GENOME-WIDE INVESTIGATION OF TUNG TREE (*VERNICIA FORDII*) WRKY TRANSCRIPTION FACTORS: EVOLUTION, EXPRESSION AND CORRELATION WITH OIL-RELATED GENES

YOU, L. N.[1,2#] – WANG, C.[1,2#] – ZHAO, Q. Y.[1,2*]

[1]*Key Laboratory of Cultivation and Protection for Non-Wood Forest Trees, Ministry of Education, Central South University of Forestry and Technology, Changsha 410004, Hunan, China*

[2]*Key Lab of Non-wood Forest Products of State Forestry Administration, College of Forestry, Central South University of Forestry and Technology, Changsha 410004, Hunan, China (phone: +86-151-9945-2075; +86-139-7566-3890)*

[#] *Co-first authors*

**Corresponding author*
e-mail: 1057330638@qq.com; phone: +86-156-0108-3052

(Received 9th Jul 2022; accepted 14th Oct 2022)

Abstract. WRKY gene family is composed of plentiful members of plant transcription factors that function in various processes of development and physiology. Nonetheless, there was little knowledge available of the WRKY proteins in tung tree (*Vernicia fordii*), a critical economic oil crop. Benefiting from the previously released draft genome of tung tree by our lab, this study was carried out on characterization and expression patterns of the *VfWRKY* family genes at the whole-genome level. In this study, we identified 54 tung tree WRKY (*VfWRKY*) genes, namely from *VfWRKY1* to *VfWRKY54*. All of the VfWRKY domains were categorized into 3 major classes along with 9 subgroups. Analysis of transcriptome data showcases different expression patterns of *VfWRKY* genes in distinct tissues/organs and developmental stages. Correlation of expression between several *VfWRKY* genes and oil biosynthesis-related genes provides some clues to further investigate the functions of WRKY genes in tung tree. Taken together, our study reported the features of evolution, structure, expression patterns and association with important oil-related genes of the 54 *VfWRKY* genes at the whole genome level, serving as the basis for going into biological functions of *VfWRKY* genes in order to improve economic traits and develop new cultivars.
Keywords: *VfWRKY, expression pattern, tung oil, TFs, bioinformatics*

Introduction

Due to sessile lifestyles, members of the plant kingdom have developed kinds of complicated physiological mechanisms to perceive and react to diverse developmental and pressure signals. One of the indispensable parts is transcription factors (TFs), which are responsible for activating or repressing the target gene expression involved in the corresponding pathway through binding upstream nucleotide sequences specifically (Jin et al., 2014; Chen and Rajewsky, 2007). For instance, the WRKY proteins, which serve as the 7th largest TF family in flowering plants, function in diverse biological processes (Birkenbihl et al., 2017; Mirabella et al., 2015). Since the first *WRKY* gene called *SPF1* was isolated from eudicot crop *Ipomoea batatas* in 1994 (Ishiguro and Nakamura, 1994), another two WRKY proteins called ABF1 and ABF2 were reported in monocot crop *Avena fatua* one year later (Rushton et al., 1995). Up to now, there are increasingly more and more experimental identifications of WRKY

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2101_207228

TFs from various model plants, such as *Arabidopsis thaliana*, *Oryza sativa* (Wang et al., 2011; Dou et al., 2014) and so on.

The WRKY proteins are characterized by a conserved WRKY domain, which is required for binding to the W-box *cis*-acting element ([C/T]TGAC[C/T]) commonly located at promoters of genes, composed of about 60 amino acids possessing a prominent WRKYGQK heptapeptide and a zinc-finger motif (Eulgem et al., 2000). Owing to both the number of WRKY domains and the type of the zinc-finger motifs, the WRKY proteins can be divided into three primary groups. Group I comprises two WRKY domains each with a Cys_2His_2 zinc-finger motif, group II comprises one WRKY domain with a Cys_2His_2 zinc-finger motif and group III comprises one WRKY domain with a Cys_2Cys_2 zinc-finger motif (Bakshi and Oelmuller, 2014; Gu et al., 2018). Furthermore, group II can be split up into five subgroups on the basis of the feature of the additional motifs beyond the WRKY domain, such as leucine-rich repeat (LLR) (Rinerson et al., 2015; Rushton et al., 2012). Group III, which is regarded as the most highly evolved and the most adaptable (Zhang and Wang, 2005), can also be clustered into two subgroups in Arabidopsis and rice (Wu et al., 2005).

Ever since the first characterization of the WRKY protein, SPF1, in 1994 (Ishiguro and Nakamura, 1994), substantial experiments on WRKY proteins have been conducted in distinct plant species. There are 72, 102 and 57 components of the WRKY family reported in *Arabidopsis thaliana*, *Oryza sativa* and *Riccinus communis*, respectively (Eulgem et al., 2000; Wu et al., 2005; Li et al., 2012). Numerous studies prove that the WRKY proteins play an essential role in different aspects of plant development (Jiang et al., 2014; Grunewald et al., 2013; Li et al., 2016). For example, *AtWRKY75*, the first identified *WRKY* family gene involved in root architecture, modulates the number of lateral roots and root hairs negatively (Devaiah et al., 2007). VvWRKY26, a TTG2-like WRKY transcription factor from *Vitis vinifera*, is involved in the metabolism of flavonoid during berry growth (Amato et al., 2017). Noticeably, *WRKY* genes also present a regulatory effect on seed development (Xiang et al., 2017), such as seed coat formation and oil accumulation (Geilen et al., 2017). Mutation of *TRANSPARENT TESSAT GLABRA2* (*TTG2*), a member of *WRKY* class genes, contributes to seed coat dysplasia, including trichome, tannin and mucilage production (Johnson et al., 2002). *MINISEED3* (*MINI3*), a pollen- and endosperm-specific expressed WRKY transcription factor from Arabidopsis, modulates seed size and oil content (Fatihi et al., 2013; Luo et al., 2005). Suppression of *OsWRKY78* leads to secondary changes in crystal structure of endosperm starch (Zhang et al., 2011). Additionally, the WRKY proteins widely participate in the response to abiotic and biotic stresses (Kloth et al., 2016; Banerjee and Roychoudhury, 2015; Levée et al., 2009), including salt stress (Hu et al., 2013), drought (Liu et al., 2011), hormone response (Phukan et al., 2016), and rice blast fungus (Chujo et al., 2013).

Tung tree, *Vernicia fordii*, is viewed as one of the four crucial wooden oil crops in China, together with tea oil tree (*Camellia oleifera*), walnut (*Juglans regia L.*) and Chinese tallow tree (*Sapium sebiferum*) (Li et al., 2017). Resulting from oxidation, quick drying and saline-alkaline resistance features, tung oil coming from seed kernels is not only widely applied as raw materials to traditional varnishes and paints, but also devoted to high-quality thermosetting polymers synthesis (Zhang et al., 2014; Park et al., 2008). Therefore, the study on the mechanism of tung oil biosynthesis has attracted large researchers' attention, underlying tung tree genetic improvement and new cultivars development. To date, functions of some significant oil-related genes have

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2101_207228

been reported, including *diacylglycerol acyltransferases* (*DGAT*) (Cao et al., 2012) and *delta-12 fatty acid conjugase* (*FADx*) (Dyer et al., 2002), but the mechanism still remains elusive owing to its complexity.

With regard to the well-annotated genomic data of Arabidopsis, the *AtWRKY* genes have been widely investigated in this model creature. However, at present, the information about *VfWRKY* still remains unknown at the whole genome-wide level. Recently, our lab published the draft genome of tung tree, providing the opportunity to conduct genome wide characterization of *VfWRKY* gene family members. In this work, we identified 54 *VfWRKY* genes and categorized them into three major groups according to sequence alignment and phylogenic analysis compared with the *AtWRKY* genes in the model plant, Arabidopsis. Gene structures, conserved motifs and chromosome location were investigated as well. To detect the candidate *VfWRKY* genes involved in some specific biological pathways, in particular of fatty acid formation and oil accumulation, gene expression patterns analysis in distinct tissues during different stages of development and correlation analysis of expression between chief oil-related genes and *VfWRKY* genes were carried out. This study offers the basic knowledge of *WRKY* family from tung tree to further research on the biological and molecular functions.

Materials and methods

Whole genome-wide identification of putative WRKY family genes in tung tree

Previously, our lab has reported the genome of tung tree (Zhang et al., 2019). To identify members of the *WRKY* family at the whole-genome level, the HMM profile corresponding to the WRKY (PF03106) was downloaded from the Pfam database (http://pfam.xfam.org/) (Finn et al., 2016) and served as a query to search the WRKY proteins from the tung tree proteome file via HMMER 3.0 software (http://www.hmmer.org/) (Finn et al., 2015). Then, 58 candidate protein sequences were obtained and submitted to SMART (http://smart.embl-heidelberg.de/) and CDD (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi) to validate the conserved WRKY domain. 54 genes were finally assigned as tung tree *WRKY* genes (*VfWRKY*) after the redundant sequences and the sequences without the core WRKY heptapeptide manually discarded. These *WRKY* genes were named as *VfWRKY1* to *VfWRKY54* based on their positions on the chromosomes. Besides, bioinformatics analysis of *VfWRKYs* were carried out by WoLF PSORT (https://wolfpsort.hgc.jp/) and ExPASy (http://www.expasy.ch/tools/pi_tool.html), including the length of opening reading frames (ORFs) and protein sequences, molecular weights, theoretical isoelectric points (pI) and prediction of subcellular localization.

Phylogenic tree construction based on the WRKY domains from tung tree and Arabidopsis

The amino acid sequences of WRKY domains from 54 *VfWRKY* genes and 71 *AtWRKY* genes, which were downloaded from TAIR10 (http://www.arabidopsis.org/), are subjected to multiple sequence alignment using ClustalW by MEGA 7.0 with default parameters. On the basis of the alignment results, the phylogenic tree was constructed via the neighbor-joining (NJ) algorithm by MEGA 7.0 with the pairwise deletion calculated under the Poisson model. And the bootstrap value was set to 1000 iterations. The initial phylogenic result was labelled manually by using the online program EvolView

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2101_207228

(https://www.evolgenius.info/evolview/#login) (He et al., 2016). Resulting from the *AtWRKY* classification principle and the phylogenic analysis between *VfWRKY* and *AtWRKY* proteins (Eulgem et al., 2000), VfWRKY family members were split into three groups (I, II, III) and five subgroups in group II (II a, II b, II c, II d, II e).

Multiple sequence alignment

The amino acid sequences of all the identified WRKY domains from tung tree and 8 selected WRKY domains from Arabidopsis were used to conduct alignment via ClustalW with the default settings. Then, the core residues in the WRKY domains were edited manually by GeneDoc (https://www.softpedia.com/get/Science-CAD/GeneDoc.shtml).

Intron-exon structure and conserved motifs analyses of VfWRKY family

The GFF3 file and WRKY domain position information for tung tree were used to portray the gene structures via TBTOOLS (http://www.tbtools.com/) and GSDS2.0 online tools (http://gsds.cbi.pku.edu.cn/index.php). The website MEME (version 5.0.5, http://meme-suite.org/tools/meme) (Bailey and Elkan, 1994) for structural characterization was used to identify the conserved motifs in VfWRKY proteins, with the following proper parameters: the number of repetitions, any; the maximum number of motifs, 20; the minimum motif width, 6 residues; the maximum motif width, 200 residues (Cheng et al., 2018; Wang et al., 2017). The result was further embellished by TBTOOLS software. Finally, the above datum was manually scheduled in accordance with order exhibited by the multiple sequence alignment.

WRKY genes distribution on chromosomes and duplication

To represent the *VfWRKY* genes assignment on chromosomes, MG2C (http://mg2c.iask.in), a bioinformatic website, was utilized to draft the physical location of genes on chromosomes. Subsequently, Adobe Illustrator software (https://www.adobe.com/cn/products/illustrator/free-trial-download.html) was applied to embellish and enrich the original result, including adding the tandem and segmental duplication information according to our lab's previously published data of collinearity analysis in tung tree genome (Zhang et al., 2019). The highly similar homologous genes located in the area no more than 200 kb on chromosome are identified as tandem duplicated genes.

Expression analyses of tung tree WRKY genes

The raw transcriptome data acquired from varieties of tung tree tissues were depicted in Ling et al. (Zhang et al., 2019). Fragments per kilobase of exon model per million mapped reads (FPKM) was applied as measurement of the expression abundance of tung tree WRKY genes. The expression pattern was represented as heatmaps which was generated by HemI software (version 1.0.3.7) on the basis of the $\log_2(FPKM + 1)$-transformed data (Deng et al., 2014).

Correlation analysis of expression

We utilized our lab's previously published transcript data of 88 oil-related genes predicted as essential genes in fatty acid formation and oil accumulation (Zhang et al.,

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2101_207228

2019). Analysis of expression associations between 9 *VfWRKY* genes (*VfWRKY6*, *-7*, *-15*, *-19*, *-25*, *-31*, *-41*, *-44*, *-50*) and 88 oil-related genes was performed via the software SPSS (version 24) with the following parameters: Bivariate correlations, Pearson correlation, Two-tailed, Flag significant correlations.

Results

Identification of WRKY transcription factors in tung tree

To identify the WRKY proteins in tung tree at the whole-genome level, the hidden Markov model (HMM) profile corresponding to the WRKY (PF03106) was used as a query to search the proteome via MEGA 7.0 software. Originally, we obtained 58 putative WRKY protein sequences. Then, 4 items (*Vf03G2493*, *Vf07G2227*, *Vf06G1610* and *Vf02G1129*) without WRKY core sequences were removed after we checked the existence of the WRKY domain of the 58 predicted sequences by SMART. Ultimately, 54 genes were annotated as members of *VfWRKY* family. 51 of the validated *VfWRKY* genes were named from *VfWRKY1* to *VfWRKY51* because of their locations on chromosomes. Another 3 *VfWRKY* genes (*Vf00G0557*, *Vf00G1604* and *Vf00G1869*) failed to be mapped to chromosomes were renamed from *VfWRKY52* to *VfWRKY54* (*Table 1*).

Table 1. *The basic information of WRKY transcription factor family in V. fordii (open reading frame, ORF; size of the protein, Size; molecular weight, MW; isoelectric point, PI; chromosome, Chr)*

Gene name	Gene ID	ORF (bp)	Size (aa)	MW (Da)	pI	Subcellular location	Chr	Conserved motif	Znf motif pattern	Znf motif type
VfWRKY1	Vf01G0251	1023	340	37436.06	9.50	Nuclear	1	WRKYGQK	C-X5-C-X23-H-X-H	C2H2
VfWRKY2	Vf01G1028	669	222	24595.10	6.75	Nuclear	1	WRKYGKK	C-X4-C-X23-H-X-H	C2H2
VfWRKY3	Vf01G1060	711	236	26890.23	9.10	Nuclear	1	WRKYGQK	C-X4-C-X23-H-X-H	C2H2
VfWRKY4	Vf01G1631	1770	589	63810.37	5.19	Nuclear	1	WRKYGQK	C-X5-C-X23-H-X-H	C2H2
VfWRKY5	Vf01G1758	1458	485	53108.94	6.30	Nuclear	1	WRKYGQK, WRKYGQK	C-X4-C-X22-H-X-H, C-X4-C-X23-H-X-H	C2H2, C2H2
VfWRKY6	Vf01G2623	642	213	24097.85	9.23	Nuclear	1	WRKYGQK	C-X4-C-X23-H-X-H	C2H2
VfWRKY7	Vf01G2658	2232	743	81402.76	5.83	Nuclear	1	WRKYGQK, WRKYGQK	C-X4-C-X22-H-X-H, C-X4-C-X23-H-X-H	C2H2, C2H2
VfWRKY8	Vf02G0697	921	306	35122.42	9.06	Nuclear	2	WRKYGQK	C-X4-C-X23-H-X-H	C2H2
VfWRKY9	Vf02G1874	957	318	35596.59	8.14	Nuclear	2	WRKYGQK	C-X5-C-X23-H-X-H	C2H2
VfWRKY10	Vf02G1938	1281	426	48442.26	6.38	Cytoplasm	2	WRKYGGR	C-X4-C-X22-H-X-H	C2H2
VfWRKY11	Vf03G0276	1656	551	61477.54	6.31	Nuclear	3	WRKYGQK	C-X5-C-X23-H-X-H	C2H2
VfWRKY12	Vf03G2491	804	267	29560.00	6.61	Nuclear	3	WRKYGQK	C-X7-C-X23-H-X-C	C2HC
VfWRKY13	Vf04G0042	1434	477	52253.87	8.93	Nuclear	4	WRKYGQK, WRKYGQK	C-X4-C-X22-H-X-H, C-X4-C-X23-H-X-H	C2H2, C2H2
VfWRKY14	Vf04G2205	1794	597	65126.98	6.16	Nuclear	4	WRKYGQK	C-X5-C-X23-H-X-H	C2H2
VfWRKY15	Vf05G1414	1599	532	57694.12	7.04	Nuclear	5	WRKYGQK, WRKYGQK	C-X4-C-X22-H-X-H, C-X4-C-X23-H-X-H	C2H2, C2H2
VfWRKY16	Vf05G1621	525	174	19566.84	10.21	Nuclear	5	WRKYGQK	-	-
VfWRKY17	Vf05G1892	969	322	36353.40	5.00	Nuclear	5	WRKYGQK	C-X5-C-X23-H-X-H	C2H2
VfWRKY18	Vf06G1476	888	295	32374.52	6.01	Nuclear	6	WRKYGQK	C-X4-C-X23-H-X-H	C2H2
VfWRKY19	Vf06G1505	1632	543	58664.63	8.44	Chloroplast	6	WRKYGQK, WRKYGQK	C-X4-C-X22-H-X-H, C-X4-C-X23-H-X-H	C2H2, C2H2
VfWRKY20	Vf06G1608	2364	787	85731.44	6.90	Plasma membrane	6	WRKYGQK	C-X5-C-X23-H-X-H	C2H2
Gene name	**Gene ID**	**ORF (bp)**	**Size (aa)**	**MW (Da)**	**pI**	**Subcellular location**	**Chr**	**Conserved motif**	**Znf motif pattern**	**Znf motif type**
VfWRKY21	Vf06G1750	1044	347	38982.10	9.75	Nuclear	6	WRKYGQK	C-X5-C-X23-H-X-H	C2H2

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2101_207228

VfWRKY22	Vf06G2170	825	274	31018.29	5.05	Nuclear	6	WRKYGQK	C-X5-C-X23-H-X-H	C2H2
VfWRKY23	Vf06G2255	1449	482	51977.35	7.64	Nuclear	6	WRKYGQK	C-X5-C-X23-H-X-H	C2H2
VfWRKY24	Vf06G2418	951	316	36321.39	5.66	Nuclear	6	WRKYGQK	C-X7-C-X23-H-X-C	C2HC
VfWRKY25	Vf06G2465	570	189	21673.48	9.11	Peroxisome	6	WRKYGQK	C-X4-C-X23-H-X-H	C2H2
VfWRKY26	Vf06G2485	984	327	37516.77	5.32	Nuclear	6	WRKYGQK	C-X7-C-X23-H-X-C	C2HC
VfWRKY27	Vf06G2813	657	218	25137.53	8.16	Nuclear	6	WRKYGQK	C-X4-C-X23-H-X-H	C2H2
VfWRKY28	Vf07G1061	828	275	31015.14	5.63	Nuclear	7	WRKYGQK	C-X5-C-X23-H-X-H	C2H2
VfWRKY29	Vf07G1146	2160	719	83386.72	8.85	Plasma membrane	7	WRKYGQK	-	-
VfWRKY30	Vf07G1454	633	210	23865.38	4.93	Nuclear	7	WRKYGQK	C-X4-C-X23-H-X-H	C2H2
VfWRKY31	Vf07G1681	1770	589	64970.16	6.54	Nuclear	7	WRKYGQK, WRKYGQK	C-X4-C-X22-H-X-H, C-X4-C-X23-H-X-H	C2H2, C2H2
VfWRKY32	Vf07G1743	999	332	37348.17	5.95	Nuclear	7	WRKYGQK	C-X7-C-X23-H-X-C	C2HC
VfWRKY33	Vf07G2207	1077	358	40104.89	9.64	Peroxisome	7	WRKYGQK	C-X5-C-X23-H-X-H	C2H2
VfWRKY34	Vf08G0205	1386	461	50309.64	8.60	Nuclear	8	WRKYGQK, WRKYGQK	C-X4-C-X22-H-X-H, C-X4-C-X23-H-X-H	C2H2, C2H2
VfWRKY35	Vf08G1341	495	164	18680.68	5.72	Nuclear	8	WRKYGKK	C-X4-C-X23-H-X-H	C2H2
VfWRKY36	Vf08G1959	984	327	36416.82	6.38	Nuclear	8	WRKYGQK	C-X5-C-X23-H-X-H	C2H2
VfWRKY37	Vf08G1960	702	233	25998.36	9.15	Nuclear	8	WRKYGQK	C-X5-C-X23-H-X-H	C2H2
VfWRKY38	Vf08G1964	1416	471	51786.82	9.04	Nuclear	8	WRKYGQK, WRKYGQK	C-X4-C-X23-H-X-H, C-X4-C-X23-H-X-H	C2H2, C2H2
VfWRKY39	Vf08G2061	1032	343	37126.94	9.51	Nuclear	8	WRKYGQK	C-X5-C-X23-H-X-H	C2H2
VfWRKY40	Vf09G1471	603	200	23027.41	6.16	Nuclear	9	WRKYGKK	C-X4-C-X23-H-X-H	C2H2
VfWRKY41	Vf09G1513	1761	586	64383.30	6.91	Nuclear	9	WRKYGQK, WRKYGQK	C-X4-C-X22-H-X-H, C-X4-C-X23-H-X-H	C2H2, C2H2
VfWRKY42	Vf09G1623	1068	355	39077.00	9.59	Nuclear	9	WRKYGQK	C-X5-C-X23-H-X-H	C2H2
VfWRKY43	Vf10G0283	1044	347	38627.84	5.26	Nuclear	10	WRKYGQK	C-X7-C-X23-H-X-C	C2HC
VfWRKY44	Vf10G0746	447	148	17170.14	8.90	Nuclear	10	WRKYGQK	C-X4-C-X23-H-X-H	C2H2
VfWRKY45	Vf10G1102	1905	634	68952.04	6.97	Nuclear	10	WRKYGQK	C-X5-C-X23-H-X-H	C2H2
VfWRKY46	Vf10G1716	1398	465	50324.36	5.26	Nuclear	10	WRKYGQK	C-X5-C-X23-H-X-H	C2H2
VfWRKY47	Vf11G0033	972	323	36955.21	8.97	Nuclear	11	WRKYGQK	C-X4-C-X23-H-X-H	C2H2
VfWRKY48	Vf11G0043	375	124	14296.55	6.04	Nuclear	11	WRKYGQK	C-X4-C-X23-H-X-H	C2H2
VfWRKY49	Vf11G1036	1512	503	55887.80	6.51	Nuclear	11	WRKYGQK	C-X5-C-X23-H-X-H	C2H2
VfWRKY50	Vf11G1123	1728	575	62448.81	6.66	Nuclear	11	WRKYGQK, WRKYGQK	C-X4-C-X22-H-X-H, C-X4-C-X23-H-X-H	C2H2, C2H2
VfWRKY51	Vf11G1991	1005	334	37032.02	6.01	Nuclear	11	WRKYGQK	C-X4-C-X23-H-X-H	C2H2
VfWRKY52	Vf00G0557	1311	436	46888.33	4.96	Nuclear	-	WRKYGQK	C-X5-C-X23-H-X-H	C2H2
VfWRKY53	Vf00G1604	1752	583	63054.58	5.83	Nuclear	-	WRKYGQK	C-X5-C-X23-H-X-H	C2H2
VfWRKY54	Vf00G1869	969	322	36082.25	6.56	Nuclear	-	WRKYGQK	C-X4-C-X23-H-X-H	C2H2

Sequence analysis of the *VfWRKY* genes were performed, such as the molecular weight (MV) of the protein, the subcellular localization (*Table 1; Appendix 1*) and so on. The result indicates that the 54 VfWRKY proteins vary widely in MV, from 14.3 kDa to 85.7 kDa. VfWRKY48 is the smallest one with only 124 amino acids, while VfWRKY20 is the largest one with 787 amino acids. Among the 54 VfWRKY proteins, 48 proteins are predicted to locate in the nuclear region, 2 proteins (VfWRKY20 and VfWRKY29) are located in the plasma membrane, 2 proteins (VfWRKY25 and VfWRKY33) are located in the peroxisome, 1 protein (VfWRKY10) is located in the cytoplasm and 1 protein (VfWRKY19) is located in the chloroplast. Besides, the open reading frame (ORF) length, the number of amino acids, pI-values and the chromosomal location are summarized in *Table 1*.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2101_207228

Multiple sequence alignment and classification of WRKY genes in tung tree

Classification of *VfWRKY* genes is based on the sequences of WRKY domains, which are characterized by approximately 60 amino acid residues with a highly conserved WRKYGQK signature at the N-terminal and a zinc finger motif at the C-terminal (Bakshi and Oelmuller, 2014; Ulker and Somssich, 2004). The amino acid sequences of WRKY domains from the 54 VfWRKY proteins and 7 AtWRKY (AtWRKY4, AtWRKY15, AtWRKY23, AtWRKY29, AtWRKY40, AtWRKY46, AtWRKY47, AtWRKY62) proteins, which were chosen from different clades as representatives, were subjected to multiple sequence alignment. *Figure 1* shows that 50 of 54 VfWRKY proteins contain the WRKYGQK heptapeptide core sequences, whereas the WRKY signature sequences are replaced by one or two different amino acids in another 4 VfWRKY proteins. In VfWRKY10, the core heptapeptide sequences in the WRKY domain is WRKYGGR; and in VfWRKY2, -35 and -40, WRKYGKK. Moreover, a total of 52 VfWRKY proteins possess a zinc-finger motif in each WRKY domain with the exception of VfWRKY16 and -29 (*Fig. 1*). Nonetheless, VfWRKY16 and -29 were still involved in our subsequent study, as conducted by Hongsheng et al. in poplar and Mangelsen et al. in barley (He et al., 2012; Mangelsen et al., 2008).

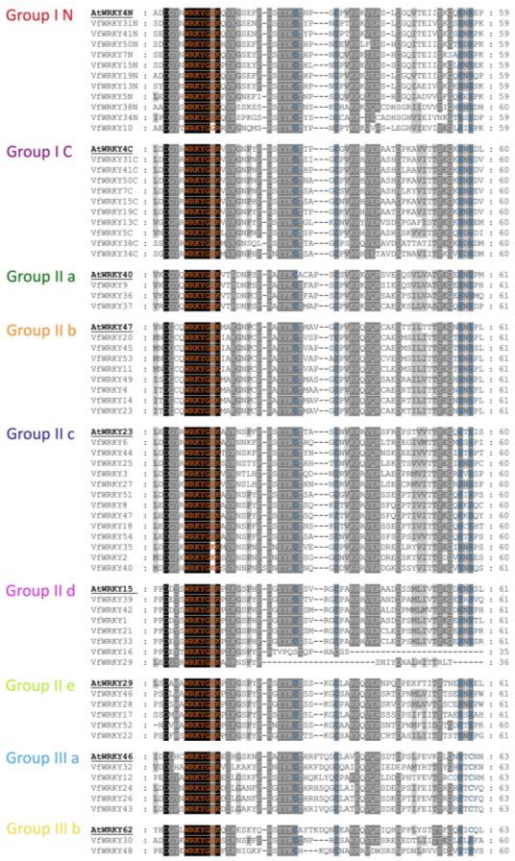

Figure 1. *Multiple sequence alignment of VfWRKY and AtWRKY proteins. Alignment of the multiple sequences of all WRKY domains from tung tree and 9 representative WRKY domains from Arabidopsis via ClustalW. The suffix –N or –C is separately corresponding to the N-terminal or C-terminal WRKY domain. The signature WRKYGQK of the WRKY domain and the core histidine and cystine of the zinc-finger motif are marked in orange and blue color, respectively*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2101_207228

In order to gain a better understanding of the classification and the evolutionary relationship of the 54 VfWRKY proteins, the phylogenetic analysis was performed by using 84 AtWRKY domains from distinct groups and subgroups as representatives. As shown in *Figure 2*, in total, 64 VfWRKY domains are categorized into 3 major groups, namely Group I, II and III, as described by Eulgem et al. (2000) in Arabidopsis. All of the 21 members from group I share the identical type of zinc-finger motif, a C2–H2 pattern (C-X_4-C-$X_{22–23}$-H-X-H). The 36 VfWRKY domains belong to group II which is the largest group and further falls into five subgroups, namely II a, II b, II c, II d and II e. Almost all members of group II contain a C_2–H_2 type zinc-finger motif (C-X_{4-5}-C-$X_{22–23}$-H-X-H) apart from two members (VfWRKY16, VfWRKY19) from subgroup II b. In terms of the phylogenic tree, group II a with 3 members and group II b with 8 members are clustered from one clade, group II d with 7 members and group II e with 5 members are clustered from one clade. Besides, tung tree group III are aligned into two subgroups with 5 in subgroup III a and 2 in III b. Instead of a C_2–HC-type zinc finger motifs (C-X_7-C-X_{23}-H-X-C), the 2 members (VfWRKY30, VfWRKY48) in subgroup III b harbor a C_2H_2 (C-X_4-C-X_{23}-H-X-H) pattern of zinc ion chelating finger motif. Varying from monocot species, the group III proteins in tung tree do not manifest extended regions in their zinc finger portions of the WRKY domains (Tripathi et al., 2012).

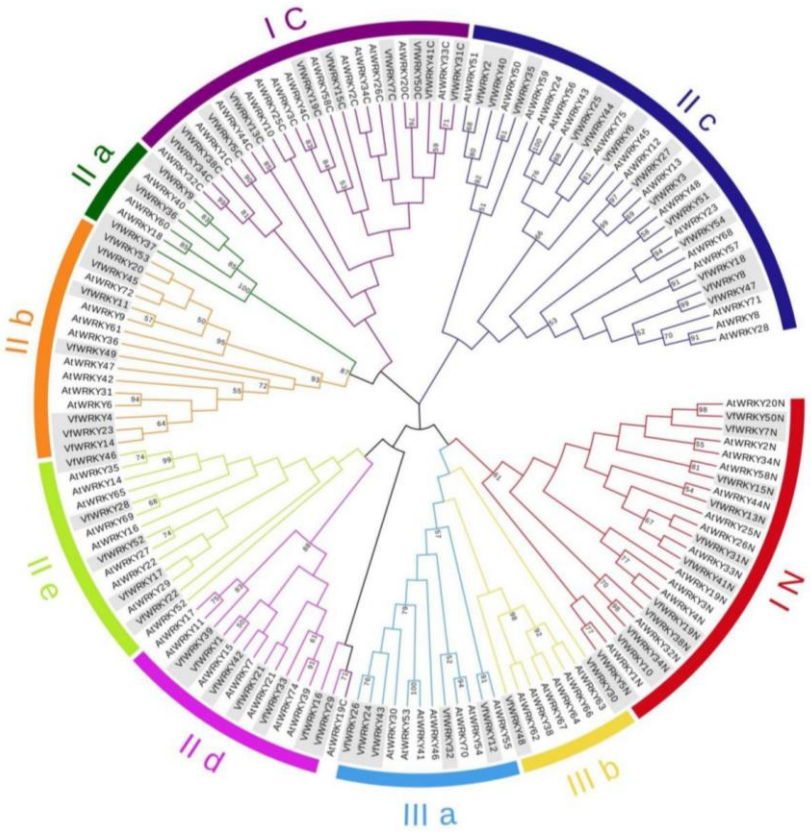

Figure 2. *Phylogenic analysis of the WRKY domains of 54 VfWRKY and 71 AtWRKY proteins. The amino acid sequences of the WRKY domains were subjected to alignment by ClustalW and the neighbor-joining bootstrap consensus tree was constructed by MEGA 7.0. Three main groups (I, II, III) and seven subgroups (II a, II b, II c, II d, II e, III a, III b) are tagged by distinctive colored branches and arcs. Sequences from group I N and I C are corresponding to the N-terminal and C-terminal WRKY domains, respectively. All of the 64 VfWRKY domains are highlighted by grey background color to distinguish from the 84 AtWRKY domains*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2101_207228

Exhibition of gene structure and analysis of conserved motifs

To get a further understanding of evolutionary relationships among the 54 WRKY components in tung tree, the analysis of intron-exon organization was performed by TBTOOLS software. As exhibited in *Figure 3a,* no *VfWRKY* genes are intronless and the quantity of introns varies widely, from 1 to 18. But majority of them harbor 1 to 5 introns, 26 (48.1%) with 2 introns, 10 (18.5%) with 4 introns, 7 (13.0%) with 3 introns, 5 (9.3%) with 5 introns, 3 (5.6%) with 1 intron. The remaining 3 genes (*VfWRKY10*, *-20* and *-29*) possess 10, 12, 18 introns, respectively. With more insight into the sequences coding the WRKY domains, we noticed that other than group I N, plenty of WRKY domains are separated by an intron. In concordance with the previous analyses in other plant species (Rinerson et al., 2015), the introns broadly distributed in the WRKY domains in tung tree are also mainly divide into two classes, R-type and V-type intron. (*Appendix 2*). *The R*-type intron, which is located within the valine codon in order to set the WRKY hallmark part apart from the zinc finger motif, is widely spread in group I C, group II c-II e, group III. Whereas the V-type intron followed by the valine which corresponds to the 6th residue after the 2nd core cysteine in the zinc finger motif, only appears in group II a and II b.

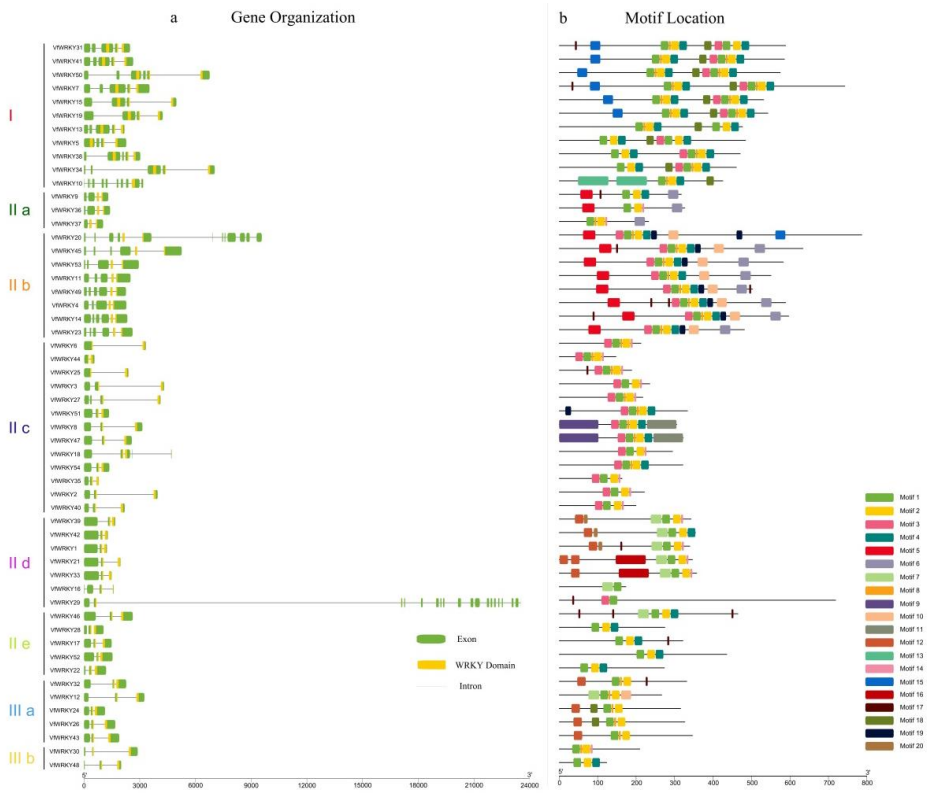

Figure 3. *Structural characterization of WRKY family in the tung tree. A. Gene structure of the 54 VfWRKY genes. The GFF3 file and WRKY domain location information for Tung tree were used to draw gene structures by TBTOOLS software. The green roundrects, black lines, yellow boxes stand for exons, introns and WRKY domains, respectively. The full-length genomic sequences can be measured by the graduated scale at the bottom. B. Patterns of the conserved motifs in the 54 VfWRKY proteins. The 20 different conserved motifs identified via MEME online program are labeled in different colored roundrects. The length of the proteins is accessible through the scale at the bottom of the figure. Appendix 3 elaborates on the sequence information of the 20 conserved motifs*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2101_207228

The composition of conserved motifs in VfWRKY proteins were examined by the MEME online system and 20 motifs were detected whose lengths range from 6 to 103 residues. The specific sequence information of the predicted motifs is summarized in *Appendix 3.* As shown in *Figure 3b*, except for VfWRKY16 and VfWRKY29 lacking of the zinc-finger motifs in the WRKY domain, the majority of VfWRKY proteins contain motif 1 in combination with motif 2 located in the WRKY domain. Apart from the motif 1 and 2 widespread, some other motifs just exist in some particular groups and subgroups, which may endow the corresponding members with specific functions. For instance, motif 5 and 6 merely occur in subgroup II a and II b; motif 12, only in subgroup II d and III a; motif 18, mostly in group I; motif 19 and 10, mainly in subgroup II b. Thus, combined with our results of exon-intron structures and phylogenetic analysis, the category of the WRKY family in tung tree is credible.

Chromosomal distribution and gene duplication of tung tree WRKY genes

To achieve the information of the physical locations of the *VfWRKY* genes, we mapped the genes on the chromosomes. *Figure 4* shows that the *VfWRKY* genes are widely distributed on the 11 chromosomes and the number of genes on each chromosome varies from 2 to 10. Chromosome 6 contains the largest number of genes (10 genes), followed by chromosome 1 (7 genes) which is the longest chromosome. 6 genes are present on chromosome 7 and 8, respectively. 5 and 4 genes separately dwell on chromosome 11 and 10. Chromosome 2, 5 and 9 are populated with 3 genes, whereas chromosome 3 and 4 only harbor the least quantity (2 genes). Although chromosome 10 is the shortest chromosome, it has more genes than some other ones (chromosome 2, 5, 9, 3, and 4), suggesting that the number of *VfWRKY* genes is not positively correlated with the length of chromosomes. Besides, the three major groups of the *VfWRKY* genes do not show preference for chromosomal allocation. *Figure 4* shows that group II members are assigned randomly across 11 chromosomes. Except for chromosome 3 and 10, the remaining 9 chromosomes have group I members. At the same time, the cause of group III members merely identified on several chromosomes (chromosome 3, -6, -7, -10, -11) may belong to the small amount of members in this group.

Expression patterns analysis of WRKY genes across different tissues in tung tree

To investigate the expression profiles of *VfWRKY* genes in different tissues/organs and different developmental stages, we used the published RNA sequencing data from 16 plant samples, including seeds, male flowers, female flowers, stems, leaves and roots, to construct a heatmap through HemI software (*Fig. 5; Appendix 4) (Zhang* et al., 2019). 34 *VfWRKY* genes are expressed throughout the 16 tested samples (FPKM > 0), among which 16 genes are expressed constitutively (FPKM > 1). Some *VfWRKY* genes display preference of expression for some tested tissues. For example, 10 genes (*VfWRKY7, -15, -20, -21, -33, -34, -38, -40, -52, -54*) exhibit the most abundant transcript levels in seed, 11 (*VfWRKY1, -4, -5, -6, -10, -13, -19, -29, -39, -41, -44*) in male flower, 7 (*VfWRKY9, -26, -31, -35, -37, -43, -51*) in female flower, 7 (*VfWRKY3, -18, -23, -27, -42, -46, -50*) in stem, 5 (*VfWRKY8, -25, -30, -32, -47*) in leaf, 14 (*VfWRKY2, -11, -12, -14, -16, -17, -22, -24, -28, -36, -45, -48, -49, -53*) in root. Besides, some remarkable tendency of *WRKY* gene expression are shown in various developmental stages. For instance, the transcript level of *VfWRKY19* is gradually rising

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2101_207228

across different developmental stages of male flowers. The expression quantities of *VfWRKY3* and *VfWRKY21* are reduced bit by bit during the growth and development of female flowers.

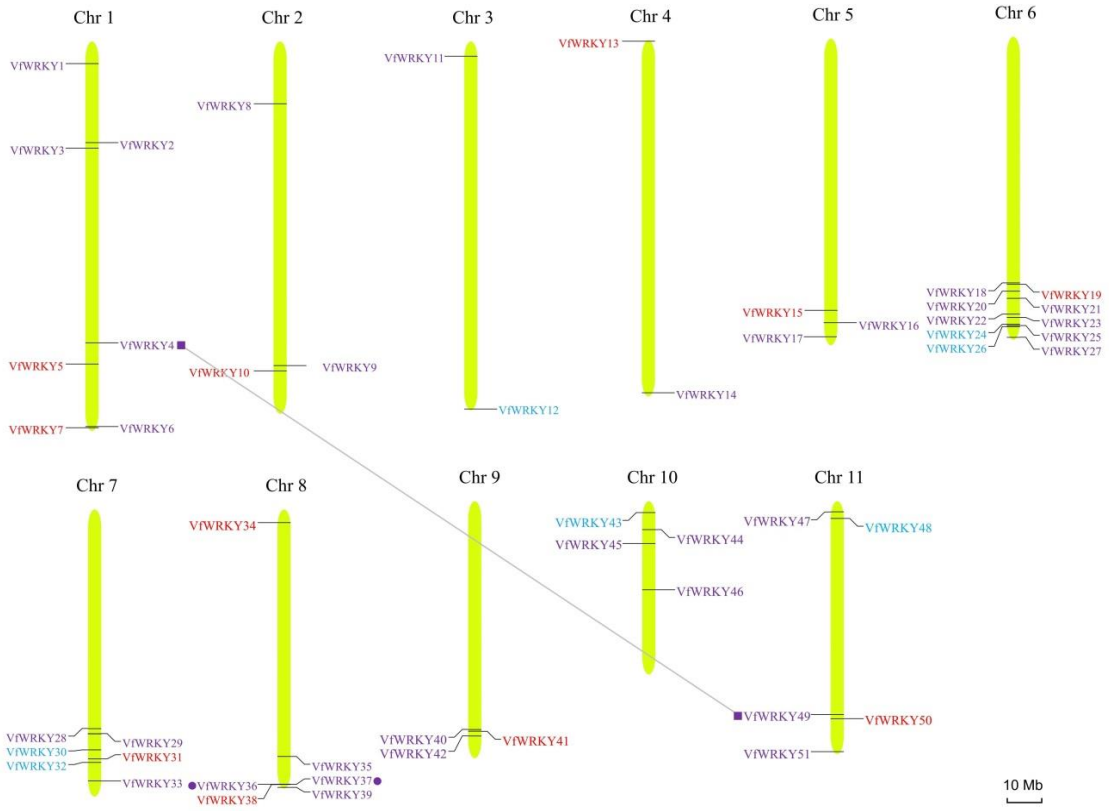

Figure 4. *Schematic representations for the physical location and region duplication of the 51 WRKY genes on 11 chromosomes in tung tree. The physical location of genes on chromosomes were mapped by MG2C bioinformatics website, and Adobe illustrator software was used to embellish and enrich the original results. The physical locations of the 51 VfWRKY genes on chromosomes are determined by the tung tree genome. The number of chromosomes is shown at the top of the corresponding chromosome (chr, short for chromosome) and each chromosome is represented by yellowish green bars. The scale is indicated at the figure lower right. Genes belonged to group I, II, III are separately painted in red, purple and blue colors. The purple circles flag the segmentally duplicated genes and the purple squares flag the tandemly duplicated genes, which are also associated by a grey line*

Correlation analysis of WRKY genes and oil biosynthesis-related genes in tung tree seeds

As an economically oil crop, the mechanism of fatty acid synthesis and oil formation is always a hot research in tung tree. Besides several functional genes investigated previously, such as *fatty acid desaturase* (*FAD*) and *diacylglycerol acyltransferase* (*DGAT*) (Cao et al., 2012, 2013; Dyer et al., 2002; Shockey et al., 2006), our lab has previously predicted 88 crucial oil biosynthesis-related genes, including *acetyl CoA carboxylase* (*ACCase*), *phosphoenolpyruvate carboxylase* (*PEPC*), *palmitoyl-acyl carrier protein thioesterase* (*FATB*), *long chain fatty aycl-CoA synthetases* (*LACS*),

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2101_207228

glycerol-3-phosphate acyltransferase (*GPAT*), *1-acyl-sn-glycerol-3-phosphate acyltransferase* (*LPAT*), *phosphatidate phosphatase* (*PP*), *diacylglycerol cholinephosphotransferase* (*DAG-CPT*), *phospholipid diacylglycerol acyltransferase* (*PDAT*) and *oleosins* (*OLEs*) (Zhang et al., 2019). In Arabidopsis, mutated (knockout) seeds in MINISEED3 gene (a *WRKY10* transcription factor) are much smaller owing to a decline in oil content (Fatihi et al., 2013), and loss of *AtWRKY43* gene leads to the increase of polyunsaturated fatty acids (Geilen et al., 2017). In tung tree, *VfWRKY7*, *-15*, *-19*, *-31*, *-41*, *-50* are the orthologs of *MINISEED3*, whereas *VfWRKY6*, *-25*, *-44* are orthologous to *AtWRKY43*, indicating the above 9 *VfWRKY* genes may be involved in the process of fatty acid formation and oil accumulation due to the fact that closely related genes are likely to possess functions in common (Li et al., 2003).

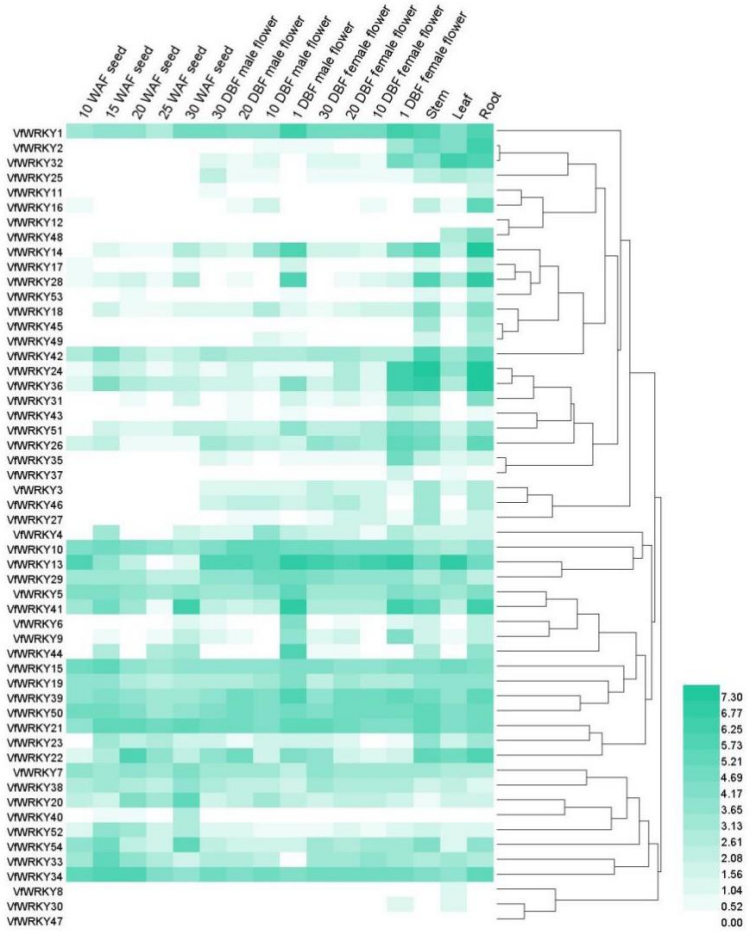

Figure 5. *Expression patterns of tung tree WRKY genes in distinct tissues at different developmental stages. The gene expression levels from 16 samples derived from different tissues and developmental stages were displayed by heatmap created by HemI software after transforming the FPKM values to space*

To determine the expression relations between 9 *VfWRKY* genes (*VfWRKY6*, *-7*, *-15*, *-19*, *-25*, *-31*, *-41*, *-44*, *-50*) and oil-related genes, the previously published transcript data of 88 important genes (*Appendix 5) predict*ed to function in oil biosynthesis were used to carry out the correlation analysis via SPSS 18.0. *Table 2* showcases that 28 of the pivotal oil-associated genes correlates with 1–3 of 9 *VfWRKY* genes significantly

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2101_207228

(<0.05) or extremely significant (<0.01). Additionally, the number of oil-related genes correlated with each of the 9 *VfWRKY* genes ranges from 1 to 11. *VfWRKY19* displays a significant or extremely significant correlation with the largest number of oil genes (11 genes), followed by *VfWRKY15* and *VfWRKY50* (9 genes). Both *VfWRKY41* and *-44* present correlations with 4 oil-related genes, *VfWRKY6* with *-3*, *VfWRKY31* with *-2*. *VfWRKY7* and *-25* only associate with 1 functional oil gene, respectively. It is rational to infer that these 9 tested *VfWRKY* genes may be involved in the pathway of oil formation in different degrees. Furthermore, 5 *WRKY* genes showcase positive correlation with oil-related genes, including *VfWRKY6*, *-7*, *-15*, *-25*, *-41*. *VfWRKY44* shows negative relation, and *VfWRKY19*, *-31*, *-50* exhibit positive and negative relation with oil-associated genes, suggesting that the *VfWRKY* genes play different roles in the process of oil synthesis.

Table 2. *Correlation analysis between 9 VfWRKY genes and 29 oil-related genes*

Oil-related genes	VfWRKY6	VfWRKY7	VfWRKY15	VfWRKY19	VfWRKY25	VfWRKY31	VfWRKY41	VfWRKY44	VfWRKY50
BCCP3	-0.533	0.198	-0.192	0.014	0.819	-0.046	-0.545	-0.883*	0.336
PEPC4	-0.772	0.067	0.576	0.779	0.614	-0.479	-0.513	-0.84	0.905*
FATB1-1	-0.557	0.027	0.777	0.920*	0.229	-0.573	-0.258	-0.443	0.826
FATB2	-0.496	-0.057	0.819	0.913*	0.064	-0.598	-0.201	-0.283	0.764
LACS1	-0.497	-0.457	0.995**	0.906*	-0.299	-0.603	-0.166	0.011	0.76
LACS2-1	-0.565	0.413	0.204	0.518	0.807	-0.316	-0.428	-0.902*	0.632
LACS2-2	-0.555	0.414	0.216	0.529	0.788	-0.325	-0.415	-0.883*	0.629
LACS6	0.554	0.942*	-0.515	-0.249	0.543	0.611	0.498	-0.213	-0.208
LACS7	0.937*	0.504	-0.19	-0.229	-0.205	0.85	0.994**	0.544	-0.284
GPAT6	-0.574	-0.637	0.549	0.446	-0.446	-0.909*	-0.537	0.124	0.2
GPAT7	0.941*	0.683	-0.491	-0.451	0.064	0.975**	0.907*	0.335	-0.408
GPAT8	-0.498	-0.522	0.962**	0.839	-0.292	-0.511	-0.168	0.005	0.762
GPAT9-2	-0.359	0.228	0.788	0.948*	0.336	-0.248	0.021	-0.456	0.928*
LPAT1-1	-0.299	-0.533	0.919*	0.74	-0.458	-0.378	0.016	0.223	0.627
LPAT4	0.959**	0.502	-0.262	-0.305	-0.208	0.871	0.984**	0.559	-0.354
PP1	0.768	0.25	0.077	-0.039	-0.331	0.721	0.910*	0.57	-0.081
PP2-4	-0.56	-0.136	0.911*	0.979**	0.062	-0.593	-0.222	-0.311	0.87
PP3-2	0.775	0.133	-0.746	-0.857	-0.404	0.51	0.467	0.674	-0.959**
DAG-CPT	-0.582	0.124	0.734	0.918*	0.39	-0.495	-0.27	-0.585	0.891*
FAD3-1	-0.471	-0.193	0.934*	0.935*	0.043	-0.349	-0.081	-0.263	0.932*
FAD3-2	-0.358	-0.533	0.943*	0.781	-0.421	-0.43	-0.037	0.167	0.669
DGAT1	0.489	-0.292	-0.629	-0.845	-0.607	0.193	0.146	0.726	-0.957*
DGAT3-1	-0.543	-0.143	0.912*	0.974**	0.035	-0.595	-0.208	-0.282	0.853
DGAT3-2	-0.438	-0.629	0.899*	0.715	-0.419	-0.473	-0.148	0.137	0.637
PDAT1-2	-0.495	0.502	0.106	0.439	0.881*	-0.197	-0.373	-0.932*	0.591
PDAT2-1	-0.671	-0.265	0.898*	0.923*	0.145	-0.503	-0.318	-0.424	0.955*
OLE3	0.721	0.098	-0.806	-0.919*	-0.361	0.502	0.386	0.62	-0.983**
OLE5	0.51	-0.03	-0.874	-0.987**	-0.257	0.396	0.128	0.453	-0.971**

*Represents a significant correlation at the 0.05 level, and ** represents a significant correlation at the 0.05 level

Discussion

As one of the ten largest families of transcription factors across the plant lineage (Rushton et al., 2010), WRKY proteins have been extensively analyzed in so many plant species at the whole-genome level, such as grape (*Vitis vinifera* L.) (Guo et al., 2014), polar (*Populus trichocarpa*) (He et al., 2012) and white pear (*Pyrus bretschneideri*) (Huang et al., 2015), whose whole genome information has been released. In this study, 54 *WRKY* genes were identified and allocated a name in tung tree

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2101_207228

genome, ranging from *VfWRKY1* to *VfWRKY54*, as determined by their location on chromosomes.

All identified 54 VfWRKY proteins were subjected to the sequence analysis of the highly conserved WRKY domain in our study. The result showed that group I contains 21 VfWRKY domains from 11 VfWRKY proteins, each of which separately has one WRKY domain at the N- and C-terminal except VfWRKY10. VfWRKY10 displays high similarity to VfWRKY34N and VfWRKY38N belonged to group I N, indicating a common origin of their WKRY domains. Remarkably, the WRKY domains located in the C- and N-terminal are clustered into group I N and group I C, respectively, suggesting that they may be evolved in parallel. Four VfWRKY members (VfWRKY2, VfWRKY10, VfWRKY35, VfWRKY40) of group I and II c contain a variant heptapeptide amino acid sequence, WRKYGKK or WRKYGGR rather than the typical WRKYGQK sequence, at the N-terminus of WRKY domain (*Fig. 1*). Previously, the conserved WRKYGQK heptad sequence was reported to make a notable contribution to the capacity of binding the W-box (TTGACC/T) DNA element (Xie et al., 2006; Xu et al., 2006). Yamasaki et al. (2005, 2012) proposed that the invariant WRKYGQK residues of AtWRKY4 form a β-sheet via which directly contacts the DNA bases in the DNA major groove as determined by the NMR solution structural analysis and the W-box binding activity of WRKY9 from tobacco is dramatically decreased resulting from the alteration of each residue in the WRKYGQK sequence by alanine (Maeo et al., 2001). Van Verk et al. (2008) reported that although NtWRKY12 with an atypical type of sequence, WRKYGKK, fails to bind the W-box, it is able to recognize the WK-box (TTTTCCAC) DNA element. Hence, these four VfWRKY proteins, VfWRKY2, VfWRKY10, VfWRKY35 and VfWRKY40, all of which contain differentiations in typical WRKYGQK heptapeptide sequence, are worth further examining their functions in combining W-box and they are likely to reveal binding favor to some other *cis*-acting elements in place of W-box in the promoters of target genes. Apart from the cognate WRKYGQK sequence, most WRKY proteins also have a zinc ion chelating finger structure, in which all of the conserved Cys and His residues are substituted by Ala leads to the abolishment of the DNA binding ability (Maeo et al., 2001), at the C-terminus of the WRKY domain. However, the zinc-finger like motif is not discovered in the VfWRKY16 and VfWRKY 29, both of which belong to group II d (*Fig. 1*). The phenomenon of zinc-finger motif loss also occurs in other species, such as pineapple and polar (He et al., 2012; Xie et al., 2018). But this may affect the recruitment of the WRKY proteins to the promoter regions of their target genes.

It is noticeable that the hallmark of the WRKY proteins contained in group I is the existence of two WRKY domains. However, VfWRKY10, a member of group I, seemingly evolves through loss of the C-terminal WRKY domain, which is consistent with the finding that the evolution of the WRKY family is along with domain loss in plant kingdom (Brand et al., 2013). Generally, both WRKY domains in group I share a common type of zinc finger motif (C-X4-5-C-X22-23-H-X-H) but harbor different affinity to DNA. Corresponding to the existing reports, the tung tree WRKY unrooted tree reveals that the N-terminal and C-terminal WRKY domains fall under two separate subtrees (*Fig. 2*). Intriguingly, group II a, II b and II c in tung tree are more tightly cluster with group I C than to group III or other subgroups, appearing that these two subgroup proteins are derived from group I C via the process of WRKY domain loss (*Fig. 2*). Meanwhile, *Figure 2* also implies that clade II d, II e and clade III may split from a common ancestor, consisted with the primary study (Zhang and Wang, 2005).

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2101_207228

Based on our annotation analysis of gene structure, we observed that, generally, an R-type or V-type intron dwells in the WRKY domain (*Fig. 3a*). *Figure 3b* shows that, to a great extent, the members clustered into the same class share similar motif patterns, which may suggest that these proteins tend to have similarity in function or be involved in the same biological pathway. In particular, the paralog pairs, including VfWRKY24 and 26, VfWRKY 8 and 47, do not only share common motif composition but also highly conserved motif arrangement. The conserved introns located in corresponding consensus positions are commonly considered to arise in ancient genes (De Roos, 2007). Therefore, our data implies that the WRKY family in tung tree originates from ancient genes containing an intron in their WRKY domains through divergence process rather than convergence, during which group members from different origins tend to be more and more alike with time going on. Strikingly, all the WRKY domains belonged to group I N are intronless and this could root in experiencing intron loss events, such as intron turnover (Lynch, 2002). We also found the intron located in the WRKY domain of VfWRKY29 and -40 is spliced exactly at the codon of serine other than arginine, and this phenomenon may result from gene mutations at the splicing sites during the evolution.

The quantity of *WRKY* genes in tung tree is considerably smaller than that in other plants, including poplar (104 *WRKY* genes) (He et al., 2012), *Setaria italica* (105 *WRKY* genes) (Muthamilarasan et al., 2015) and so on. The course of gene family distribution and expansion correlates with gene duplication comprising tandem and segmental duplication events. The homologous genes distributed on the same chromosome are assigned as tandem duplicated genes, whereas those located in distinct chromosomes are designated as segmental duplicated genes (Wang et al., 2010). During the evolutionary process, both tandem duplication and segmental duplication make a contribution to striking diversification in the size and allocation of gene family. Gene tandem duplication, which comes into being adjacent gene clusters, is defined in the light of sequence similarity and positional proximity (Cannon et al., 2004). Upon the established standards, *VfWRKY36* and *VfWRKY37* underwent local tandem event on chromosome 8. Noticeably, five genes (*VfWRKY22*, *-23*, *-24*, *-25* and *-26*) form a relatively high-density zone spanning 3.6 Mb on chromosome 6, suggesting that chromosome 6 comprises a *WRKY* gene hot spot which exists in other species genome as well (Xie et al., 2005). Moreover, *VfWRKY4* and *VfWRKY49*, which are separately located on chromosome 1 and 11, are segmental duplicated genes according to our lab's previously published data of collinearity analysis in tung tree genome (Zhang et al., 2019) (*Fig. 4*). Compared to the large number of gene duplication events in other species (Wu et al., 2005; Xie et al., 2018), the WRKY family in tung tree only contains one pair of tandem and segmental duplicated genes (*VfWRKY36/VfWRKY37*, *VfWRKY4/49*), respectively (*Fig. 4*), indicating that the presence of low-frequency gene duplications in *VfWRKY* family. In addition, gene duplication is involved in generating novel *WRKY* genes in group III, which makes an underlying contribution to the multiplicity of *WRKY* family (Wang et al., 2015). In this analysis, differing from other plants, group III *WRKY* genes in tung tree do not experience gene duplication event and the number of it is comparatively smaller, concordant with the smaller *WRKY* family size in tung tree.

According to the expression profiles of *VfWRKY* genes in different tissues/organs and different developmental stages (*Fig. 5*), Almost all of the 54 *VfWRKY* genes are expressed in at least one tissue/organ, but RNA transcripts of *VfWRKY12* and *VfWRKY47* are only slightly detected, indicating that both of them may be only driven to express in some specific tissues at some particular stages. The investigation of *WRKY*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2101_207228

gene expression patterns at distinct developmental stages in tung tree provides some clues of value for further research in the functions of *VfWRKY* genes in varieties of biological processes. For instance, the expression level of *VfWRKY54* is highest in mature seeds, plus its WRKY domain shares high similarity with *AtWRKY23* which is able to mediate Arabidopsis embryo development (Grunewald et al., 2013), suggesting that *VfWRKY54* is likely to play similar roles in tung tree. *VfWRKY27* is predominantly expressed in stem tissue compared to the other tested tung tree samples, in accord with its ortholog in Arabidopsis, *AtWRKY12*, which is a stem-expressed gene as well. As a result, we speculate that *VfWRKY27* and *AtWRKY12* may share the common functions in mediating pith secondary wall formation and stem biomass accumulation negatively (Wang et al., 2010). *VfWRKY32* shows an expression peak in leaves tissue among the detected tissues, and possesses homologous correlation with *AtWRKY54* and *AtWRKY70*, both of which are identified as negative regulatory factors of leaf senescence in Arabidopsis, suggesting that *VfWRKY32* may be also implicated in repressing the process of leaf senescence (Besseau et al., 2012; Ulker et al., 2007).

As a major oil plant in industry, its main product, tung oil, has raised global concern recently resulting from its advantages in terms of human health, environmental protection (Itakura et al., 2003) apart from outstanding performance as paints, varnishes, etc. Identifying pivotal oil-associated genes, verifying their functions and elucidating the mechanism of tung oil biosynthesis have been the focus of many scientific researchers, with the final goal of boosting quality of tung oil products and efficiency of breeding. The expression relations between 28 of the pivotal oil-associated genes and 9 *VfWRKY* genes was analyzed, of these expression-associated oil synthesis-related genes, all come from *BCCP* (subunits of *ACCase*), *PEPC*, *FATB*, *LACS*, *GPAT*, *LPAT*, *PP*, *DAG-CPT*, *FAD*, *DGAT*, *PDAT*, *OLE* gene families, which may be regulated by the 9 *VfWRKY* genes during the process of fatty acid formation and oil accumulation. In this study, our correlation analysis of expression between chief oil formation-related genes between *VfWRKY* genes offers a hint for further testing the capability of candidate *WRKY* genes in modulating the process of oil formation. Among the analyzed 9 *VfWRKY* genes, we realize that *VfWRKY19* displays a significant or extremely significant correlation with the largest number of oil genes (11 genes), followed by *VfWRKY15* and *VfWRKY50* (9 genes), suggesting that these three *VfWRKY* genes may make a more important contribution to the pathway of oil biosynthesis. Furthermore, *VfWRKY19*, *-31*, *-50* have both positive and negative correlation with oil-related genes, indicating that they may be versatile and bifunctional during different stages of oil formation (*Table 2*).

Together, our present discoveries afford basic information of understanding *WRKY* gene family at the whole genome level in tung tree. The global analyses are also beneficial for picking some potential candidate genes for biological function investigation laying a foundation for further genetic improvement and new cultivars breeding.

Acknowledgements. This work was supported by Science Fund for Distinguished Young Scholars of Central South University of Forestry and Technology (Grant no: 2018YC002), Youth Program of National Natural Science Foundation of China (Grant no: 31901349), and Scientific Research Starting Foundation of Central South University of Forestry and Technology (Grant no: 2018YJ001).

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2101_207228

REFERENCES

[1] Amato, A., Cavallini, E., Zenoni, S., Finezzo, L., Begheldo, M., Ruperti, B., Tornielli, G. B. (2017): A grapevine TTG2-Like WRKY transcription factor is involved in regulating vacuolar transport and flavonoid biosynthesis. – Front Plant Sci 7: 1979.
[2] Bailey, T. L., Elkan, C. (1994): Fitting a mixture model by expectation maximization to discover motifs in biopolymers. – Proc Int Conf Intell Syst Mol Biol 2: 28-36.
[3] Bakshi, M., Oelmuller, R. (2014): WRKY transcription factors: jack of many trades in plants. – Plant Signal Behav 9(2): e27700.
[4] Banerjee, A., Roychoudhury, A. (2015): WRKY proteins: signaling and regulation of expression during abiotic stress responses. – Scientific World Journal 2015: 807560.
[5] Besseau, S., Li, J., Palva, E. T. (2012): WRKY54 and WRKY70 co-operate as negative regulators of leaf senescence in Arabidopsis thaliana. – J Exp Bot 63(7): 2667-2679.
[6] Birkenbihl, R. P., Liu., S., Somssich, I. E. (2017): Transcriptional events defining plant immune responses. – Curr Opin Plant Biol 38: 1-9.
[7] Brand, L. H., Fischer, N. M., Harter, K., Kohlbacher, O., Wanke, D. (2013): Elucidating the evolutionary conserved DNA-binding specificities of WRKY transcription factors by molecular dynamics and in vitro binding assays. – Nucleic Acids Res 41(21): 9764-9778.
[8] Cannon, S. B., Mitra, A., Baumgarten, A., Young, N. D., May, G. (2004): The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana. – BMC Plant Biol 4: 10.
[9] Cao, H., Chapital, D. C., Howard, O. D., Jr, Deterding, L. J., Mason, C. B., Shockey, J. M., Klasson, K. T. (2012): Expression and purification of recombinant tung tree diacylglycerol acyltransferase 2. – Appl Microbiol Biotechnol 96(3): 711-727.
[10] Cao, H., Shockey, J. M., Klasson, K. T., Chapital, D. C., Mason, C. B., Scheffler, B. E. (2013): Developmental regulation of diacylglycerol acyltransferase family gene expression in tung tree tissues. – PLoS One 8(10): e76946.
[11] Chen, K., Rajewsky, N. (2007): The evolution of gene regulation by transcription factors and microRNAs. – Nat Rev Genet 8(2): 93-103.
[12] Cheng, X., Xiong, R., Liu, H., Wu, M., Chen, F., Hanwei Yan, Xiang, Y. (2018): Basic helix-loop-helix gene family: genome wide identification, phylogeny, and expression in Moso bamboo. – Plant Physiol Biochem 132: 104-119.
[13] Chujo, T., Miyamoto, K., Shimogawa, T., Shimizu, T., Otake, Y., Yokotani, N., Nishizawa, Y., Shibuya, N., Nojiri, H., Yamane, H., Minami, E., Okada, K. (2013): OsWRKY28, a PAMP-responsive transrepressor, negatively regulates innate immune responses in rice against rice blast fungus. – Plant Mol Biol 82(1-2): 23-37.
[14] de Roos, A. D. (2007): Conserved intron positions in ancient protein modules. – Biol Direct 2: 7.
[15] Deng, W., Wang, Y., Liu, Z., Cheng, H., Xue, Y. (2014): HemI: a toolkit for illustrating heatmaps. – PLoS One 9(11): e111988.
[16] Devaiah, B. N., Karthikeyan, A. S., Raghothama, K. G. (2007): WRKY75 transcription factor is a modulator of phosphate acquisition and root development in Arabidopsis. – Plant Physiol 143(4): 1789-1801.
[17] Dou, L., Zhang, X., Pang, C., Song, M., Wei, H., Fan, S., Yu, S. (2014): Genome-wide analysis of the WRKY gene family in cotton. – Mol Genet Genomics 289(6): 1103-1121.
[18] Dyer, J. M., Chapital, D. C., Kuan, J. C., Mullen, R. T., Turner, C., McKeon, T. A., Pepperman, A. B. (2002): Molecular analysis of a bifunctional fatty acid conjugase/desaturase from tung. Implications for the evolution of plant fatty acid diversity. – Plant Physiol 130(4): 2027-2038.
[19] Eulgem, T., Rushton, P. J., Robatzek, S., Somssich, I. E. (2000): The WRKY superfamily of plant transcription factors. – Trends Plant Sci 5(5): 199-206.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2101_207228

[20] Fatihi, A., Zbierzak, A. M., Dormann, P. (2013): Alterations in seed development gene expression affect size and oil content of Arabidopsis seeds. – Plant Physiol 163(2): 973-985.
[21] Finn, R. D., Clements, J., Arndt, W., Miller, B. L., Wheeler, T. J., Schreiber, F., Bateman, A., Eddy, S. R. (2015): HMMER web server: 2015 update. – Nucleic Acids Res 43(W1): W30-38.
[22] Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., Bateman, A. (2016): The Pfam protein families database: towards a more sustainable future. – Nucleic Acids Res 44(D1): D279-285.
[23] Geilen, K., Heilmann, M., Hillmer, S., Böhmer, M. (2017): WRKY43 regulates polyunsaturated fatty acid content and seed germination under unfavourable growth conditions. – Sci Rep 7(1): 14235.
[24] Grunewald, W., De Smet, I., De Rybel, B., Robert, H. S., van de Cotte, B., Willemsen, V., Gheysen, G., Weijers, D., Friml, J., Beeckman, T. (2013): Tightly controlled WRKY23 expression mediates Arabidopsis embryo development. – EMBO Rep 14(12): 1136-1142.
[25] Gu, L., Li, L., Wei, H., Wang, H., Su, J., Guo, Y., Yu, S. (2018): Identification of the group IIa WRKY subfamily and the functional analysis of GhWRKY17 in upland cotton (Gossypium hirsutum L.). – PLoS One 13(1): e0191681.
[26] Guo, C., Guo, R., Xu, X., Gao, M., Li, X., Song, J., Zheng, Y., Wang, X. (2014): Evolution and expression analysis of the grape (Vitis vinifera L.) WRKY gene family. – J Exp Bot 65(6): 1513-1528.
[27] 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 Rep 31(7): 1199-1217.
[28] He, Z., Zhang, H., Gao, S., Lercher, M. J., Chen, W. H., Hu, S. (2016): Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. – Nucleic Acids Res 44(W1): W236-241.
[29] Hu, Y., Chen, L., Wang, H., Zhang, L., Wang, F., Yu, D. (2013): Arabidopsis transcription factor WRKY8 functions antagonistically with its interacting partner VQ9 to modulate salinity stress tolerance. – Plant J 74(5): 730-745.
[30] Huang, X., Li, K., Xu, X., Yao, Z., Jin, C., Zhang, S. (2015): Genome-wide analysis of WRKY transcription factors in white pear (Pyrus bretschneideri) reveals evolution and patterns under drought stress. – BMC Genomics 16: 1104.
[31] Ishiguro, S., Nakamura, K. (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 beta-amylase from sweet potato. – Mol Gen Genet 244(6): 563-571.
[32] Itakura, T., Airey, D. W., Leo, C. J. (2003): The diffusion and sorption of volatile organic compounds through kaolinitic clayey soils. – J Contam Hydrol 65(3-4): 219-243.
[33] Jiang, Y., Liang, G., Yang, S., Yu, D. (2014): Arabidopsis WRKY57 functions as a node of convergence for jasmonic acid- and auxin-mediated signaling in jasmonic acid-induced leaf senescence. – Plant Cell 26(1): 230-245.
[34] Jin, J., Zhang, H., Kong, L., Gao, G., Luo, J. (2014): PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors – Nucleic Acids Res 42(Database issue): D1182-1187.
[35] Johnson, C. S., Kolevski, B., Smyth, D. R. (2002): TRANSPARENT TESTA GLABRA2, a trichome and seed coat development gene of Arabidopsis, encodes a WRKY transcription factor. – Plant Cell 14(6): 1359-1375.
[36] Kloth, K. J., Wiegers, G. L., Busscher-Lange, J., van Haarst, J. C., Kruijer, W., Bouwmeester, H. J., Dicke, M., Jongsma, M. A. (2016): AtWRKY22 promotes

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2101_207228

susceptibility to aphids and modulates salicylic acid and jasmonic acid signalling. – J Exp Bot 67(11): 3383-3396.

[37] Levée, V., Major, I., Levasseur, C., Tremblay, L., MacKay, J., Séguin, A. (2009): Expression profiling and functional analysis of Populus WRKY23 reveals a regulatory role in defense. – New Phytol 184(1): 48-70.

[38] Li, L., Stoeckert, C. J., Jr., Roos, D. S. (2003): OrthoMCL: identification of ortholog groups for eukaryotic genomes. – Genome Res 13(9): 2178-2189.

[39] Li, H. L., Zhang, L. B., Guo, D., Li, C. Z., Peng, S. Q. (2012): Identification and expression profiles of the WRKY transcription factor family in Ricinus communis. – Gene 503(2): 248-253.

[40] Li, W., Wang, H., Yu, D. (2016): Arabidopsis WRKY transcription factors WRKY12 and WRKY13 oppositely regulate flowering under short-day conditions. – Mol Plant 9(11): 1492-1503.

[41] Li, Z., Long, H., Zhang, L., Liu, Z., Cao, H., Shi, M., Tan, X. (2017): The complete chloroplast genome sequence of tung tree (Vernicia fordii): organization and phylogenetic relationships with other angiosperms. – Sci Rep 7(1): p. 1869.

[42] Liu, H., Yang, W., Liu, D., Han, Y., Zhang, A., Li, S. (2011): Ectopic expression of a grapevine transcription factor VvWRKY11 contributes to osmotic stress tolerance in Arabidopsis. – Mol Biol Rep 38(1): 417-427.

[43] Luo, M., Dennis, E. S., Berger, F., Peacock, W. J., Chaudhury, A. (2005): a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in Arabidopsis. – Proc Natl Acad Sci U S A 102(48): 17531-17536.

[44] Lynch, M. (2002): Intron evolution as a population-genetic process. – Proc Natl Acad Sci USA 99(9): 6118-6123.

[45] Maeo, K., Hayashi, S., Kojima-Suzuki, H., Morikami, A., Nakamura, K. (2001): Role of conserved residues of the WRKY domain in the DNA-binding of tobacco WRKY family proteins. – Biosci Biotechnol Biochem 65(11): 2428-2436.

[46] Mangelsen, E., Kilian, J., Berendzen, K. W., Kolukisaoglu, U. H., Harter, K., Jansson, C., Wanke, D. (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: 194.

[47] Mirabella, R., Rauwerda, H., Allmann, S., Scala, A., Spyropoulou, E. A., de Vries, M., Boersma, M. R., Breit, T. M., Haring, M. A., Schuurink, R. C. (2015): WRKY40 and WRKY6 act downstream of the green leaf volatile E-2-hexenal in Arabidopsis. – Plant J 83(6): 1082-1096.

[48] Muthamilarasan, M., Bonthala, V. S., Khandelwal, R., Jaishankar, J., Shweta, S., Nawaz, K., Prasad, M. (2015): Global analysis of WRKY transcription factor superfamily in Setaria identifies potential candidates involved in abiotic stress signaling. – Front Plant Sci 6: 910.

[49] Park, J. Y., Kim, D. K., Wang, Z. M., Lu, P., Park, S. C., Lee, J. S. (2008): Production and characterization of biodiesel from tung oil. – Appl Biochem Biotechnol 148(1-3): 109-117.

[50] Phukan, U. J., Jeena, G. S., Shukla, R. K. (2016): WRKY transcription factors: molecular regulation and stress responses in plants. – Front Plant Sci 7: 760.

[51] Rinerson, C. I., Rabara, R. C., Tripathi, P., Shen, Q. J., Rushton, P. J. (2015): The evolution of WRKY transcription factors. – BMC Plant Biol 15: 66.

[52] Rushton, P. J., Macdonald, H., Huttly, A. K., Lazarus, C. M., Hooley, R. (1995): Members of a new family of DNA-binding proteins bind to a conserved cis-element in the promoters of alpha-Amy2 genes. – Plant Mol Biol 29(4): 691-702.

[53] Rushton, P. J., Somssich, I. E., Ringler, P., Shen, Q. J. (2010): WRKY transcription factors. – Trends Plant Sci 15(5): 247-258.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2101_207228

[54] Rushton, D. L., Tripathi, P., Rabara, R. C., Lin, J., Ringler, P., Boken, A. K., Langum, T. J., Smidt, L., Boomsma, D. D., Emme, N. J., Chen, X., Finer, J. J., Shen, Q. J., Rushton, P. J. (2012): WRKY transcription factors: key components in abscisic acid signalling. – Plant Biotechnol J 10(1): 2-11.

[55] Shockey, J. M., Gidda, S. K., Chapital, D. C., Kuan, J. C., Dhanoa, P. K., Bland, J. M., Rothstein, S. J., Mullen, R. T., Dyer, J. M. (2006): Tung tree DGAT1 and DGAT2 have nonredundant functions in triacylglycerol biosynthesis and are localized to different subdomains of the endoplasmic reticulum. – Plant Cell 18(9): 2294-2313.

[56] Tripathi, P., Rabara, R. C., Langum, T. J., Boken, A. K., Rushton, D. L., Boomsma, D. D., Rinerson, C. I., Rabara, J., Reese, R. N., Chen, X., Rohila, J. S., Rushton, P. J. (2012): The WRKY transcription factor family in Brachypodium distachyon. – BMC Genomics 13: 270.

[57] Ulker, B., Somssich, I. E. (2004): WRKY transcription factors: from DNA binding towards biological function. – Curr Opin Plant Biol 7(5): 491-498.

[58] Ulker, B., Shahid Mukhtar, M., Somssich, I. E. (2007): The WRKY70 transcription factor of Arabidopsis influences both the plant senescence and defense signaling pathways. – Planta 226(1): 25-37.

[59] van Verk, M. C., Pappaioannou, D., Neeleman, L., Bol, J. F., Linthorst, H. J. (2008): A Novel WRKY transcription factor is required for induction of PR-1a gene expression by salicylic acid and bacterial elicitors. – Plant Physiol 146(4): 1983-1995.

[60] Wang, H., Avci, U., Nakashima, J., Hahn, M. G., Chen, F., Dixon, R. A. (2010a): Mutation of WRKY transcription factors initiates pith secondary wall formation and increases stem biomass in dicotyledonous plants. – Proc Natl Acad Sci U S A 107(51): 22338-22343.

[61] Wang, Y., Deng, D., Bian, Y., Lv, Y., Xie, Q. (2010b): Genome-wide analysis of primary auxin-responsive Aux/IAA gene family in maize (Zea mays. L.). – Mol Biol Rep 37(8): 3991-4001.

[62] Wang, Q., Wang, M., Zhang, X., Hao, B., Kaushik, S. K., Pan, Y. (2011): WRKY gene family evolution in Arabidopsis thaliana. – Genetica 139(8): 973-983.

[63] Wang, Y., Feng, L., Zhu, Y., Li, Y., Yan, H., Xiang, Y. (2015): Comparative genomic analysis of the WRKY III gene family in Populus, grape, Arabidopsis and rice. – Biol Direct 10: 48.

[64] Wang, Y., Liu, H., Zhu, D., Gao, Y., Yan, H., Xiang, Y. (2017): Genome-wide analysis of VQ motif-containing proteins in Moso bamboo (Phyllostachys edulis). – Planta 246(1): 165-181.

[65] Wu, K. L., Guo, Z. J., Wang, H. H., Li, J. (2005): The WRKY family of transcription factors in rice and Arabidopsis and their origins. – DNA Res 12(1): 9-26.

[66] Xiang, J., Tang, S., Zhi, H., Jia, G., Wang, H., Diao, X. (2017): Loose Panicle1 encoding a novel WRKY transcription factor, regulates panicle development, stem elongation, and seed size in foxtail millet [Setaria italica (L.) P. Beauv.]. – PLoS One 12(6): e0178730.

[67] Xie, Z., Zhang, Z. L., Zou, X., Huang, J., Ruas, P., Thompson, D., Shen, Q. J. (2005): Annotations and functional analyses of the rice WRKY gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells. – Plant Physiol 137(1): 176-189.

[68] Xie, Z., Zhang, Z. L., Zou, X., Yang, G., Komatsu, S., Shen, Q. J. (2006): Interactions of two abscisic-acid induced WRKY genes in repressing gibberellin signaling in aleurone cells. – Plant J 46(2): 231-242.

[69] Xie, T., Chen, C., Li, C., Liu, J., Liu, C., He, Y. (2018): Genome-wide investigation of WRKY gene family in pineapple: evolution and expression profiles during development and stress. – BMC Genomics 19(1): 490.

[70] Xu, X., Chen, C., Fan, B., Chen, Z. (2006): Physical and functional interactions between pathogen-induced Arabidopsis WRKY18, WRKY40, and WRKY60 transcription factors. – Plant Cell 18(5): 1310-1326.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2101_207228

[71] Yamasaki, K., Kigawa, T., Inoue, M., Tateno, M., Yamasaki, T., Yabuki, T., Aoki, M., Seki, E., Matsuda, T., Tomo, Y., Hayami, N., Terada, T., Shirouzu, M., Tanaka, A., Seki, M., Shinozaki, K., Yokoyama, S. (2005): Solution structure of an Arabidopsis WRKY DNA binding domain. – Plant Cell 17(3): 944-956.
[72] Yamasaki, K., Kigawa, T., Watanabe, S., Inoue, M., Yamasaki, T., Seki, M., Shinozaki, K., Yokoyama, S. (2012): Structural basis for sequence-specific DNA recognition by an Arabidopsis WRKY transcription factor. – J Biol Chem 287(10): 7683-7691.
[73] Zhang, Y., Wang, L. (2005): The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants. – BMC Evol Biol 5: 1.
[74] 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.
[75] Zhang, L., Jia, B., Tan, X., Thammina, C. S., Long, H., Liu, M., Wen, S., Song, X., Cao, H. (2014): Fatty acid profile and unigene-derived simple sequence repeat markers in tung tree (Vernicia fordii). – PLoS One 9(8): e105298.
[76] Zhang, L., Liu, M., Long, H., Dong, W., Pasha, A., Esteban, E., Li, W., Yang, X., Li, Z., Song, A., Ran, D., Zhao, G., Zeng, Y., Chen, H., Zou, M., Li, J., Liang, F., Xie, M., Hu, J., Wang, D., Cao, H., Provart, N. J., Zhang, L., Tan, X. (2019): Tung tree (Vernicia fordii) genome provides a resource for understanding genome evolution and improved oil production. – Genomics Proteomics & Bioinformatics 17(6).

APPENDIX

Appendix 1. *Used softwares and online tools*

Chart	Software and online tools	Database and website
Figure 1. Multiple sequence alignment of VfWRKY and AtWRKY proteins	Pfam, NCBI-CDD, HMMER3.0, SMART, TAIR10, MEGA X(ClustalW), GeneDoc	WRKY (PF03106) in Pfam(http://pfam.xfam.org/), NCBI-CDD (https://www.ncbi.nlm.nih.gov/Structure/cdd/docs/cdd_search.html), HMMER3.0 (http://www.hmmer.org/), SMART (http://smart.embl-heidelberg.de/), TAIR10 (http://www.arabidopsis.org/), GeneDoc (https://www.softpedia.com/get/Science-CAD/GeneDoc.shtml)
Figure 2. Phylogenic analysis of the WRKY domains of 54 VfWRKY and 71 AtWRKY proteins	MEGA 7.0(neighbor-joining (NJ)), EvolView	MEGA 7.0 (https://www.megasoftware.net/), EvolView (https://www.evolgenius.info/evolview/#login)
Figure 3. Structural characterization of WRKY family in the tung tree	GSDS2.0, TBTOOLS, MEME	GSDS2.0 (http://gsds.cbi.pku.edu.cn/index.php), TBTOOLS (http://www.tbtools.com/), MEME (version 5.0.5, http://meme-suite.org/tools/meme)
Figure 4. Schematic representations for the physical location and region duplication of the 51 WRKY genes on 11 chromosomes in tung tree	MG2C, Adobe Illustrator	MG2C (http://mg2c.iask.in), Adobe Illustrator (https://www.adobe.com/cn/products/illustrator/free-trial-download.html)
Figure 5. Expression patterns of tung tree WRKY genes in distinct tissues at different developmental stages	HemI (version 1.0.3.7)	$\log_2$(FPKM + 1)-transformed data
Table 1. The basic information of WRKY transcription factor family in V. fordii	WoLF PSORT, ExPASy	WoLF PSORT (https://wolfpsort.hgc.jp/), ExPASy (http://www.expasy.ch/tools/pi_tool.html)
Table 2. Correlation analysis between 9 VfWRKY genes and 29 oil-related genes	SPSS (version 24)	

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2101_207228

ELECTRONIC APPENDICES

Appendix 1. *The basic information of WRKY transcription factor family in V. fordii*

Appendix 2. *Schematic representation of the intron position in the WRKY domain*

Appendix 3. *Detailed sequence information of the 20 motifs in the 54 VfWRKY proteins*

Appendix 4. *The RNA transcript levels of 54 VfWRKY genes were analyzed in the present study*

Appendix 5. *Expression level of 88 important oil-related genes (FPKM value)*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2101_207228