

Genome-Wide Identification and Expression Profile Analysis of WRKY Family Genes in Teak (*Tectona Grandis*)

Xi-Yang Wang

Guangdong eco-Engineering Polytechnic

Jie Song (✉ jsong@caf.ac.cn)

Research Institute of Tropical Forestry Chinese Academy of Forestry <https://orcid.org/0000-0002-7438-4882>

Jia-Hui Xing

Guangdong Academy of Sciences

Jun-Feng Liang

Chinese Academy of Forestry

Bi-ying Ke

Guangdong eco-Engineering Polytechnic

Research

Keywords: Expression profiling, Teak, WRKY, Evolution, Abiotic stress

Posted Date: November 19th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-108100/v1>

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

[Read Full License](#)

Abstract

Background: WRKY proteins comprise a large family of transcription factors that play vital roles in many aspects of physiological processes and adaptation to environment. However, little information was available about the WRKY genes in teak (*Tectona grandis*). The recent release of the whole-genome sequence of teak allowed us to perform a genome-wide investigation into the organization and expression profiling of teak WRKY genes.

Results: In the present study, 102 teak WRKY (TgWRKY) genes were identified and renamed as per their positions on chromosome and scaffolds. According to their structural and phylogenetic analysis, the 102 TgWRKYs were further classified into three main groups with several subgroups. The segmental duplication event played a major role in the expansion of teak WRKY gene family and three WGD events were inferred. Expression profiles derived from transcriptome data exhibited distinct expression patterns of TgWRKY genes in various tissues and in response to different abiotic stress.

Conclusions: 102 TgWRKY genes were identified in teak and the structure of their encoded proteins, their evolutionary characteristics and expression patterns were examined in this study. This study generated an important resource that will provide helpful information for further exploration of the TgWRKY genes role in the regulatory mechanism in response to abiotic stresses.

Background

Transcription factors (TF) are a class of proteins that regulate gene expression in all organisms, and bind to specific DNA sequences in the promoter regions of genes to activate or repress the transcription of multiple target genes (Ross et al., 2007). WRKY is a transcription factor first discovered in plants, which has a highly conserved domain of 60 amino acids, with one or two conserved amino acid sequences composed of WRKYGQK at the N-terminus and a zinc finger structure at the C-terminus (Agarwal et al., 2011). The WRKY transcriptional regulator specifically binds to the DNA sequence of the promoter region of the target gene [(T)TGAC(C/T)], regulates the expression of the target gene, and plays an important role in pathogen defense, abiotic stress, and the development of the target gene (Chen et al., 2010). Based on the number of WRKY domains and the zinc finger sequences, WRKY transcription factors are classified into 3 groups: group I members have two WRKY domains and their zinc finger structures are C2H2(C-X₄₋₅-C-X₂₂₋₂₃-H-X-H); group II members have a single WRKY domain and their zinc finger structures are C2H2(C-X₄₋₅-C-X₂₂₋₂₃-H-X-H); and group III members have a single WRKY domain and their zinc finger structures are C2H2(C-X₄₋₅-C-X₂₂₋₂₃-H-X-H). The type IIa, IIb, IIc, II d, and II e are divided into five subfamilies (C-X₂₂₋₂₃-H-X-H); the group III members have a single WRKY domain and their zinc finger structure is C2HC (C-X₇-C-X₂₃-H-X-C) (Chen et al., 2017; Kalde et al., 2003; Wei et al., 2012).

The WRKY transcription factor is involved in various biotic and abiotic stress processes in plants, especially during external drought stress, it can act as a positive or negative regulator to regulate the expression of related genes (Dong et al., 2003). Ishiguro and Nakamura first discovered the WRKY

transcription factor in sweet potato and named it SPF1 (sweet potato factor 1), and subsequently identified a large number of WRKY transcription factors from various plants (Ishiguro & Nakamura, 1994). For example, 74 WRKY genes were identified in *Arabidopsis thaliana* in which overexpression of TaWRKY146, a homolog of AtWRKY46, resulted in hypersensitivity to drought and salt stress in *Arabidopsis* (Birkenbihl et al., 2018). 109 WRKY genes were identified in rice (*Oryza sativa* L.). Overexpression of OsWRKY47 improves yield and drought tolerance (Ma et al., 2017), while OsWRKY13 can bind to WBOX1 (TTGACT) and WBOXa (TTGAC) sequences in the promoter regions of SNAC1 and WRKY45-1, inhibiting the expression of SNAC1 and WRKY45-1 and negatively regulating the drought tolerance of rice (Birkenbihl et al., 2018). In soybean, GmWRKY6, GmWRKY46, GmWRKY56, GmWRKY106, and GmWRKY149 genes were differentially expressed between drought tolerance and susceptible soybean genotypes under drought stress conditions (Dias et al., 2016). In addition, related genes involved in drought stress were also identified in sesame (Li et al., 2017), cucumber, maize, tomato and cotton (Dou et al., 2014; Ling et al., 2011; Wang et al., 2014; Wei et al., 2016).

Natural teak forests has been found in India, Myanmar, Laos, and northern Thailand, with an area of 29 million ha, about 46.4% of which in Myanmar, 30.1% in Thailand, and 23.4% in India (Pandey and Brown, 2000). It is known as the national tree in Myanmar, and has been introduced to Indonesia, Sri Lanka, Vietnam, Malaysia and other Southeast Asian countries, as well as to more than 60 countries and regions in Africa and Latin America (Pérez and Kanninen, 2005). Because of its fast-growing characteristics and excellent timber properties, the area of teak plantation is increasing to 6.8 million ha. Teak is an excellent greening tree for sidewalks and gardens. It is an important timber for military and marine purposes, such as warships and ocean liners, and is used for docks, bridges, buildings, carriages, high-grade furniture, wooden floors, carvings, carpentry, veneer and paneling, etc. It is one of the most valuable timber in the world (Kollert and Kleine, 2017).

In recent years, the price of natural teak logs in Myanmar has increased by 3.0%~4.5% annually, especially in the Indian market, from 615 dollars m^{-3} in 2005 to 1000 dollars m^{-3} in 2014 (Kollert and Kleine, 2017). Continuous increasing international demand has led to the expansion of teak timber supply from natural forest logs in Asia to fast-growing, small-diameter timber plantations in Africa and Latin America. With high quality plantation timber currently costing 600 ~ 1000 dollars m^{-3} for large-diameter timber and 350 ~ 500 dollars m^{-3} for small-diameter timber, teak is already one of the most expensive hardwood species in the world. The world's traditional teak exporters, Thailand, India and Indonesia, have stopped exporting natural teak logs since the 1980s (Kollert and Cherubini, 2012), and Myanmar has also implemented a ban on natural forest log exports since 2014. Resource gaps and high prices have spurred the plantation and management of teak plantations. Internationally, large areas of teak plantation forests have been developed in Brazil, Costa Rica and Ecuador in South America, and Ghana, Congo and Madagascar in Africa. Since 2000, in Guangdong, Guangxi, Fujian, Hainan, Yunnan and other provinces in China, many individuals and private enterprises have invested heavily in teak plantation for an area of 35,000 hectares. The market needs a large number of excellent seedlings with high stress resistance.

Therefore, the identification of the WRKY gene family in teak is important for the improvement of stress tolerance and genetic breeding in the future.

Results

Identification of WRKY Gene Family in Teak

Initially a total of 109 WRKY gene family members were identified in teak genome. Out of 109 genes, 102 showed the occurrence of a well-conserved WRKY domain confirmed by NCBI-Conserved Domain Database searches (see Additional file 1). The WRKY proteins with no conserved WRKY domain or zinc-finger motif were excluded from our analysis. Finally, 102 WRKY proteins which consist of a conserved WRKY domain were named as TgWRKY1 to TgWRKY102, as per their positions on *T. grandis* Chromosome numbers 1–18 and Scaffolds (see Additional file 3). The length of TgWRKY proteins ranged between 97 to 726 amino acids. The molecular mass of the predicted TgWRKY proteins ranged from 11208.73 to 78769.26 kDa and also showed a broad range of isoelectric points range from 4.73 to 9.98 pKa. The aliphatic index indicates the thermostability of proteins, which ranges from 41.12 to 75.16. The TgWRKY76 was found to be the most thermostable among the predicted WRKY proteins (see Additional file 3). The predicted subcellular localization results showed that all the TgWRKY proteins were located in the nuclear region.

Multiple Sequences Alignment And Phylogenetic Analysis

The phylogenetic relationship of the TgWRKY proteins was examined by multiple sequence alignment of their WRKY domains, which span approximately 60 amino acids (see Additional file 2). The WRKY domain of seven different *Arabidopsis* WRKY proteins (AtWRKY58, 40, 61, 50, 74, 65, 54) from each of the groups or subgroups, were randomly selected as representatives for the further comparison. From the 102 TgWRKY proteins, 96 have a highly conserved heptapeptide domain 'WRKYGQK', whereas 'WRKYGKK', 'WRKYGEK' or 'WHKYGQK' domains are present in the rest of the TgWRKY proteins. A total of 19 TgWRKY genes were classified into Group I. All WRKY proteins of Group I having two WRKY domains with (C2H2)-type zinc finger pattern, except for TgWRKY6 and TgWRKY 7. Apparently, gene loss occurred in teak, as the AtWRKY10 in *Arabidopsis thaliana*. The remaining 65 members contained a C2H2-type and 18 members contained a C2HC-type zinc finger, which were ordered as groups II and III respectively.

The phylogenetic analysis (Fig. 2) indicated that the teak WRKY domain could be divided into three large groups corresponding to group I, II and III in *Arabidopsis* as defined by Eulgem et al. (2000). The N-terminal and C-terminal WRKY domains of Group I were clustered in different clades, which may reflect the parallel evolution of the two domains. TgWRKY6 and TgWRKY7 were clustered together with I-C and I-N clades respectively, which suggested a cross relationship. The WRKY members in group II can be further clustered into five subgroups (IIa-IIe), 6 WRKY proteins belong to IIa, 16 to IIb, 24 to IIc, 8 to IId and

11 to Ile. All the members in group III contain the C2HC-type zinc fingers (C-X₄₋₇-C-X₂₃-H-X-C) and clustered into an independent clade.

Gene Structure And Motif Composition Of Teak Wrky Proteins

The distribution of exon-intron was analyzed to gain more insight into the structural features of TgWRKY genes. As shown in Fig. 3C, the number of exon varies from one to nine among TgWRKY gene categories (1 with one exon, seven with two exons, 48 with three exons, 13 with four exons, 22 with five exon and 10 with six exons and 1 with nine exons). The TgWRKY genes from the same WRKY group commonly shared a similar gene structure in terms of intron/exon organization. For example, all group Ile members contained three exons and two introns. The distribution of introns and the intron phase were coincident with the alignment clusters of TgWRKY genes. The V-type intron, a phase-0 intron, was only observed in group Ila and Ilb. R-type intron (a phase-1 intron) was widely distributed in all the other groups (group I, Ilc, Ild, Ile and III), similar with that in rice and arabidopsis (Wu et al. 2005). No introns were found in the N-terminal WRKY domains of genes belonging to group I.

For a better understanding, the diversity of TgWRKY protein motifs, the conserved motif analysis of TgWRKY were investigated by using the MEME software. As exhibited in Fig. 3B, motif 1 and 2 are widely distributed in all the TgWRKY genes, and TgWRKY members within the same groups were usually found to share a similar motif composition (Fig. S3). For example, motif 11 is unique to group I, motif 15 is unique to Ild, motif 20 is only found in 13 of the 18 members of group III, whereas motif 8 is specific to group Ila and Ilb. The clustered TgWRKY pairs, i.e. TgWRKY50/99, showed highly similar motif distribution. The similar motif arrangement among TgWRKY proteins within subgroups indicated that the protein architecture is conserved within a specific subfamily. The functions of most of these conserved motifs remain to be elucidated. Overall, the conserved motif compositions and similar gene structures of the WRKY members in the same group, together with the phylogenetic analysis results, could strongly support the reliability of the group classifications.

Chromosomal Distribution And Synteny Analysis Of Tgwrky Genes

Among the 102 identified TgWRKY genes, 92 were distributed across the eighteen pseudomolecules of teak genome (pseudomolecules 1–18), and the remaining 10 TgWRKY genes were separately mapped onto 9 unassigned scaffolds (Fig. 4). Pseudomolecules 1 contained the largest number of TgWRKY genes, with 10 out of 102 genes (9.8%), whereas pseudomolecules 4 have only one gene. The TgWRKY genes of subgroup Ilc were the most widely distributed, and found on 13 Pseudomolecules. The results revealed a non-uniform distribution of WRKY genes among the pseudomolecules and scaffolds of teak.

According to the descriptions of Holub (2001) and Gu et al. (2002), a chromosomal region shorter than 200 kb containing two or more genes with a similarity greater than 75% is defined as a tandem duplication event. Four TgWRKY genes (TgWRKY 2/3, TgWRKY 4/5) in Group III were clustered into two tandem duplication event regions on pseudomolecule 1, indicating a hot spot of WRKY gene distribution. Besides the tandem duplication events, 63 segmental duplication events with 78 WRKY genes were also identified with BLASTP and MCScanX methods (Fig. 5, Additional file 4). These results indicated that some TgWRKYs were possibly generated by gene duplication and the segmental duplication events played a major driving force for TgWRKY evolution. In addition, some TgWRKY genes were found to be associated with at least three syntenic gene pairs (TgWRKY 34, 46, 68, 72, 73, 88, 92), suggested that these genes may have played an important role in the extension of TgWRKY genes.

To better understand the evolutionary constraints acting on WRKY gene family, the Ka/Ks ratios of the duplicated WRKY gene pairs were calculated. Most segmental and tandem duplicated TgWRKY gene pairs had Ka/Ks < 1, suggesting that the teak WRKY gene family might have experienced strong purifying selective pressure during evolution (see Additional file 4). Intriguingly, the divergence times maybe suggested three whole-genome duplications (WGDs): 32 segmental duplicated gene pairs were diverged during 15.3–33.9 million years ago (Mya), 16 segmental and 2 tandem duplicated gene pairs were diverged during 50.5–69.1 Mya, and 6 segmental duplicated gene pairs were diverged during 82.5–119.3 Mya.

Expression Profiling Of Tgwrky Genes In Mature Teak Plant

In order to discover the tissue specificity in teak, the expression pattern of 102 TgWRKY genes were investigated in this study. RNA data of various tissues from wild mature individuals including roots, stems, leaves, flowers, pulps and cores were sequenced. Among the 102 TgWRKY genes, TgWRKY 19 was not expressed in all detected samples, which maybe pseudogenes or had special temporal and spatial expression patterns not examined in our libraries (see Additional file 5). Some genes exhibited preferential expression across the detected tissues. TgWRKY53/55, TgWRKY 61, TgWRKY94 and TgWRKY98 were only expressed in flower, root, stem and leaf respectively. 25 genes in root, 16 genes in stem, 37 genes in leaf, 13 genes in flower, 3 genes in pulp and 2 genes in core showed the highest transcript abundances (Fig. 6, Additional file 5).

Expression profiling of TgWRKY genes in seedlings under abiotic stress

To further confirm whether the expression of TgWRKY genes was influenced by different abiotic stresses, the transcriptional levels of all 102 genes in different tissues of seedlings were also investigated. Overall, our analysis showed that the expression pattern of TgWRKY genes were tissue-specific, complex and

dynamically changing (Fig. 6, Additional file 5), and more than half of genes were considered as stress response genes ($\log_2FC > 1$ or $\log_2FC < -1$).

Under drought stress, 14 TgWRKY genes in roots keep a rising expression level over time, 6 genes keep downregulated, 8 genes were upregulated during the first 24 h and then decreased, 62 genes were downregulated during the first 24 h and then increased, 1 gene was switched off and 3 genes were turned on at 48 h. The corresponding quantities in the stems and leaves are 12, 6, 28, 42, 4, 2 and 4, 23, 39, 24, 1, 4, respectively. The highest upregulated expression (53-fold increase) in roots was detected at 24 h in TgWRKY 32, followed by TgWRKY 90 (15-fold increase) at 48 h. The most remarkable downregulated gene is TgWRKY84 with 16-fold decrease at 48 h. In stems, TgWRKY 22 was the most upregulated gene at 24 h with 56- fold change, and TgWRKY 84 was the most down-regulated gene at 24 h with 390-fold decrease. In leaves, TgWRKY 54 was significantly induced at 24 h with 15-fold increase, and TgWRKY 49 was repressed more than 18-fold at 48 h. Besides, TgWRKY14 in root, TgWRKY10 in stem and leaves showed the highest transcript abundances and also were obviously induced under stress.

Under salt condition, 8 TgWRKY genes in roots keep a rising expression level over time, 7 genes keep repressed, 52 genes were upregulated during the first 24 h and then decreased, 23 genes were downregulated during the first 24 h and then increased, 0 gene was switched off and 2 genes were turned on at 48 h. In stems and leaves, the corresponding quantities are 25, 5, 14, 48, 1, 0 and 8, 9, 29, 45, 1, 1, respectively. The highest induced expression (97-fold increase) in roots was detected in TgWRKY 32 at 48 h. The most significant downregulated gene is TgWRKY84 with 35-fold decrease at 24 h. In stems, TgWRKY 22 was the most upregulated gene at 24 h with 125- fold change, and TgWRKY 84 was the most down-regulated gene at 24 h with 436-fold decrease. In leaves, TgWRKY 100 was significantly induced at 24 h with 18-fold increase, and TgWRKY 96 was repressed more than 42-fold at 24 h. Besides, TgWRKY79 in root, TgWRKY10 in stem and leaves showed the highest transcript abundances and also were obviously induced under stress.

In addition, three TgWRKY genes (TgWRKY 1/2/50) were did not expressed in seedlings, which was obviously distinct with the expression of mature individuals. TgWRKY 19, which was unexpressed in mature individuals, was expressed in seedlings with high TPM (Transcripts Per Kilobase of exon model per Million mapped reads) values. TgWRKY 34 was only weakly detected in roots and leaves under stress. TgWRKY 55 and 94 were weakly detected only in leaves under drought stress. TgWRKY 61 was weakly detected in stems under stress. TgWRKY 102 in roots, TgWRKY 53/87 in stems and TgWRKY 80/98 in leaves of seedlings were unexpressed.

Discussion

The WRKY gene family is one of the largest transcription factors families that play an important role in growth and development, and abiotic stress resistance in plant. Previously, many studies have been carried out to exploring the evolution and gene function of WRKY genes (Cannon et al. 2004, Ross et al. 2007, Xie at al. 2018). In the present study, extensive analysis of the WRKY gene family in teak identified

102 members, which were designated TgWRKY1-TgWRKY102 on the basis of their chromosomal location (Fig. 4). The 102 FtWRKYs could be divided into three main groups, which was consistent with previous studies. From the phylogenetic tree, we identified at least one TgWRKY and AtWRKY in each subgroup (Fig. 2).

The domain gain and loss is a divergent force for expansion of the WRKY gene family. The loss of WRKY domain in Group I seems to be common in many monocotyledons, which appear infrequently in dicotyledonous (Wei et al. 2012). In our analysis, gene loss can be observed in TgWRKY6 and TgWRKY7, suggesting the different characteristics of this group during teak evolution. The phylogenetic tree shows that the N-terminal WRKY domains in Group I clustered as a monophyletic branch, whereas the C-terminal WRKY domains, II-a, II-b and II-c were clustered together (Fig. 2). This result consisted with previous studies and suggest that the single WRKY domains of group II family members are more closely related to the C-terminal domains of group I than to the N-terminal domains (Wei et al. 2012, Xie et al. 2018).

The conserved structural domains of the predicted teak WRKY proteins were assessed in this study. Multiple sequences alignments revealed that six TgWRKY proteins in group I-N, II-c and III had sequences variation in their WRKY domain (see Additional file 2). The WRKYGKK sequence present in TgWRKY21, 39 and 75; the WRKYGEK variants present in TgWRKY37 and 96; the WHKYGQK variant was only observed in TgWRKY7. According to previous studies, variation in the WRKYGQK motif in WRKY domain might influence normal interactions of WRKY genes with downstream target genes, and therefore these six proteins might be worthy to further investigate their functions and binding specificities (Xiong et al. 2013, van Verk et al. 2008, Zhou et al. 2008). The TgWRKY members of the specific subgroups likely shared closely related motif compositions and functional similarities, which was supported by the subsequent gene structure analysis (Fig. 3).

The synteny analysis inferred that tandem and segmental duplication events

have played a critical role in the expansion of WRKY gene family. Ancient WGDs, also referred to as paleopolyploidizations, are widespread in plant lineages and represent a powerful evolutionary force for the development of novel gene functions and the emergence of new species (Vanneste et al. 2014). In this study, Three WGDs were identified and the majority of gene duplications (32 segmental duplicated) were occurred during 15.3–33.9 Mya, around the split of *Tectona grandis* with other Lamiaceae species (21.4508–33.283 Mya, Yasodha et al. 2018). These duplication maybe significantly associated with the speciation of genus *Tectona*.

Six segmental duplicated were occurred during 82.5–119.3 Mya, which maybe represented the γ WGD event that was shared by all eudicots and estimated in 117 Mya (Jiao et al. 2012). The remained 18 duplicated gene pairs occurred around the K–Pg (Cretaceous–Paleogene) boundary (50.5–69.1 Mya), during which an estimated ~ 75% of all living species became extinct (Raup 1994, Renne et al. 2013). This WGD maybe result from environment and ecological challenges in the K–Pg extinction (Vanneste et al. 2014).

Like the expression pattern of WRKY in other species, different genes showed incongruous expression patterns, suggesting they have various function and diversity regulatory mechanisms (He et al. 2019, Ma et al. 2015). According to the RNA-seq data of teak (see Additional file 5), all the genes were detected in tissues, but four genes (TgWRKY50, 61, 94, 98) were only weakly expressed in few tissues ($0 < \text{TPM} < 1$). 56 TgWRKY genes were expressed in all samples tested ($\text{TPM} > 0$). 33 genes were identified as constitutive expression ($\text{TPM} > 1$ in all samples), which were continuous and stable expressed at different growth stages, tissues and under stress. Remarkably, TgWRKY19 was not expressed in mature teak plant, whereas TgWRKY1, 2 and 50 were not expressed in seedlings. These genes may have functions in different growing times. Our analysis also revealed that the tissue-specific expression pattern exhibited by certain TgWRKY members. For example, TgWRKY55 was only highly expressed in flowers, which inferred its association with flower development. TgWRKY51 were exhibited the highest expression in pulp, indicating that TgWRKY51 could mediate the fruit development. TgWRKY10 was observed in root, stem, leaf and flower with high expression, however, it is weakly detected in pulp and core.

Substantial evidence has shown that WRKY TFs can improve the stress tolerance of plants by modulating their molecular and physiological metabolism (Karanja et al. 2017, Yu et al. 2016). In our study, more than half of the TgWRKY genes were considered as stress response genes as these genes could be markedly induced under salt and drought stress. TgWRKY22, 32, 49, 54 and 84 were highly regulated by drought stress, suggesting that these genes might act as important drought stress regulators responses factors. TgWRKY22, 32, 84, 96 and 100 were highly induced by salt stress and might act as important salt stress regulators responses factors. TgWRKY10, 14 and 79, which have the highest transcript abundances in different tissues of seedlings under drought or salt stress, were also obviously induced, indicating that these genes also play a vital role in plant stress tolerance.

Conclusions

In this study, we identified a total of 102 teak WRKY genes from the *T. grandis* genome sequence, which could be divided into three groups. The chromosomal localization, phylogenetic distribution, exon-intron structure analysis, and prospective motif composition provided a basis for the understanding of WRKY gene family evolution. The results also suggested that duplication events played a critical role in the expansion of WRKY gene family and three WGD events were inferred. The expression analysis of WRKY genes in response to drought and salt stress were studied. The phylogenetic and gene expression analysis will shed light on the functional analysis of TgWRKY genes. This study generated an important resource that will provide helpful information for further exploration of the TgWRKY genes role in the regulatory mechanism in response to abiotic stresses.

Materials And Methods

Genome-wide identification, nomenclature and distribution pattern of WRKY family genes.

To accurately identify WRKY TFs in teak, both all the teak proteins and whole genome sequences were retrieved from GigaScience Database (Zhao et al. 2019). The Hidden Markov Model (HMM) profiles of the WRKY domain (PF03106) was downloaded from the Pfam family database (<http://pfam.xfam.org/>). All possible WRKY proteins were searched using HMMER 3.0 (Finn et al., 2011) with default parameters and the cutoff was set to 0.01. All the identified WRKY sequences were further confirmed separately against the Pfam database and NCBI-Conserved Domains Database for conserved domains. The naming of the teak WRKY genes was performed according to their chromosomal locations. The properties of the proteins, including length of sequences, molecular weights, isoelectric points, aliphatic index, instability index, and GRAVY index were calculated using the ExPasy website (<http://web.expasy.org/protparam/>), while the subcellular location of each protein was predicted with Plant-mPLoc (<http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/>).

Sequence analysis and Phylogeny.

The WRKY family members of *Arabidopsis thaliana* were retrieved from TAIR database (<https://www.arabidopsis.org/>). The WRKY domain sequences of the characterized WRKY proteins were used to create multiple protein sequence alignments using MAFFT version 7 (Kato et al. 2019, <https://mafft.cbrc.jp/alignment/server/index.html>) with default parameters. Sequence alignment was used to generate a phylogenetic tree using the MEGA X software tool. The neighbor-joining method with 1000 bootstrap replicates interactively using iTOL (<https://itol.embl.de/>). All identified teak WRKY genes were divided into different groups according to the AtWRKY classification scheme and the alignment of WRKY domains of TgWRKY and AtWRKY proteins.

Determination of gene structures and conserved motifs of teak WRKY.

The exon-intron structures of TgWRKY genes were determined through the comparison of the coding sequence of each TgWRKY gene with its genomic sequence using Gene Structure Display Server (GSDS v2.0; <http://gsds.cbi.pku.edu.cn/>) (Hu et al., 2015). The conserved motifs were found for the identified TgWRKYs using MEME server v5.1.1 (Bailey et al., 2009). The optimized parameters were employed as the following: -nostatus -time 30000 -maxsize 6000000 -mod anr -nmotifs 20 -minw 6 -maxw 100.

Chromosomal Location And Gene Duplication

The physical location of TgWRKYs on chromosomes was retrieved from the annotation file of teak genome, and all genes were plotted separately onto pseudomolecules and scaffolds based on the order of their physical position. Finally, the image of their physical location was generated using MapChart software (<https://www.wur.nl/en/show/Mapchart.htm>). The gene duplication events were analyzed using MCScanX and the synteny relationship of the orthologous WRKY genes were exhibited using Circos (Krzywinski et al. 2009). Non-synonymous (Ka) and synonymous (Ks) substitution of each duplicated WRKY genes were calculated using KaKs_Calculator 2.0 (Wang et al. 2010). The divergence time (T) was calculated by $T = Ks / (2 \times 1.5 \times 10^{-8})$ Mya (Koch et al. 2000).

Plant Materials And Treatments

Tectona grandis cv. Relin7029, a typical cultivated variety, was used throughout the study. The root, stem, leaf, flower, pulp and core from mature teak plants, were collected separately for RNA extraction and used for further RNA-seq analysis. In order to discover stress response genes in teak, the seeds of *Tectona grandis* cv. Relin7029 were got from the Research Institute of Tropical Forestry, Chinese Academy of Forestry. Plants were grown in pots containing soil and vermiculite mixture (3:1) in greenhouse. Stress treatments were initiated in 12-week-old normal seedlings, and the seedlings were disposed with following treatment as described by Xie et al. (2018) and Li et al. (2017). For salt treatment, seedlings were irrigated with 150 mM NaCl. For drought stress treatments, water was withheld for 2 days and the seedlings were temporary wilting. The roots, stems and leaves were harvested after 24 and 48 h, while tissues at 0 h were used as the control. Samples from six biological replicates were mixed and frozen immediately in liquid nitrogen and stored at -80°C until further analysis.

Rna Extraction And Gene Expression Analysis

Total RNA was isolated from frozen tissues with an RNAprep Pure Plant Kit (TIANGEN, Beijing, China) following the manufacturer's recommended protocol. All RNA was analyzed by agarose gel electrophoresis and then quantified with a Nanodrop ND-1000 spectrophotometer. For the construction of RNA-seq libraries, 2 µg of total RNA was processed using the TruSeq RNA Sample Preparation kit (Illumina) followed by sequencing on the Illumina HiSeq 2500 platform. Transcript expression levels were calculated by Kallisto v 0.46.2 software (Bray et al. 2016) in TPM units as transcripts per kilobase of exon model per Million mapped reads. TPM value were transformed by log₂ and the heatmap was performed by R package Pheatmap v1.0.12 (Kolde 2019).

Abbreviations

TPM: Transcripts Per Kilobase of exon model per Million mapped reads;

TF: Transcription factor

Ka: The rate of nonsynonymous substitutions

Ks: The rate of synonymous substitutions

WGD: Whole-genome duplications

Mya: Million years ago

Declarations

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and materials: Supporting data and materials are available in the NCBI SRA database under BioProject accession number PRJNA671563.

Competing interests: The authors declare that they have no competing interests.

Funding: The research is supported by the National Natural Science Foundation of China (Project Nos. 31800529, 31900016).

Authors' contributions

J.S. and X.W. designed the sequencing strategy. J.S. and J.L. prepared and analysed the samples. J.X. drew the figures. J.S. and X.W. wrote the manuscript. All authors read and approved the final manuscript..

Acknowledgements

Special thanks are due to Dr. Xue-Yu Pan and Mrs. Zhong-Hui Zhang for help in collecting specimens.

References

- Agarwal P, Reddy MP, Chikara J. WRKY: Its structure, evolutionary relationship, DNA-binding selectivity, role in stress tolerance and development of plants. *Molecular Biology Reports*. 2011;38:3883–3896.
- Bailey TL, Bonden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*. 2009;37:202–208.
- Birkenbihl RP, Kracher B, Ross A, Kramer K, Finkemeier I, Somssich IE. Principles and characteristics of the Arabidopsis WRKY regulatory network during early MAMP-triggered immunity. *Plant Journal*. 2018;525–527.
- Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*. 2016;34: 525–527.
- Cannon SB, Mitra A, Baumgarten A, Young ND, May G. The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana. *BMC Plant Biology*. 2004;4:10.
- Chen F, Hu Y, Vannozzi A, Wu K, Cai H, Qin Y, Mullis A, Lin Z, Zhang L. The WRKY Transcription Factor Family in Model Plants and Crops. *Critical Reviews in Plant Sciences*. 2017;36: 311–335.
- Chen H, Lai Z, Shi J, Xiao Y, Chen Z, Xu X. Roles of arabidopsis WRKY18, WRKY40 and WRKY60 transcription factors in plant responses to abscisic acid and abiotic stress. *BMC Plant Biology*. 2010;281.
- Dias LP, de Oliveira-Busatto LA, Bodanese-Zanettini MH. The differential expression of soybean [Glycine max (L.) Merrill] WRKY genes in response to water deficit. *Plant Physiology and Biochemistry*. 2016;107:

288–300.

Dong J, Chen C, Chen Z. Expression profiles of the Arabidopsis WRKY gene superfamily during plant defense response. *Plant Molecular Biology*. 2003;51: 21–37.

Dou L, Zhang X, Pang C, Song M, Wei H, Fan S, Yu S. Genome-wide analysis of the WRKY gene family in cotton. *Molecular Genetics and Genomics*. 2014;289(6): 1103–1121.

Eulgem T, Rushton PJ, Robatzek S, Somssich IE. The WRKY superfamily of plant transcription factors. *Trends Plant Science*. 2000;5:199–206.

Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*. 2011;39:29–37.

Gu Z, Cavalcanti A, Chen FC, Bouman P, Li WH. Extent of gene duplication in the genomes of *Drosophila*, nematode, and yeast. *Molecular Biology and Evolution*. 2002;19:256–262.

He X, Li JJ, Chen Y, Yang JQ, Chen XY. Genome-wide analysis of the WRKY gene family and its response to abiotic stress in buckwheat (*Fagopyrum tataricum*). *Open Life Science*. 2019;14:080–96.

Holub EB. The arms race is ancient history in Arabidopsis, the wildflower. *Nature Reviews Genetics*. 2001;2:516.

Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*. 2015;31:1296–1297.

Ishiguro S, Nakamura K. Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5' upstream regions of genes coding for sporamin and β -amylase from sweet potato. *MGG Molecular and General Genetics*. 1994;244: 563–571.

<https://doi.org/10.1007/BF00282746>

Jiao YN, Leebens-Mack J, Ayyampalayam S, Bowers JE, McKain MR, McNeal J, et al. A genome triplication associated with early diversification of the core eudicots. *Genome Biology*. 2012;13:R3.

Kalde M, Barth M, Somssich IE, Lippok B. Members of the Arabidopsis WRKY group III transcription factors are part of different plant defense signaling pathways. *Molecular Plant-Microbe Interactions*. 2003;16:295–305.

<https://doi.org/10.1094/MPMI.2003.16.4.295>

Karanja BK, Fan L, Liang X, Yan W, Zhu X, Tang M, et al. Genome-wide characterization of the WRKY gene family in radish (*Raphanus sativus* L.) reveals its critical functions under different abiotic stresses. *Plant Cell Reports*. 2017;36:1757–1773.

Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*. 2019;20:1160–1166.

Koch MA, Haubold B, Mitchell-Olds T. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Molecular Biology and Evolution*. 2000;17:1483–1498.

Kolde R. pheatmap: Pretty Heatmaps. R package version 1.0.12. 2019. <https://CRAN.R-project.org/package=pheatmap>.

Kollert W, Cherubini L. 2012. Teak Resources and Market Assessment 2010. FAO Planted Forests and Trees Working Paper FP/47/E, Rome. Available at

<http://www.fao.org/forestry/plantedforests/67508@170537/en/>.

Kollert W, Kleine M. The Global Teak Study. Analysis, Evaluation and Future Potential of Teak Resources, International Union of Forest Research Organizations (IUFRO), World Series, Vienna. 2017.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. *Genome Research*. 2009;19:1639–45.

Li DH, Liu Pan, Yu JY, Wang LH, Dossa K, Zhang YX, et al. Genome-wide analysis of WRKY gene family in the sesame genome and identification of the WRKY genes involved in responses to abiotic stresses. *BMC Plant Biology*. 2017;17:152.

Ling J, Jiang W, Zhang Y, Yu H, Mao Z, Gu X, et al. Genome-wide analysis of WRKY gene family in *Cucumis sativus*. *BMC Genomics*. 2011;12:471. <https://doi.org/10.1186/1471-2164-12-471>

Ma J, Gao X, Liu Q, Shao Y, Zhang D, Jiang L, Li C. Overexpression of TaWRKY146 increases drought tolerance through inducing stomatal closure in *Arabidopsis Thaliana*. *Frontiers in Plant Science*. 2017;8:2036

Ma J, Lu J, Xu J, Duan B, He X, Liu J. Genome-wide identification of wrky genes in the desert poplar *Populus euphratica* and adaptive evolution of the genes in response to salt stress. *Evolutionary Bioinformatics*. 2015;11:47–55.

Pandey D, Brown C. Teak: A global overview. *Unasylva*. 2000;51:3-13.

Pérez D, Kanninen M. Stand growth scenarios for *Tectona grandis* plantations in Costa Rica. *Forest Ecology and Management*. 2005; 210: 425–441.

Ross CA, Liu Y, Shen QJ. The WRKY gene family in rice (*Oryza sativa*). *Journal of Integrative Plant Biology*. 2007;49:827–842.

- Van Verk MC, Pappaioannou D, Neeleman L, Bol JF, Linthorst HJ. A novel WRKY transcription factor is required for induction of PR-1a gene expression by salicylic acid and bacterial elicitors. *Plant Physiology*. 2008;146:1983–95.
- Vanneste K, Baeke G, Maere S, de Peer YV. Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous–Paleogene boundary. *Genome Research*. 2014;24:1334–1347.
- Wang D, Zhang Y, Zhang Z, Zhu J, Xu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics*. 2010;8:77–80.
- Wang M, Vannozzi A, Wang G, Liang YH, Tornielli GB, Zenoni S, et al. Genome and transcriptome analysis of the grapevine (*Vitis vinifera* L.) WRKY gene family. *Horticulture Research*. 2014;1: 14016.
- Wei KF, Chen J, Chen YF, Wu LJ, Xie DX. Molecular Phylogenetic and Expression Analysis of the Complete WRKY Transcription Factor Family in Maize. *DNA Research*. 2012;19:153–164.
- Wei Y, Shi H, Xia Z, Tie W, Ding Z, Yan Y, et al. Genome-wide identification and expression analysis of the WRKY gene family in cassava. *Front. Plant Sci*. 2016;
- Wu KL. The WRKY Family of Transcription Factors in Rice and Arabidopsis and Their Origins. *DNA Research*. 2005;12:9–26.
- Xie T, Chen C, Li C, Liu J, Liu C, He Y. Genome-wide investigation of WRKY gene family in pineapple: evolution and expression profiles during development and stress. *BMC Genomics*. 2018;19:490.
- Xiong W, Xu X, Zhang L, Wu P, Chen Y, Li M, et al. Genome-wide analysis of the WRKY gene family in physic nut (*Jatropha curcas* L.). *Gene*. 2013;524:124–132.
- Yasodha R, Vasudeva R, Balakrishnan S, Sakthi AR, Abel N, Binai N, Rajashekar B, Bachpai VKW, Pillai C, Dev SA. Draft genome of a high value tropical timber tree, Teak (*Tectona grandis* L. f): insights into SSR diversity, phylogeny and conservation. *DNA Research*. 2018;25:409–419.
- Yu Y, Nan W, Hu R, Xiang F. Genome-wide identification of soybean WRKY transcription factors in response to salt stress. *Springerplus*. 2016;5:920–935.
- Zhao D, Hamilton JP, Bhat WW, Johnson SR, Godden GT, Kinser TJ, Boachon B, Dudareva N, Soltis DE, Soltis PS, Hamberger B, Buell CR. A chromosomal-scale genome assembly of *Tectona grandis* reveals the importance of tandem gene duplication and enables discovery of genes in natural product biosynthetic pathways. *GigaScience*. 2019;8:1–10.
- Zhou QY, Tian AG, Zou HF, Xie ZM, Lei G, Huang J, Wang CM, Wang HW, Zhang JS, Chen SY. Soybean WRKY-type transcription factor genes, GmWRKY13, GmWRKY21, and GmWRKY54, confer differential

Figures

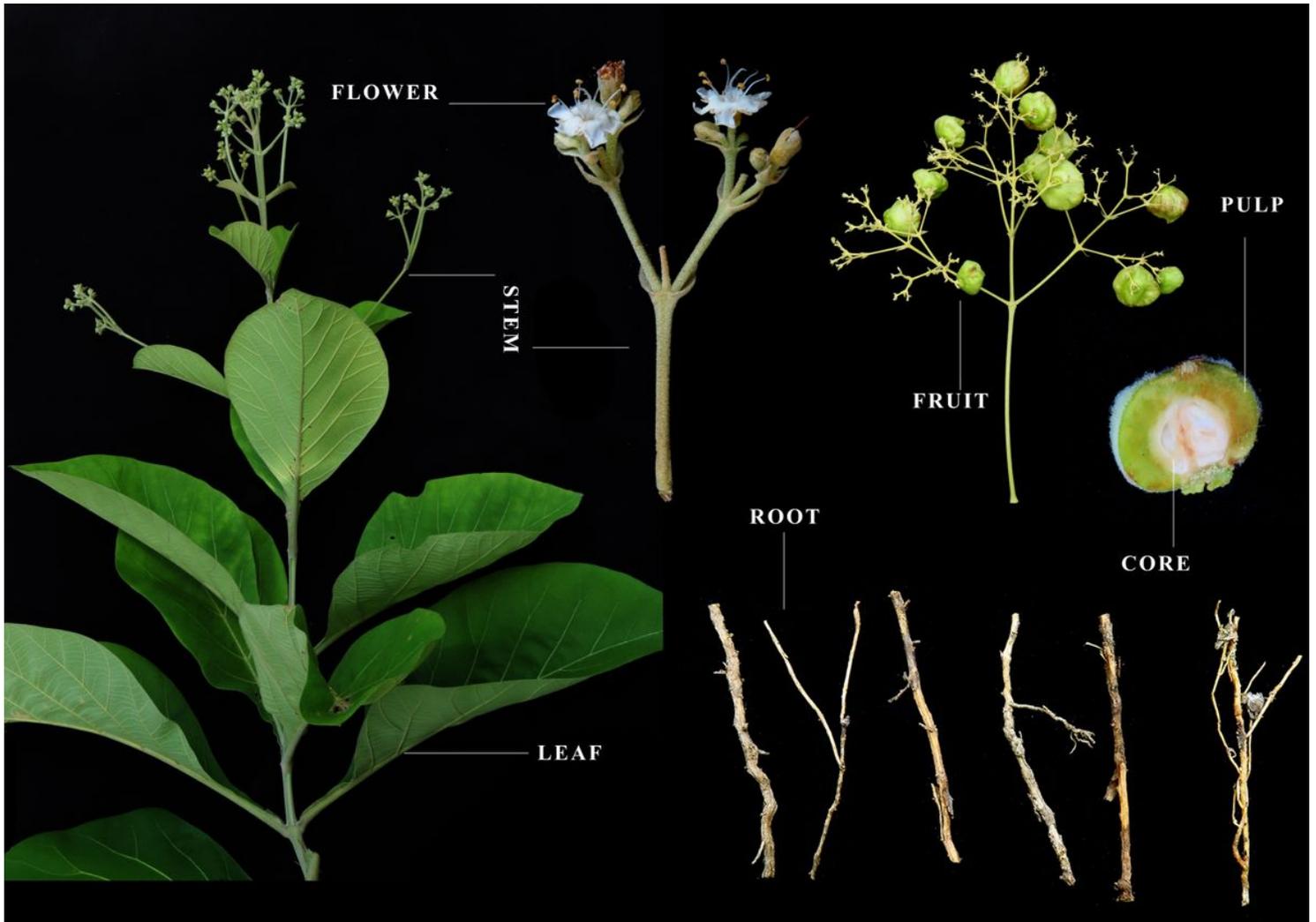


Figure 1

Tissus from a 15 years old mature teak tree.

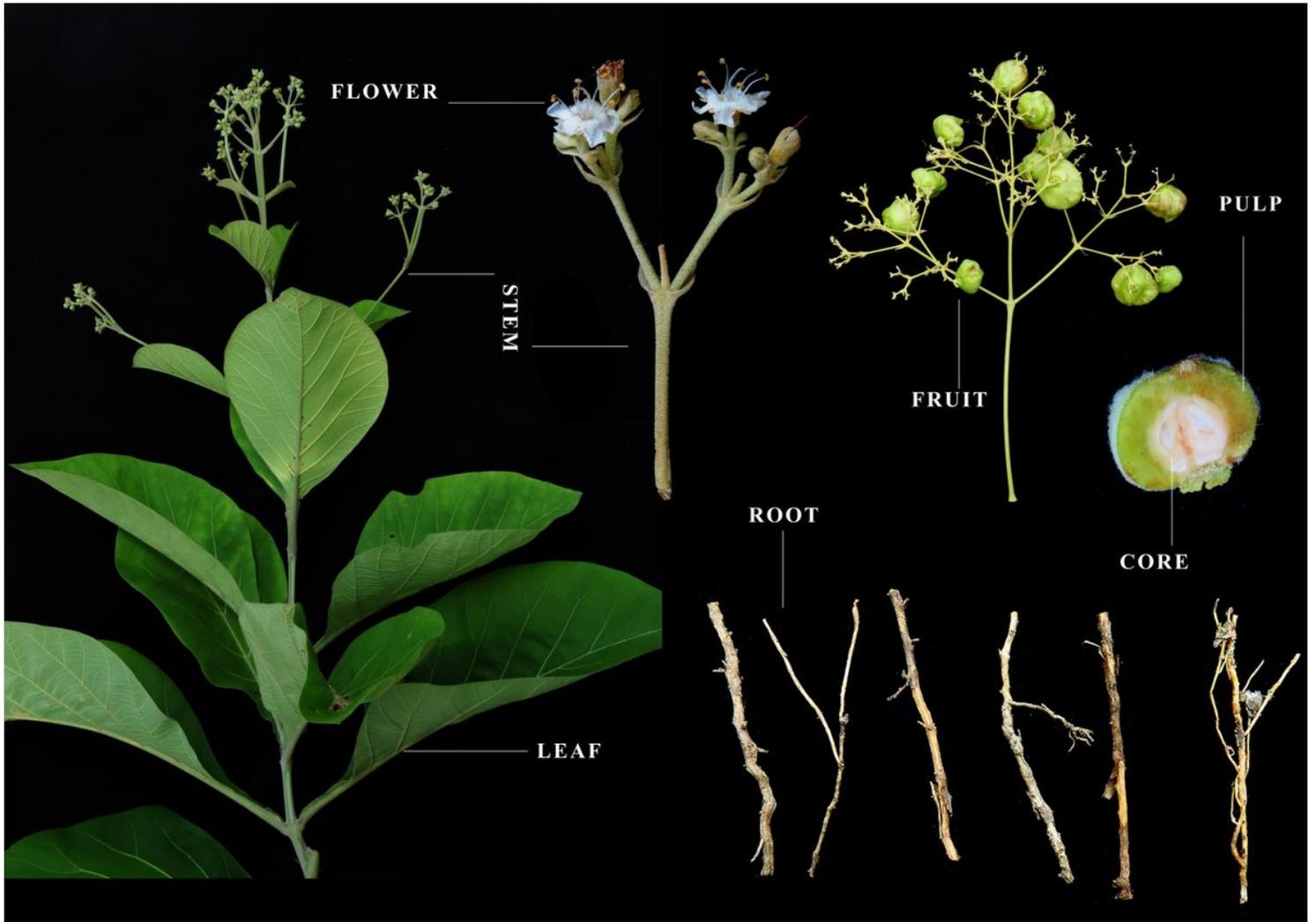


Figure 1

Tissus from a 15 years old mature teak tree.

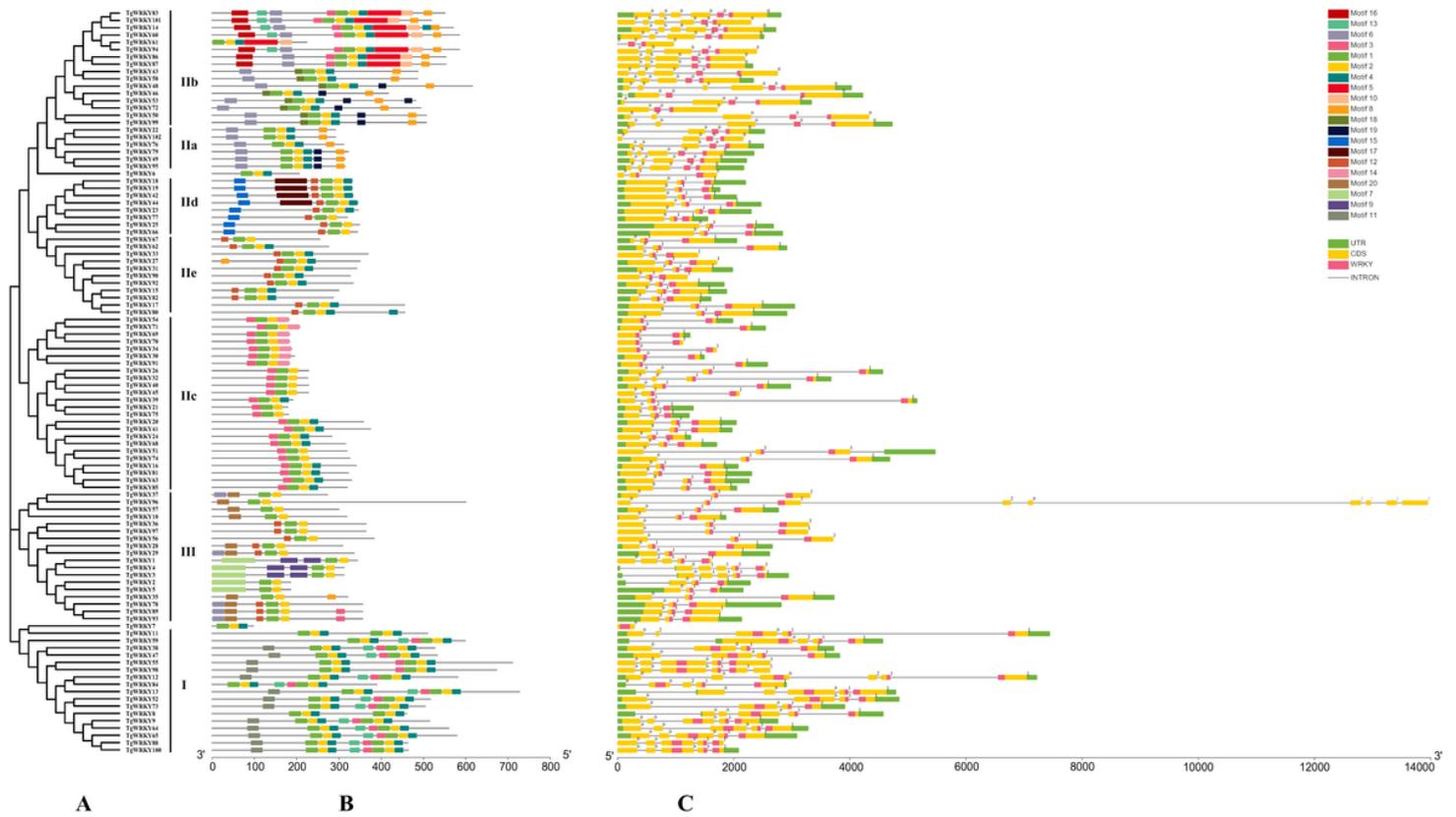


Figure 3

Phylogenetic relationships, architecture of conserved protein motifs and gene structure in WRKY genes from pineapple. A. The phylogenetic tree was constructed based on the full-length sequences of teak WRKY proteins using MEGA 5 software. B. The motif composition of teak WRKY proteins. The motifs, numbers 1–20, are displayed in different colored boxes. C. Exon-intron structure of teak WRKY genes. Green boxes indicate untranslated 5'- and 3'-regions; yellow boxes indicate exons; black lines indicate introns. The WRKY domains are highlighted by red boxes. The number indicates the phases of corresponding introns.

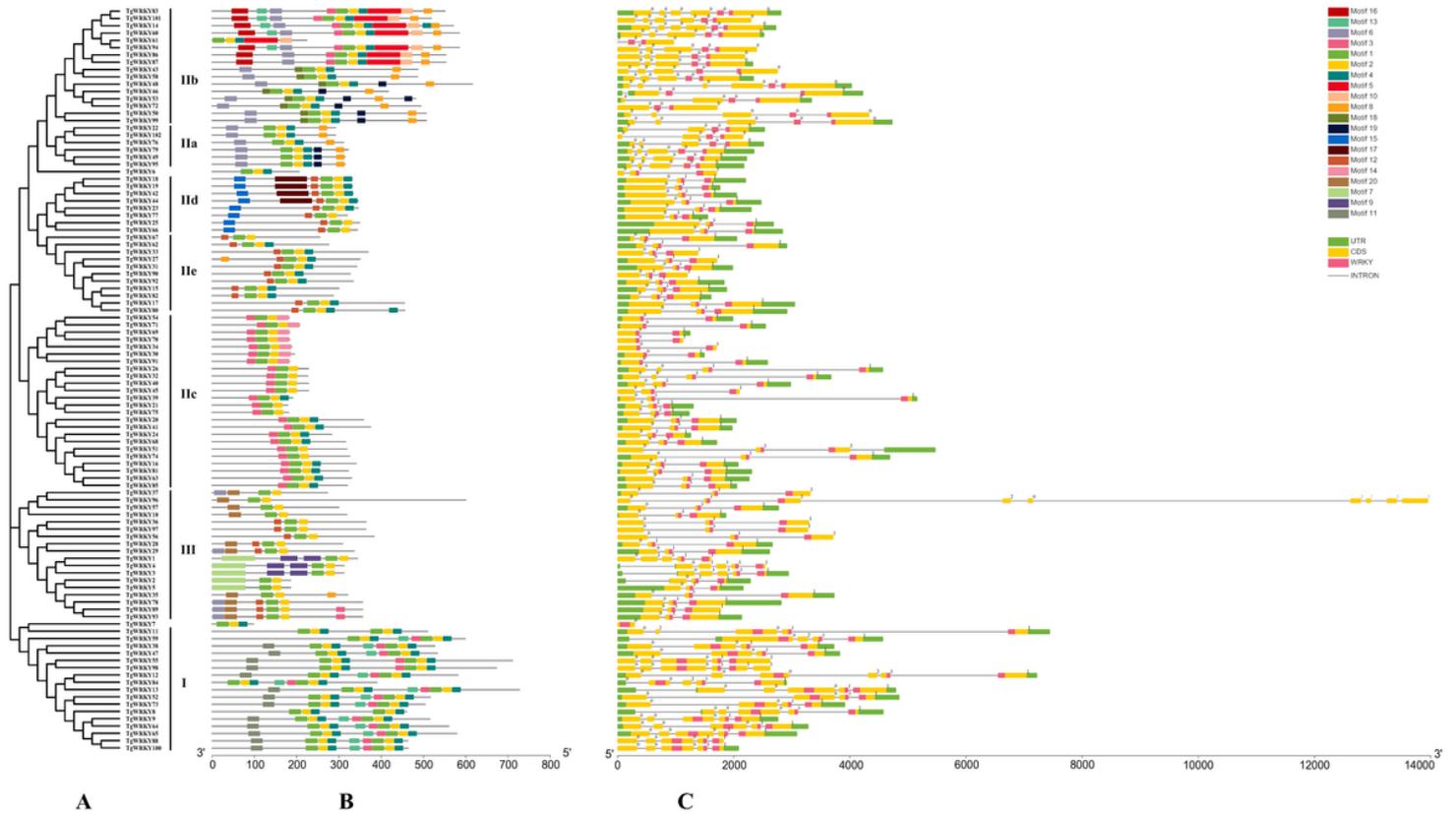


Figure 3

Phylogenetic relationships, architecture of conserved protein motifs and gene structure in WRKY genes from pineapple. A. The phylogenetic tree was constructed based on the full-length sequences of teak WRKY proteins using MEGA 5 software. B. The motif composition of teak WRKY proteins. The motifs, numbers 1–20, are displayed in different colored boxes. C. Exon-intron structure of teak WRKY genes. Green boxes indicate untranslated 5'- and 3'-regions; yellow boxes indicate exons; black lines indicate introns. The WRKY domains are highlighted by red boxes. The number indicates the phases of corresponding introns.

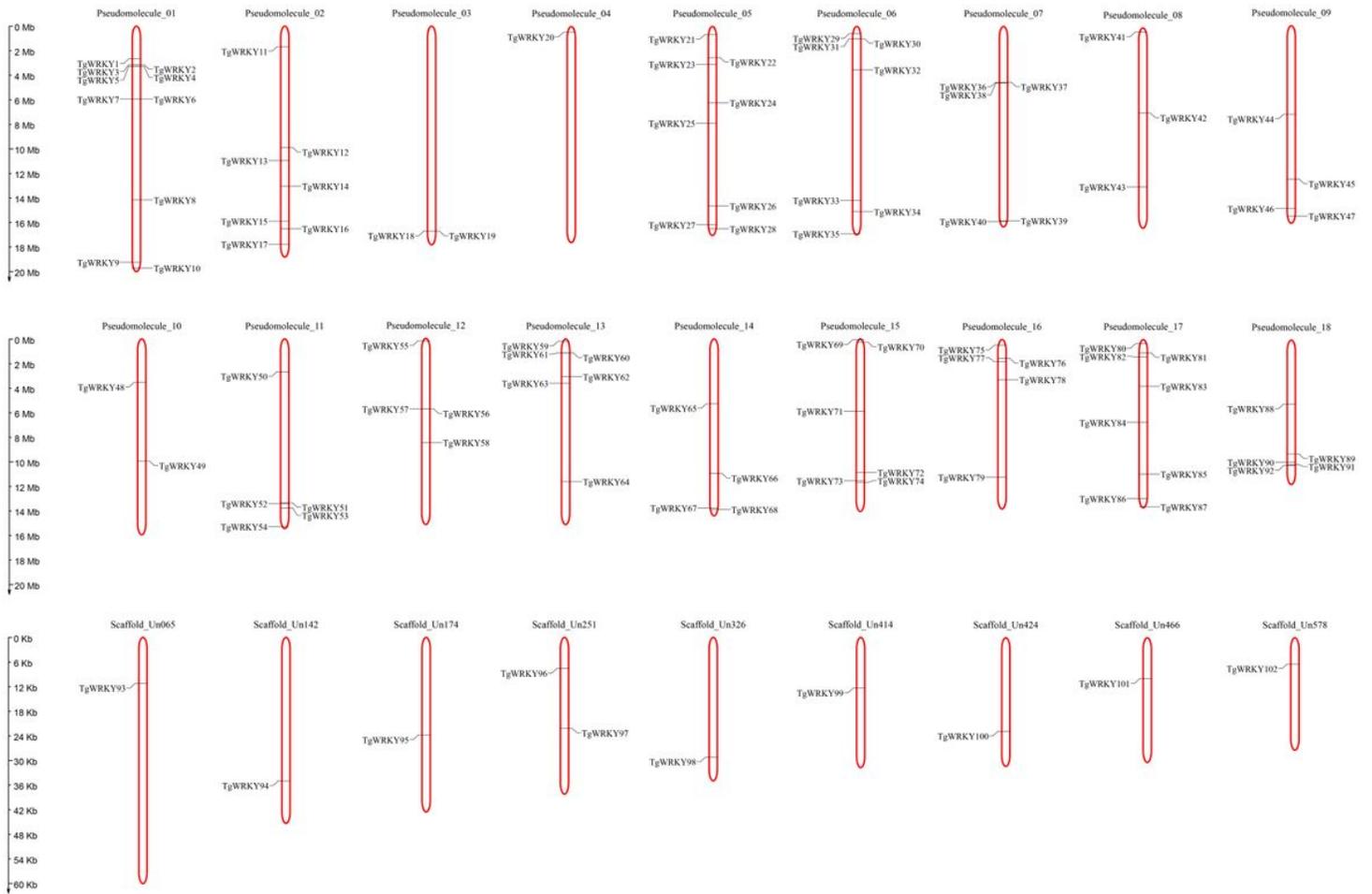


Figure 4

Chromosomal location of TgWRKY genes in 18 Pseudomolecules and 9 Scaffolds.

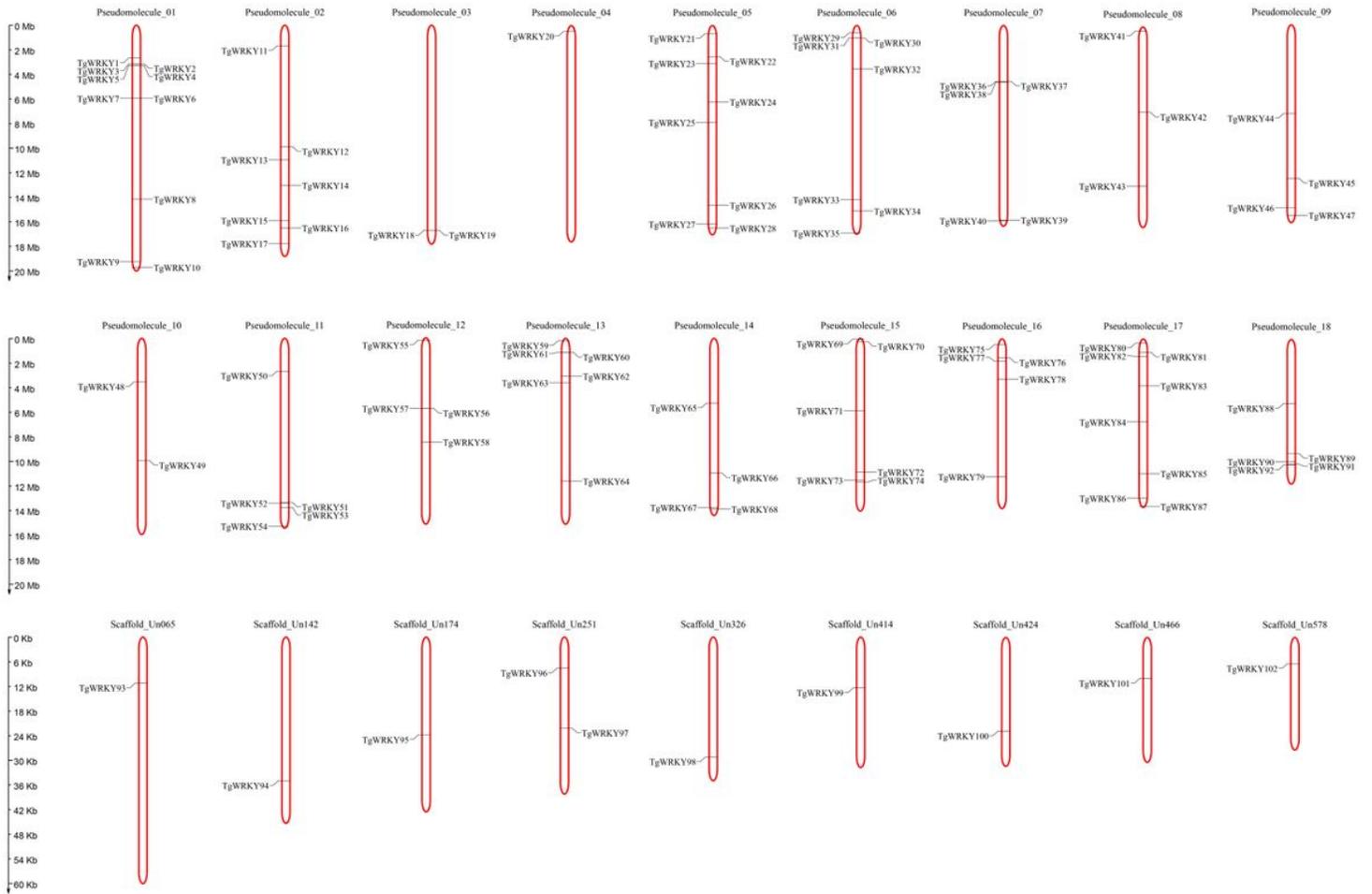


Figure 4

Chromosomal location of TgWRKY genes in 18 Pseudomolecules and 9 Scaffolds.

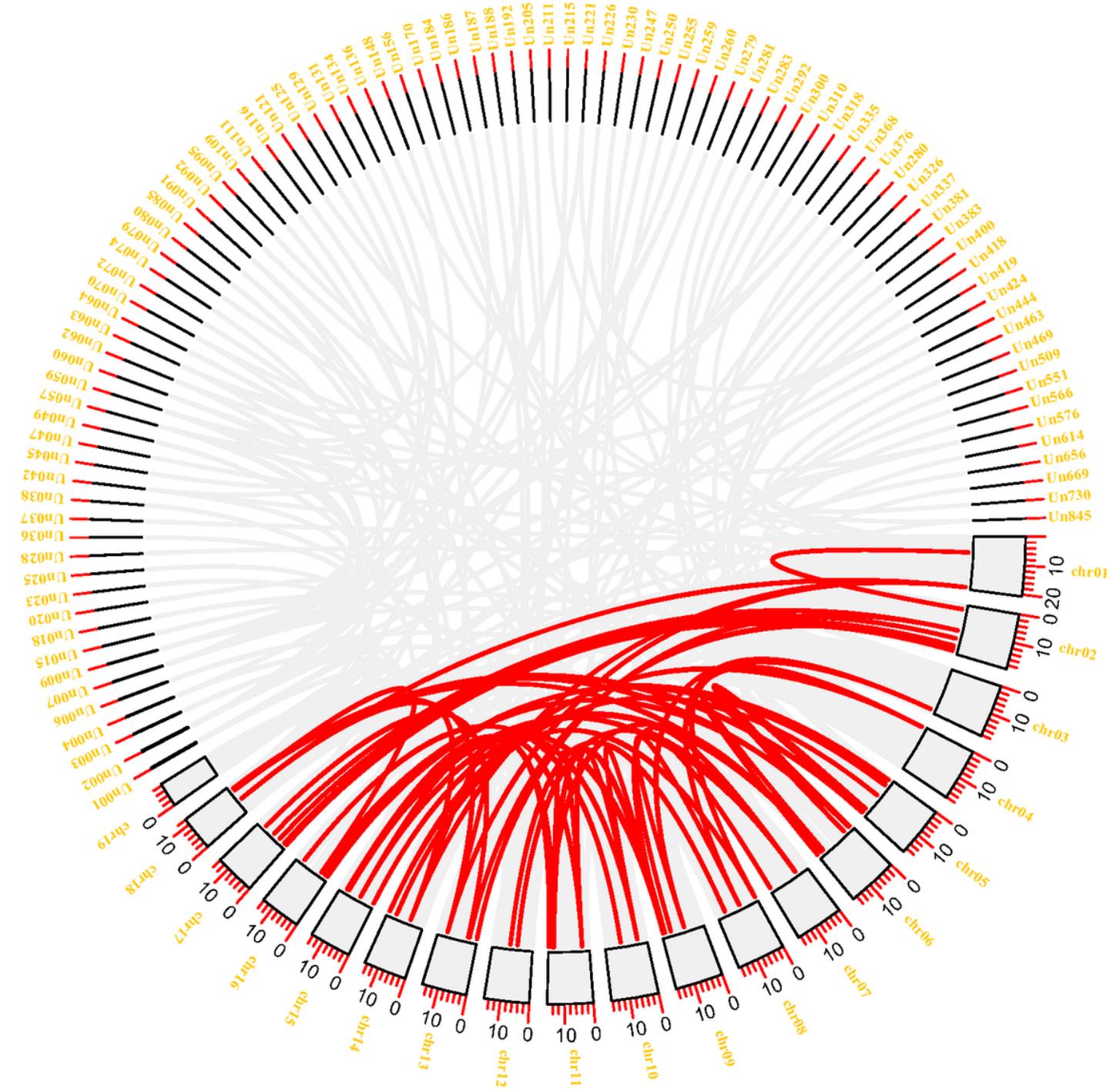


Figure 5

Schematic representations for the chromosomal distribution and interchromosomal relationships of teak WRKY genes. Gray lines indicate all syntenic blocks in the teak genome, and the red lines indicate duplicated WRKY gene pairs. The chromosome number is indicated at the bottom of each chromosome.

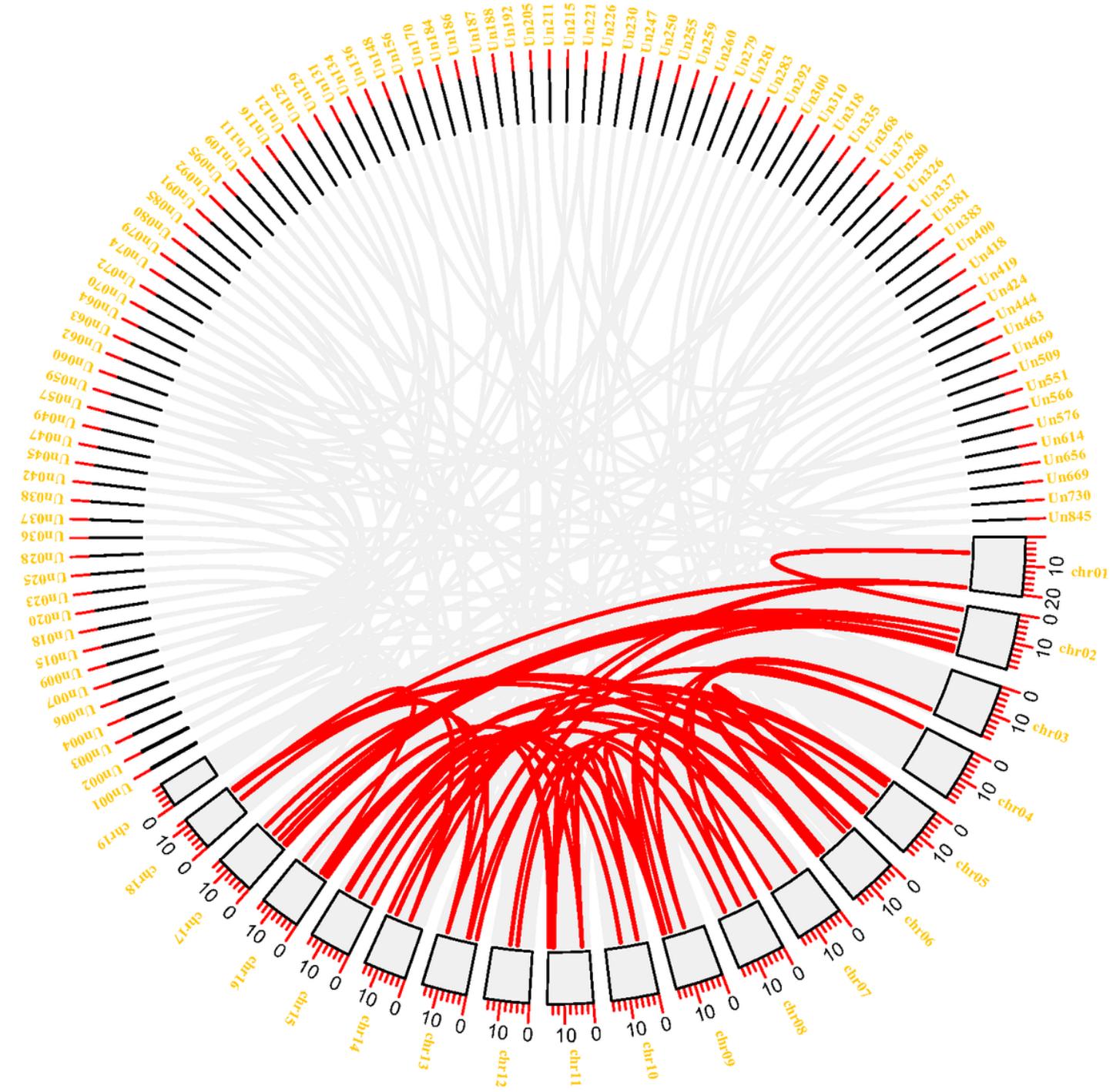


Figure 5

Schematic representations for the chromosomal distribution and interchromosomal relationships of teak WRKY genes. Gray lines indicate all syntenic blocks in the teak genome, and the red lines indicate duplicated WRKY gene pairs. The chromosome number is indicated at the bottom of each chromosome.

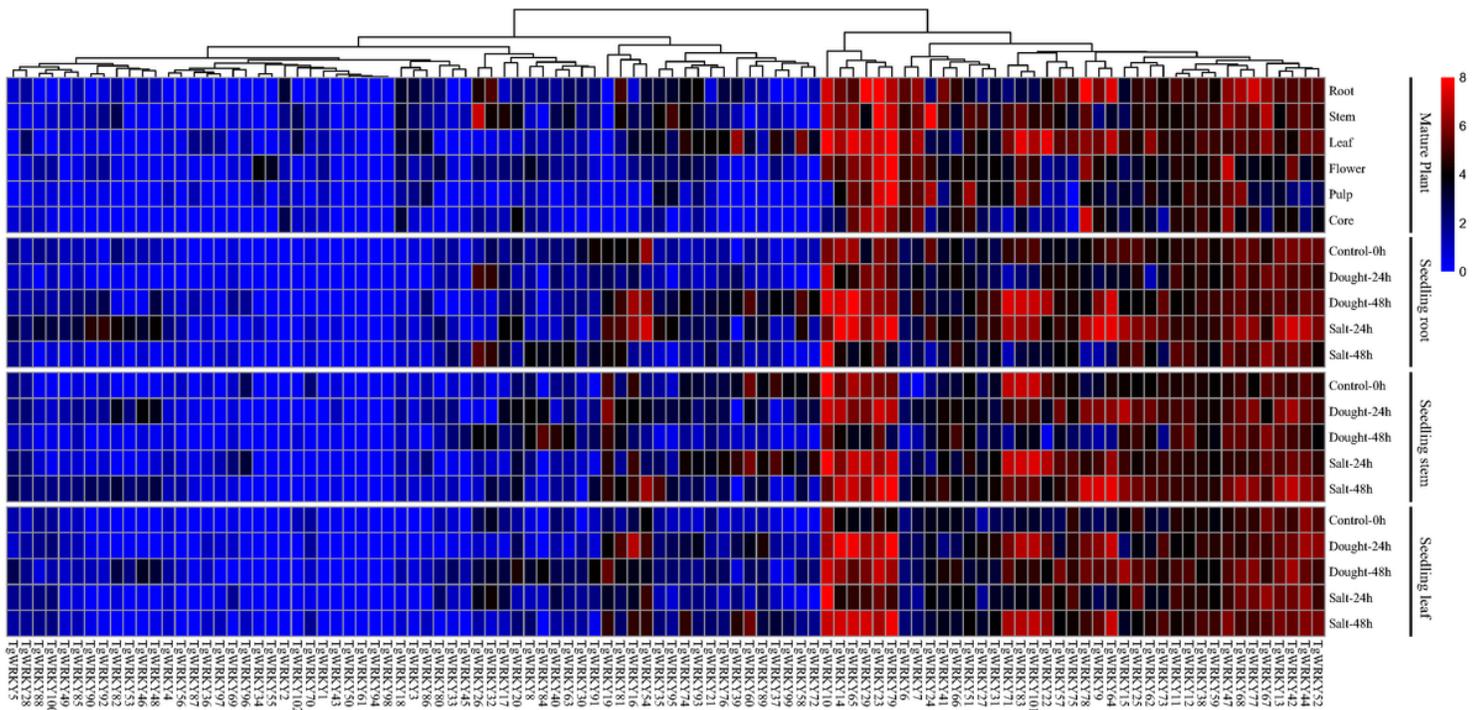


Figure 6

Hierarchical clustering of expression profiles of teak WRKY genes in different samples including different tissues and under stress.

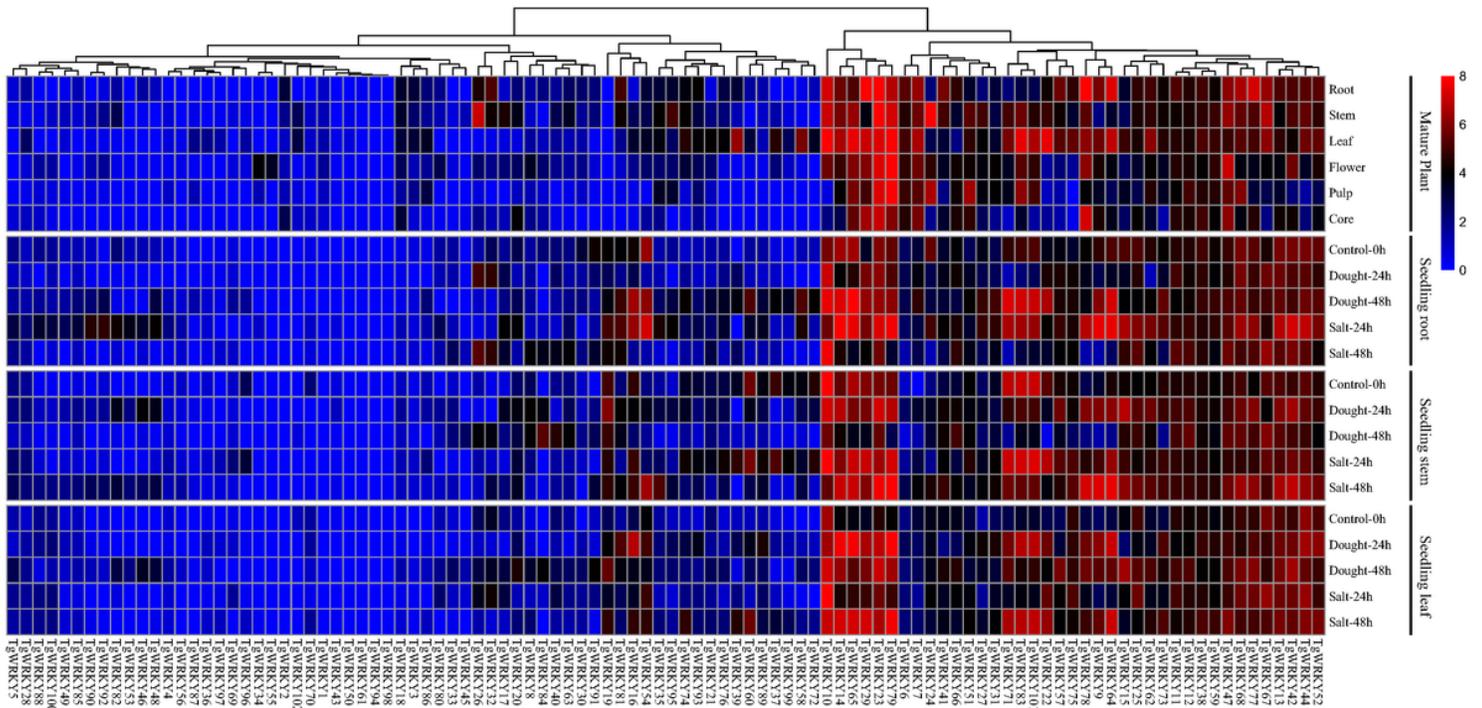


Figure 6

Hierarchical clustering of expression profiles of teak WRKY genes in different samples including different tissues and under stress.

Supplementary Files

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

- [Additionalfile1.docx](#)
- [Additionalfile1.docx](#)
- [Additionalfile2.docx](#)
- [Additionalfile2.docx](#)
- [Additionalfile3.xlsx](#)
- [Additionalfile3.xlsx](#)
- [Additionalfile4.xlsx](#)
- [Additionalfile4.xlsx](#)
- [Additionalfile5.xlsx](#)
- [Additionalfile5.xlsx](#)