
CiWRKY51	CiPaw.08G006400	IIa	273	30585.53	8.86	59.86	73.22	-0.696	peroeus
CiWRKY52	CiPaw.08G006500	IIa	310	34718.92	8.38	49.39	62.68	-0.709	nucleus
CiWRKY53	CiPaw.08G031100	IId	316	34698.34	9.82	45.88	65.44	-0.574	nucleus
CiWRKY54	CiPaw.08G087600	IIc	218	24291.21	9.40	63.60	54.54	-0.877	nucleus
CiWRKY55	CiPaw.08G089100	IIe	256	29242.87	5.81	52.42	68.52	-0.915	nucleus
CiWRKY56	CiPaw.08G115100	I	720	78097.63	5.22	44.76	68.68	-0.540	nucleus
CiWRKY57	CiPaw.09G072600	IIe	248	27480.58	5.19	57.71	54.23	-0.706	nucleus
CiWRKY58	CiPaw.09G095400	IIb	588	63732.02	6.18	46.86	65.77	-0.604	nucleus
CiWRKY59	CiPaw.09G128600	IIc	332	36331.24	6.66	63.55	44.70	-0.790	nucleus
CiWRKY60	CiPaw.09G147900	IIe	459	48923.19	5.22	47.96	52.96	-0.583	nucleus
CiWRKY61	CiPaw.09G184100	I	475	51568.10	8.85	51.39	57.85	-0.826	nucleus
CiWRKY62	CiPaw.09G194700	None	338	38197.16	5.57	58.02	61.21	-0.856	nucleus
CiWRKY63	CiPaw.09G216000	I	542	60274.04	6.30	59.25	46.24	-0.908	nucleus
CiWRKY64	CiPaw.09G222100	III	309	34750.04	6.55	63.03	71.97	-0.657	nucleus
CiWRKY65	CiPaw.10G061400	IIe	251	27895.11	5.04	51.26	54.78	-0.710	nucleus
CiWRKY66	CiPaw.10G073300	IIb	588	63290.08	5.95	45.67	65.27	-0.615	nucleus
CiWRKY67	CiPaw.10G094700	IIc	317	35918.59	6.92	62.88	45.58	-1.018	nucleus
CiWRKY68	CiPaw.10G138200	I	474	51575.63	9.43	43.46	54.63	-0.887	nucleus
CiWRKY69	CiPaw.10G146700	None	297	33436.37	4.89	60.67	66.7	-0.677	nucleus

Continued Table 1

Gene name	Gene ID	Group	Number of amino acid	Molecular weight	Theoretical pI	Instability index	Aliphatic index	Grand average of hydropathicity	Subcellular localization
CiWRKY73	CiPaw.11G039600	IIb	610	66385.11	6.31	53.89	51.28	-0.854	nucleus
CiWRKY74	CiPaw.11G123000	III	358	40233.73	5.43	66.44	66.45	-0.680	nucleus
CiWRKY75	CiPaw.11G129700	IIe	307	34391.91	5.84	61.27	65.41	-0.593	nucleus
CiWRKY76	CiPaw.11G200800	I	498	54779.69	6.58	39.83	62.79	-0.729	nucleus
CiWRKY77	CiPaw.11G211300	I	745	80032.94	5.70	58.71	59.17	-0.679	nucleus
CiWRKY78	CiPaw.12G015800	IIb	427	47970.75	8.62	44.68	63.28	-0.734	nucleus
CiWRKY79	CiPaw.12G084300	III	357	39654.96	5.89	51.31	60.36	-0.614	nucleus
CiWRKY80	CiPaw.12G089100	IIc	180	20585.3	9.64	39.56	66.61	-0.803	nucleus
CiWRKY81	CiPaw.12G089200	IIe	311	34879.22	6.12	64.62	66.14	-0.625	nucleus
CiWRKY82	CiPaw.12G135700	I	584	63140.97	6.93	46.70	60.60	-0.689	nucleus
CiWRKY83	CiPaw.13G025100	IIc	227	25660.05	6.99	43.81	43.70	-0.993	nucleus
CiWRKY84	CiPaw.13G068200	I	592	64845.83	7.33	58.02	48.63	-0.912	nucleus
CiWRKY85	CiPaw.13G178200	None	162	18571.91	8.89	39.63	69.14	-0.690	chloroplast
CiWRKY86	CiPaw.14G053400	I	586	64529.93	6.45	57.66	43.99	-1.017	nucleus
CiWRKY87	CiPaw.15G052500	IId	320	34681.44	9.48	55.60	71.03	-0.503	nucleus
CiWRKY88	CiPaw.15G087700	IIc	163	18755.44	4.92	50.06	47.12	-1.133	nucleus
CiWRKY89	CiPaw.16G111300	I	517	56828.83	5.52	68.46	57.93	-0.969	nucleus

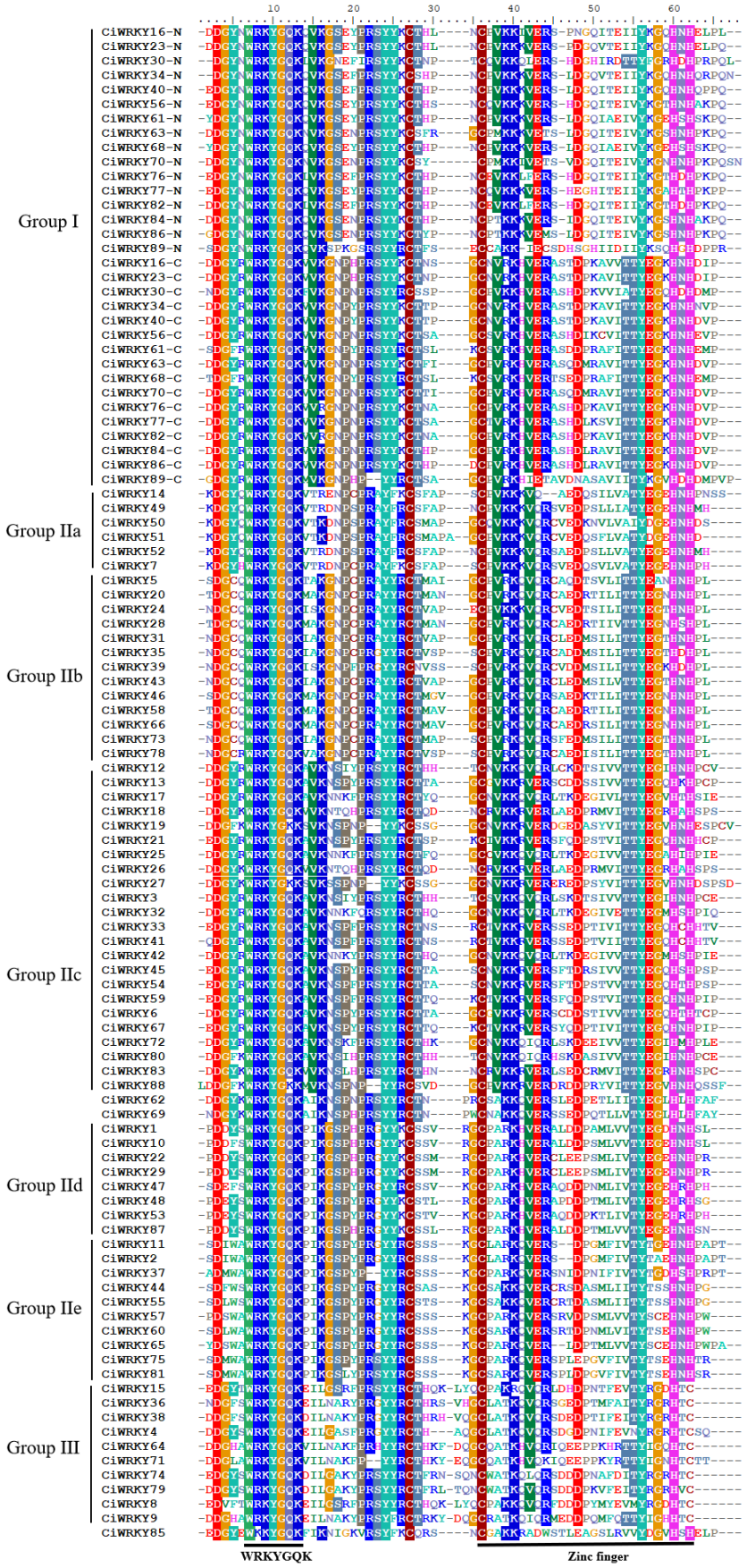


Figure 1 Multiple sequence alignment of the WRKY domains from 89 CiWRKY proteins

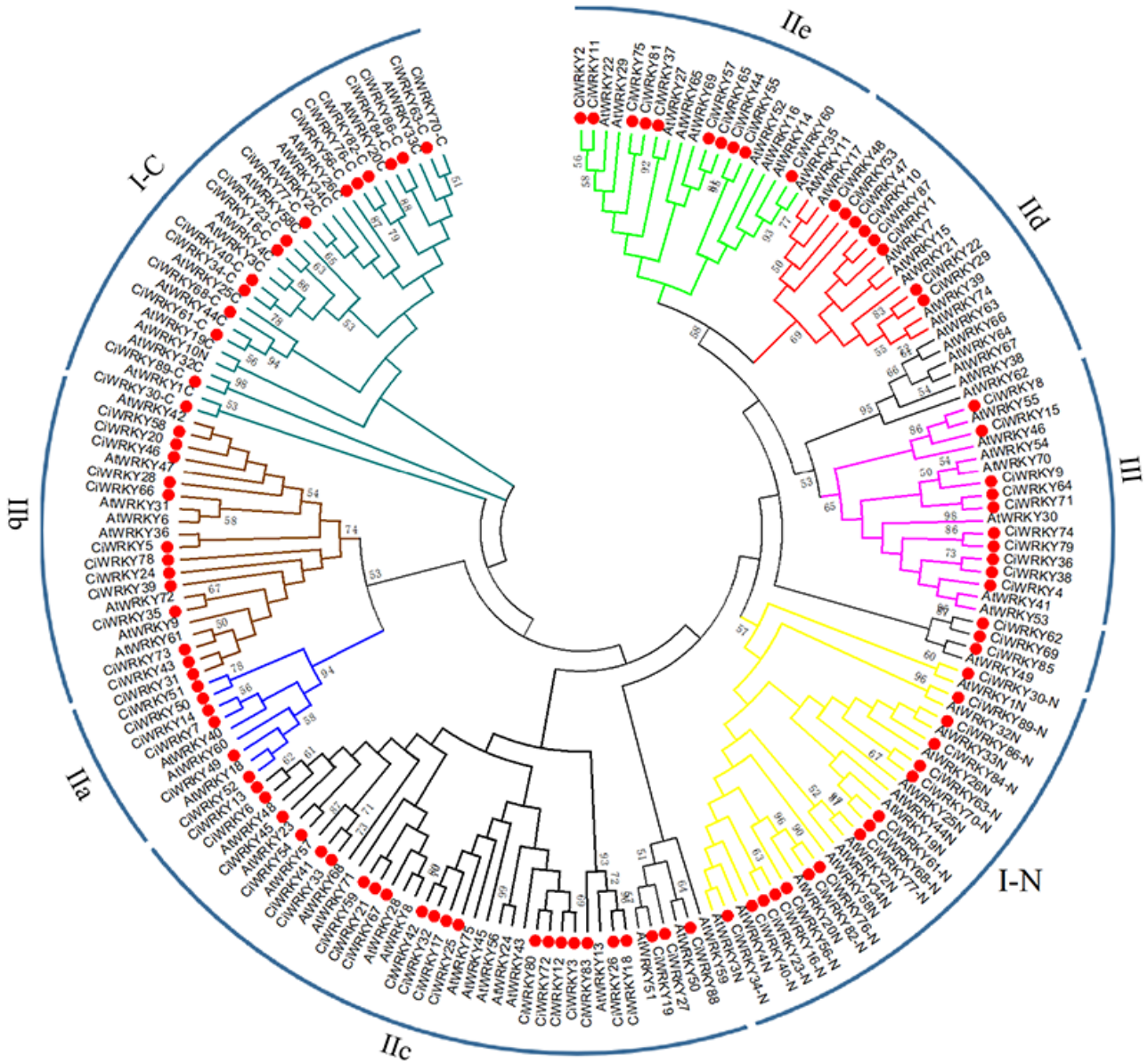


Figure 2 Phylogenetic tree of WRKY domains from pecan and Arabidopsis

Note: The name of groups (I, II, and III) and subgroup (a–e) were shown at the outside of the circle. The WRKY named with suffix-N or -C indicated the N-terminal WRKY domain or the C-terminal WRKY domain in one CiWRKY proteins with two WRKY domains

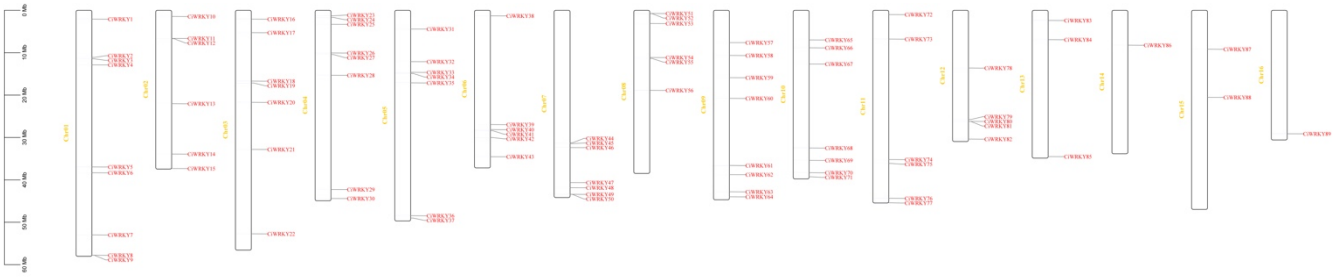


Figure 3 Chromosomal location of *CiWRKY* genes in pecan

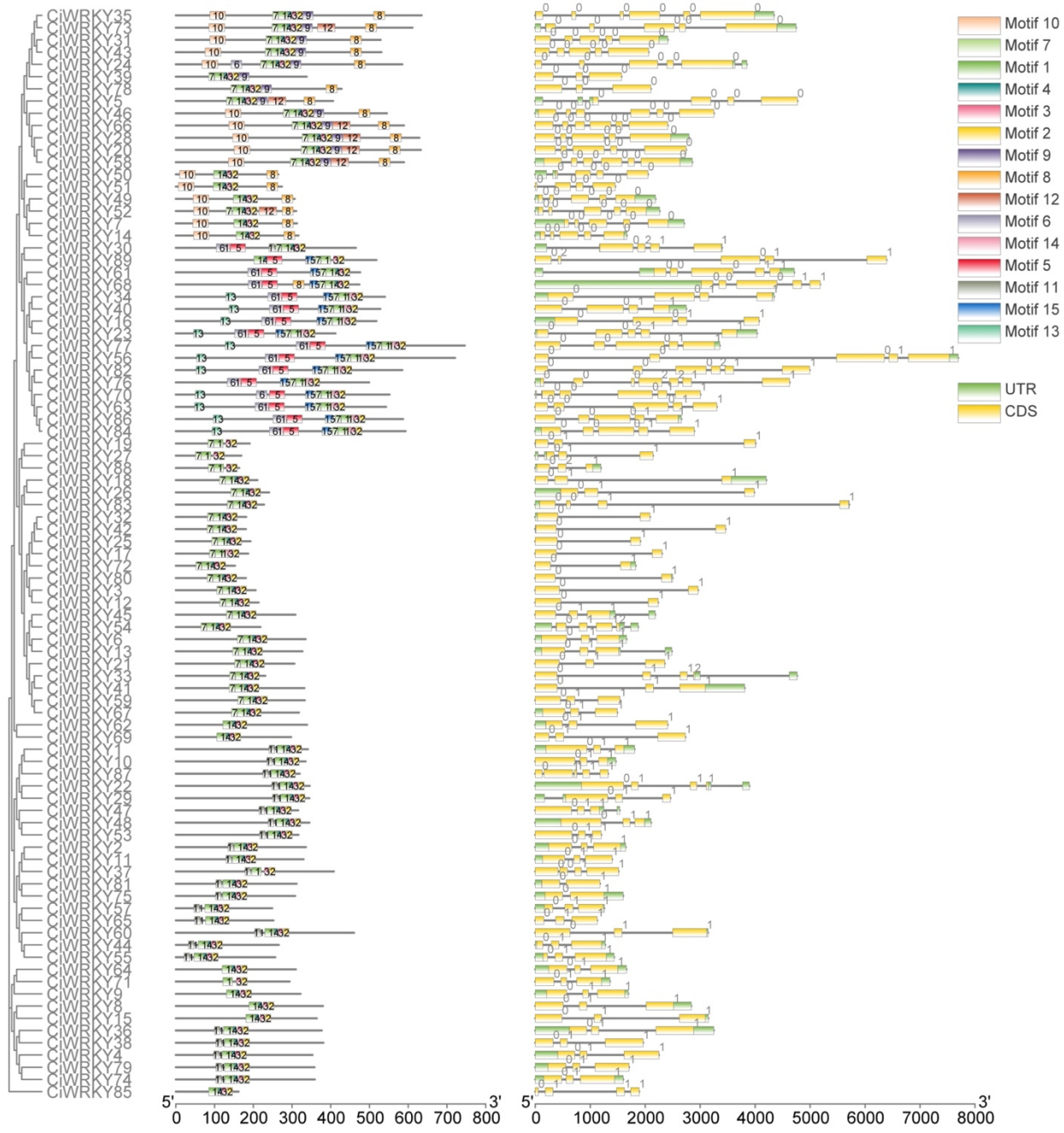


Figure 4 Conserved motifs distribution of CiWRKY proteins and exon-intron organization of *CiWRKY* genes

Note: The phylogenetic tree of full length CiWRKY proteins on the left; the conserved motifs in pecan WRKY proteins in the middle, exon-intron compositions of *CiWRKY* genes on the right

1.4 Expression profiles of *CiWRKYs* during flower and embryo development process

To further understand the function of *CiWRKYs*, the global expression patterns of *CiWRKYs* at different stages of flower development were systematically analyzed. The expression profiles of *CiWRKYs* can be divided into four types (Figure 5). 20 genes were included in type 1, which were almost not expressed during flower development. Genes within type 2 (17 genes) displayed high expression at the five stages. Especially, the expression level of *CiWRKY16*, *CiWRKY30*, *CiWRKY42*, *CiWRKY62*, and *CiWRKY69* were the highest. The other *CiWRKYs* exhibited varied expression levels. The expression of *CiWRKYs* were also investigated during the embryo development of pecan (Figure 6). 85.4% (76/89) of *CiWRKYs* were expressed during the embryo development. *CiWRKY14*, *CiWRKY58*, *CiWRKY68*, and *CiWRKY70* were only expressed during the early stage of cotyledon

development, indicating they mainly participate in the organ differentiation process. Five *CiWRKY* genes (*CiWRKY47*, *CiWRKY36*, *CiWRKY79*, *CiWRKY55*, and *CiWRKY73*) showed higher expression levels in the fully matured stage of the embryos. *CiWRKY41*, *CiWRKY9*, *CiWRKY42*, *CiWRKY80*, *CiWRKY21*, and *CiWRKY29* were highly expressed throughout the embryo development, these genes maybe closely related to the process of nutrients accumulation and embryonic tissue development.

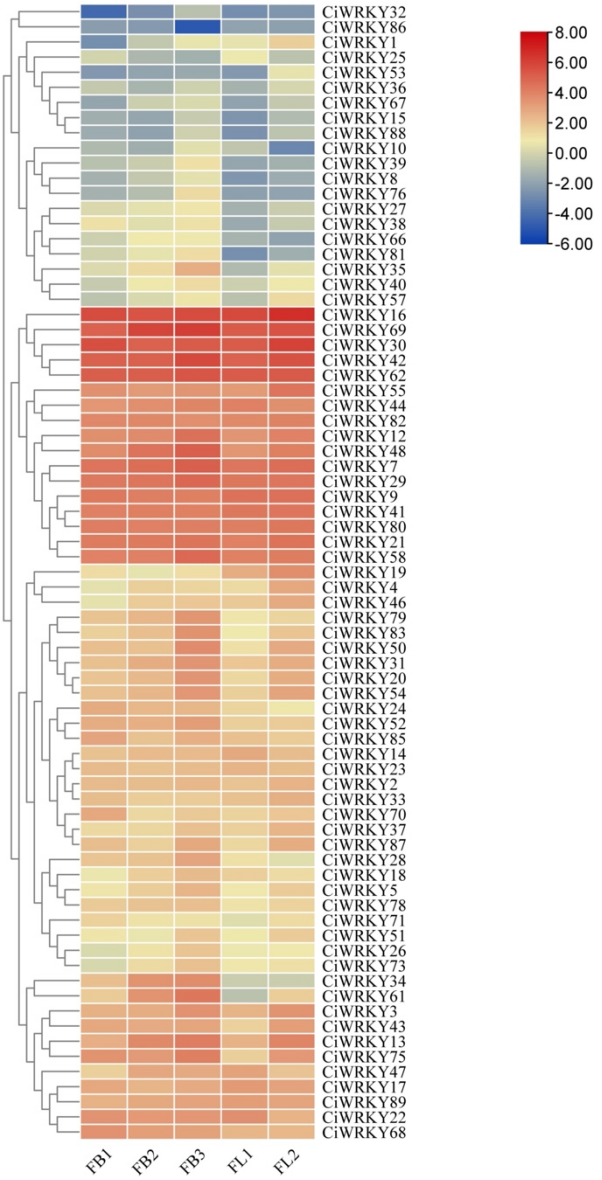


Figure 5 Expression profiles of the *CiWRKY* genes during different stages of female flower development

Note: FB1, initial stage of female flower bud differentiation; FB2, formation stage of female inflorescence; FB3, the formation stage of female flower involucre; FL1, initial flowering stage of female flowers; FL2, blooming period of female flowers

1.5 Analysis of *cis*-acting elements in the promoter regions of *CiWRKY* genes

Ten *CiWRKYs* highly expressed during flower and embryo development were selected for further *cis*-element analysis (Figure 7). Nine meristem expression elements (CAT-box) were identified in *CiWRKY41*, *CiWRKY62*, *CiWRKY69*, and *CiWRKY80* promoters. The four *CiWRKYs* were all had abscisic acid responsiveness elements (ABRE). The seed-specific regulation elements (RY-element) were found in the promoter regions of *CiWRKY62* and *CiWRKY80*, indicating these two genes were very likely to participate in the embryo development process. Additionally, MeJA-responsiveness and salicylic acid responsiveness (TCA-element) regulatory elements were located in the promoter regions of seven and five *CiWRKYs*, respectively.

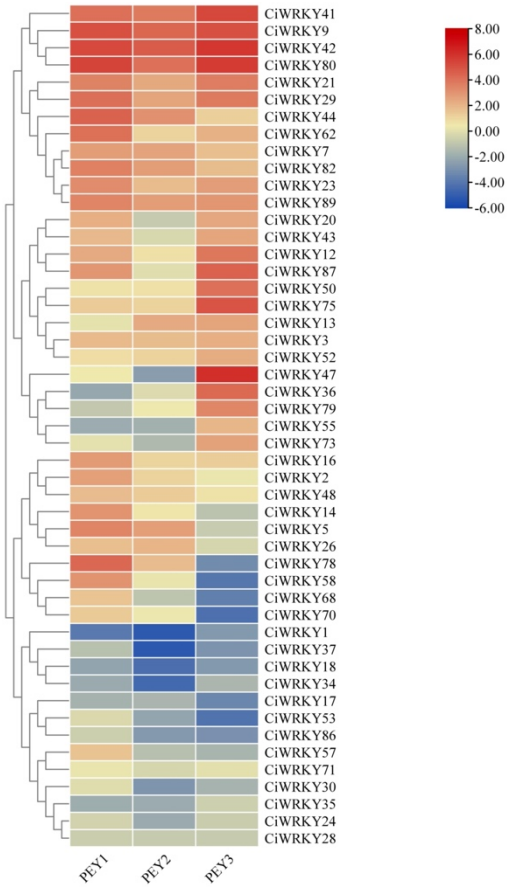
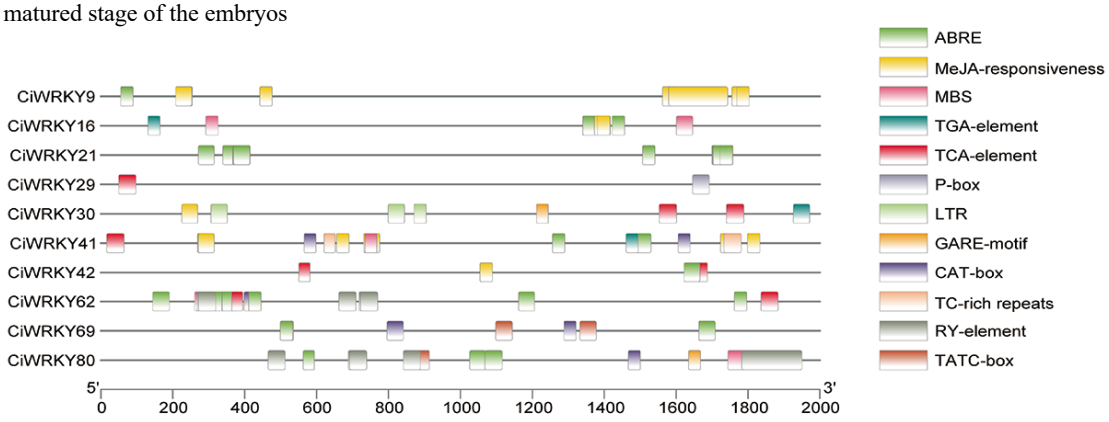


Figure 6 Expression profiles of the *CiWRKY* genes during embryo development

Note: PEY1, the early stage of cotyledon development; PEY2, the fully developed stage of cotyledon development; PEY3, the fully matured stage of the embryos



Gene name	CAT-box	RY-element	ABRE	TGA-element	CGTCA-motif	TCA-element	GARE-motif	TATC-box	P-box	MBS	LTR	TC-rich repeats
CiWRKY9			2		6							
CiWRKY16			3	1	2					2		
CiWRKY21			10		1							
CiWRKY29						1			1			
CiWRKY30				1	1	2	1				3	
CiWRKY41	2		4	1	5	1				1		2
CiWRKY42			1		1	2						
CiWRKY62	1	5	7			2				1		
CiWRKY69	2		3		1			2				
CiWRKY80	1	5	3				1	1		1		

Figure 7 *Cis*-acting elements in promoter regions of ten *CiWRKY* genes

Note: ABRE: Abscisic acid responsiveness; MBS: Drought-inducibility; TGA-element: Auxin-responsive; TCA-element: Salicylic acid responsiveness; GARE-motif, P-box, TATC-box: Gibberellin-responsive; LTR: Low-temperature responsiveness; CAT-box: Meristem expression; TC-rich repeats: Defense and stress responsiveness; RY-element: Seed-specific regulation

2 Discussion

2.1 *CiWRKY* genes in pecan

WRKY TFs are one of the largest gene families in higher plants, which play critical roles in multiple developmental processes. The characterization analysis of the WRKY gene family in many plant species have been carried out. For the first time, 89 *CiWRKYs* were identified in pecan from the latest version of genome assembly of the pecan cultivar 'Paween'. Compared with pecan (89 CiWRKYs; genome size 674 Mb), the number of WRKYs was more in rice (103; genome size 389 Mb), poplar (103; genome size 483 Mb) and fewer in in tomato (81; genome size 900 Mb) and grapevine (59; genome size 487 Mb), indicating the number of *WRKY* genes may not only related to the size of genome (Ramamoorthy et al., 2008; International Rice Genome Sequencing, 2005; Tomato Genome, 2012; Huang et al., 2012; Wang et al., 2014). The conserved domain of WRKY proteins in pecan were evaluated. Out of 89 CiWRKY members, 85 were properly conservative in the 'WRKYGQK' domain. However, three CiWRKY proteins belong to Group IIc, CiWRKY19, CiWRKY27, and CiWRKY88 (WRKYGKK) "Q" were replaced by "K". This WRKYGKK is a common variant in previous studies and usually present in Group IIc (Song et al., 2014; Song et al., 2016a; Song et al., 2016b). In a few WRKY proteins, the WRKYGQK sequence were replaced by WKKY, WRRY, WSKY, WKRY, WVKY, WRIC, WRMC, WIKY, and WKRY (Jiang et al., 2017). As shown in this study, the WKKYGQK variant appeared in CiWRKY88. The *CiWRKY* genes were categorized into three groups (I, II, and III), Group II were further classified into five distinct subgroups (IIa-e). Chen et al. (2017) proposed that IIa and IIb could be merged as a single subfamily, and the IId and IIe can also be merged into one subgroup. The phylogenetic analysis in this study showed the *CiWRKY* genes in Group IIa were closely related to IIb, and Group IIe genes were clustered with genes in IId, which support this classification.

2.2 *CiWRKY* genes function in flower and embryo development

Numerous studies have proved that *WRKY* genes regulate plant growth and development. This study focuses on the expression of *WRKY* genes during flower and embryo development. *CiWRKY21* clustered with Arabidopsis *AtWRKY71*, which positively promotes flowering via the direct modulation of *AtFT* and *AtLFY* expression (Yu et al., 2016). In this study, *CiWRKY21* was highly expressed in the whole process of female flower, suggesting that this gene is related to flower bud differentiation and flower development. Arabidopsis *AtWRKY75* is a positive factor in regulating flowering through the GA signaling pathway (Zhang et al., 2018). Moreover, *CiWRKY42* exhibited relatively higher expression throughout the whole flower development process and was closely related to Arabidopsis *AtWRKY75*, indicating *CiWRKY42* as Arabidopsis homologs maybe the key regulators of flower development. *CiWRKY42*, *CiWRKY21*, *CiWRKY80*, *CiWRKY12*, and *CiWRKY 41*, all belonging to Group IIc, were also highly expressed, we speculated that these Group IIc WRKY proteins may play a role in flower development. Embryo development is a very important stage in the research of pecan. The expression changes of *CiWRKYs* at three stages of embryo development varied greatly. *CiWRKY68* exhibited higher expression in the early stage of cotyledon development, which indicates its potential role in embryo development. *AtWRKY2*, a *ClWRKY68* homolog, which mediates seed germination and postgermination developmental arrest by ABA (Jiang et al., 2009). *CiWRKY36* clustered together with *AtWRKY41*, which positively regulates ABA signaling and seed maturation genes during early post-germination seedling growth (Ding et al., 2014). In the expression profile, *CiWRKY36* was highly expressed in the fully matured stage of the embryos, suggesting that this gene may have similar functions as *AtWRKY41*.

3 Materials and Methods

3.1 Identification and annotation of *WRKY* genes in pecan genome

The genome sequences of pecan and Arabidopsis were downloaded from Phytozome 13 (https://phytozome next.jgi.doe.gov/info/CillinoinensisPawnee_v1_1) (Lovell et al., 2021) and TAIR (http://www.arabidopsis.org), respectively. The Hidden Markov Model (HMM）profile for the WRKY domain (PF03106) was downloaded from the Pfam database (http://pfam.xfam.org/). Then HMMER3.0 program was used to search against pecan protein database with the E-value≦1e-5. Meantime, the Arabidopsis WRKY proteins used as the query, local BLASTp

were scanned for WRKY domains in pecan genome using BioEdit, the E value was set to 1e-2. The two data sets were merged to remove the repetitive sequence, then the NCBI-CDD (https://www.ncbi.nlm.nih.gov/cdd) were used to further verify. The characteristic of pecan WRKY proteins were analyzed using the ExPASy software (https://web.expasy.org/protparam/), and the WoLF PSORT (https://www.genscript.com/tools/wolf-psort) was used to predict the subcellular localization.

3.2 Phylogenetic tree analysis and classification of the pecan WRKY family

Multiple sequence alignments of WRKY domains of CiWRKY proteins were performed using BioEdit software. The WRKY proteins from pecan and Arabidopsis were compared using the ClustalW tool in MEGA5.0 software, the phylogenetic tree was constructed with neighbor joining (NJ) (Bootstrap=1000). The phylogenetic tree of full-length sequences of pecan WRKY proteins was built with the same method. The chromosome distribution map of pecan WRKY gene family was drew by TBtools software (Chen et al., 2020).

3.3 Motif analysis and exon-intron structures

The conserved motifs in the 89 CiWRKY proteins were detected by MEME (http://meme.nbcr.net/meme/cgibin/meme.cgi), with a maximum motif number of 15; the optimum motif width was 6-50 amino acid residues. The phylogenetic tree, gene structure, and conserved motif of WRKY family genes in pecan were visualized by TBtools software (Chen et al., 2020).

3.4 Expression analysis of *CiWRKY* genes during flower and fruit development

To reveal the expression pattern of *CiWRKY* genes during the flower development, the transcriptome data comes from our previous research, which contained early stage of female flower differentiation, female inflorescence differentiation stage, female flower involucre formation stage, bud stage, and female flower in full bloom (Wang et al., 2019). The *CiWRKYs* expression data (Fragments per kilobase of transcript per million mapped fragments, FPKM) during embryo development was downloaded from RNA transcriptome data (BioProject ID PRJNA435846, Huang et al., 2019). The FPKM values were used to estimate the expression level of each gene. The log2(FPKM) values of *CiWRKY* genes were used to draw heat maps by TBtools (Chen et al., 2020).

3.5 Category and number of *cis*-acting elements in the promoters of *CiWRKYs*

The 1 500 bp sequences upstream from the start codon of *CiWRKYs*, extracted from the pecan genome data by Tbtools, were labeled as putative promoter regions. The online program PlantCARE (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was used to analyze the *cis*-acting elements of ten selected *CiWRKYs* (Lescot et al., 2002).

Authors' contributions

WM was the executor of experimental design and research in this study. WM completed the data analysis and wrote the first draft of the manuscript. CJX, TX, and ZXW collected the data. BTY and ZCC guided experimental design and manuscript revision. All authors read and approved the final manuscript.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (32001350).

References

Chen C., Chen H., Zhang Y., Thomas H. R., Frank M. H., He Y., Xia R., 2020, TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data, Molecular plant, 13(8): 1194-1202
https://doi.org/10.1016/j.molp.2020.06.009

Chen F., Hu Y., Vannozzi A., Wu K. C., Cai H. Y., Qin Y., Mullis A., Lin Z. G., Zhang L. S., 2017, The WRKY Transcription Factor Family in Model Plants and Crops, Critical Reviews in Plant Sciences, 36(5-6): 311-335
https://doi.org/10.1080/07352689.2018.1441103

Ding Z., Yan J., Li G., Wu Z., Zhang S., Zheng S., 2014, WRKY41 controls Arabidopsis seed dormancy via direct regulation of *ABI3* transcript levels not downstream of ABA, Plant Journal, 79(5): 810-823
https://doi.org/10.1111/tpj.12597

Eulgem T., Rushton P. J., Robatzek S., Somssich I. E., 2000, The WRKY superfamily of plant transcription factors, Trends in Plant Science, 5(5): 199-206
https://doi.org/10.1016/S1360-1385(00)01600-9

Goel R., Pandey A., Trivedi P. K., Asif M. H., 2016, Genome-Wide Analysis of the Musa WRKY Gene Family: Evolution and Differential Expression during Development and Stress, Frontiers in plant science, 7: 299
https://doi.org/10.3389/fpls.2016.00299

He H., Dong Q., Shao Y., Jiang H., Zhu S., Cheng B., Xiang Y, 2012, Genome-wide survey and characterization of the WRKY gene family in *Populus trichocarpa*, Plant Cell Report, 31: 1199-1217
https://doi.org/10.1007/s00299-012-1241-0

Huang S., Gao Y., Liu J., Peng X., Niu X., Fei Z., Cao S., Liu Y., 2012, Genome-wide analysis of WRKY transcription factors in *Solanum lycopersicum*, Molecular Genetics and Genomics, 287: 495-513
https://doi.org/10.1007/s00438-012-0696-6

Huang R., Liu D., Huang M., Ma J., Li Z., Li M., Sui S., 2019, CpWRKY71, a WRKY Transcription Factor Gene of Wintersweet (*Chimonanthus praecox*), Promotes Flowering and Leaf Senescence in Arabidopsis, International journal of molecular sciences, 20(21): 5325
https://doi.org/10.3390/ijms20215325

Huang Y., Xiao L., Zhang Z., Zhang R., Wang Z., Huang C., Huang R., Luan Y., Fan T., Wang J., Shen C., Zhang S., Wang X., Randall J., Zheng B., Wu J., Zhang Q., Xia G., Xu C., Chen M., Zhang L., Jiang W., Gao L., Chen Z., Leslie C.A., Grauke L.J., Huang J., 2019, The genomes of pecan and Chinese hickory provide insights into *Carya* evolution and nut nutrition, GigaScience 8(5): giz036
https://doi.org/10.1093/gigascience/giz036

International Rice Genome Sequencing Project. P., 2005, The map-based sequence of the rice genome, Nature, 436(7052): 793-800
https://doi.org/10.1038/nature03895

Jiang J., Ma S., Ye N., Jiang M., Cao J., Zhang J., 2017, WRKY transcription factors in plant responses to stresses, Journal of Integrative Plant Biology, 59(2): 86-101
https://doi.org/10.1111/jipb.12513

Jiang W., and Yu D., 2009, *Arabidopsis WRKY2* transcription factor mediates seed germination and postgermination arrest of development by abscisic acid, BMC Plant Biol, 9(1): 1-14
https://doi.org/10.1186/1471-2229-9-96

Lei Y., Sun Y., Wang B., Yu S., Dai H., Li H., Zhang Z., Zhang J., 2020, Woodland strawberry WRKY71 acts as a promoter of flowering via a transcriptional regulatory cascade, Horticulture research, 7: 137
https://doi.org/10.1038/s41438-020-00355-4

Lescot M., Déhais P., Thijs G., Marchal K., Moreau Y., Van de Peer Y., Rouzé P., Rombauts S., 2002, PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic acids research, 30(1): 325-327
https://doi.org/10.1093/nar/30.1.325

Li W., Wang H., Yu D., 2016, Arabidopsis WRKY transcription factors WRKY12 and WRKY13 oppositely regulate flowering under short-day conditions, Molecular plant, 9(11), 1492-1503
https://doi.org/10.1016/j.molp.2016.08.003

Luo M., Dennis E. S., Berger F., Peacock W. J., Chaudhury A., 2005, *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, 102(48): 17531-17536
https://doi.org/10.1073/pnas.0508418102

Lovell J. T., Bentley N. B., Bhattarai G., Jenkins J. W., Sreedasyam A., Alarcon Y., Bock C., Boston L. B., Carlson J., Cervantes K., Clermont K., Duke S., Krom N., Kubenka K., Mamidi S., Mattison C. P., Monteros M. J., Pisani C., Plott C., Rajasekar S., Rhein H. S., Rohla C., Song M., Hilaire R. S., Shu S, Wells L, Webber J, Heerema R. J., Klein P. E., Conner P, Wang X, Grauke L. J., Grimwood J, Schmutz J, Randall J. J., 2021, Four chromosome scalegenomes and a pan-genome annotation to accelerate pecan tree breeding, Nature Communications, 12(1): 4125
https://doi.org/10.1038/s41467-021-24328-w

Ramamoorthy R., Jiang S. Y., Kumar N., Venkatesh P. N., Ramachandran S., 2008, A comprehensive transcriptional profiling of the WRKY gene family in rice under various abiotic and phytohormone treatments, Plant Cell Physiol, 49(6): 865-879
https://doi.org/10.1093/pcp/pcn061

Rushton P. J., Somssich I. E., Ringler P., Shen Q. J., 2010, WRKY transcription factors, Trends in Plant Science, 15(5): 247-258
https://doi.org/10.1016/j.tplants.2010.02.006

Song H., Wang P., Nan Z., Wang X., 2014, The WRKY transcription factor genes in *lotus japonicus*, International journal of genomics, 2014: 420128
https://doi.org/10.1155/2014/420128

Song H., Wang P., Hou L., Zhao S., Zhao C., Xia H., Li P., Zhang Y., Bian X., Wang X., 2016a, Global analysis of WRKY genes and their response to dehydration and salt stress in soybean, Frontiers in plant science, 7: 9
https://doi.org/10.3389/fpls.2016.00009

Song H., Wang P., Lin J. Y., Zhao C., Bi Y., Wang X., 2016b, Genome-wide identification and characterization of WRKY gene family in peanut, Frontiers in plant science, 7: 534
https://doi.org/10.3389/fpls.2016.00534

Tomato Genome C., 2012, The tomato genome sequence provides insights into fleshy fruit evolution, Nature, 485(7400): 635-641
https://doi.org/10.1038/nature11119

Tiika R. J., Wei J., Ma R., Yang H., Ma Y., 2020, Identification and expression analysis of the *WRKY* gene family during different developmental stages in *lycium ruthenicum* Murr. Fruit, Peer J, 8: e10207
https://doi.org/10.7717/peerj.10207

Ulker B., and Somssich I. E., 2004, WRKY transcription factors: from DNA binding towards biological function, Current Opinion in Plant Biology, 7(5): 491-498
https://doi.org/10.1016/j.pbi.2004.07.012

Wang H., Chen W., Xu Z., Chen M., Yu D., 2023, Functions of WRKYs in plant growth and development, Trends in plant science, S1360-1385(22)00335-1
https://doi.org/10.1016/j.tplants.2022.12.012

Wang M., Vannozzi A., Wang G., Liang Y. H., Tornielli, G. B., Zenoni S., Cavallini E., Pezzotti M., Cheng Z. M., 2014, Genome and transcriptome analysis of the grapevine (*Vitis vinifera* L.) WRKY gene family, Horticulture Research, 1: 16
https://doi.org/10.1038/hortres.2014.16

Wang M., Xi D., Chen Y., Zhu C., Zhao Y., Geng G., 2019, Morphological characterization and transcriptome analysis of pistillate flowering in pecan (*Carya illinoinensis*), Scientia Horticulturae, 257:108674
https://doi.org/10.1016/j.scienta.2019.108674

Yang Y., Zhou Y., Chi Y., Fan B., Chen Z., 2017, Characterization of Soybean WRKY Gene Family and Identification of Soybean WRKY Genes that Promote Resistance to Soybean Cyst Nematode, Scientific reports, 7(1): 17804
https://doi.org/10.1038/s41598-017-18235-8

Yin G., Xu H., Xiao S., Qin Y., Li Y., Yan Y., Hu Y., 2013, The large soybean (*Glycine max*) WRKY TF family expanded by segmental duplication events and subsequent divergent selection among subgroups, BMC plant biology, 13(1): 1-19
https://doi.org/10.1186/1471-2229-13-148

Yu Y., Hu R., Wang H., Cao Y., He G., Fu C., Zhou G., 2013, Mlwrky12, a novel miscanthus transcription factor, participates in pith secondary cell wall formation and promotes flowering, Plant Science, 212: 1-9
https://doi.org/10.1016/j.plantsci.2013.07.010

Yu Y., Liu S., Wang L., Kim S., Seo P., Qiao M., Wang N., Li S., Cao X., Park C., Xiang F., 2016, WRKY71 accelerates flowering via the direct activation of FLOWERING LOCUS T and LEAFY in *Arabidopsis thaliana*, Plant Journal, 85(1): 96-106
https://doi.org/10.1111/tpj.13092

Zhang C. Q., Xu Y., Lu Y., Yu H. X., Gu M. H., Liu Q. Q., 2011, The WRKY transcription factor OsWRKY78 regulates stem elongation and seed development in rice, Planta, 234(3): 541-554
https://doi.org/10.1007/s00425-011-1423-y

Zhang L., Chen L., Yu D., 2018, Transcription factor WRKY75 interacts with della proteins to affect flowering, Plant physiology, 176(1): 790-803

Zhang W. W., Zhao S. Q., Gu S., Cao X. Y., Zhang Y., Niu J. F., Liu L., Li A. R., Jia WS., Qi B. X., Xing Y., 2022, FvWRKY48 binds to the pectate lyase *FvPLA* promoter to control fruit softening in *Fragaria vesca*, Plant Physiology, 189(2): 1037-1049
https://doi.org/10.1093/plphys/kiac091