Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Genome-wide identification and integrative analysis of the WRKY gene family reveal candidate regulators involved in biotic stress responses in Hass avocado

Xiafei Wang
Yunnan Institute of Tropical Crops
Zekun Shi
Avocado Science and Technology Research Centre of Menglian County
Bo Yuan
Yunnan Institute of Tropical Crops
Liang Tao
Yunnan Institute of Tropical Crops
Yangmei Zhang
Yunnan Institute of Tropical Crops
Zhonghua Wu
Yunnan Institute of Tropical Crops
Hongchang A
Yunnan Institute of Tropical Crops
Guoyuan Zhu
Yunnan Institute of Tropical Crops
Liyue Xu
Yunnan Institute of Tropical Crops
Yongke Zhang
Yunnan Institute of Tropical Crops
Jinqiang Wang
wjinq@126.com

Yunnan Institute of Tropical Crops

Research Article

Keywords: Hass avocado, Persea americana, WRKY transcription factor, Gene family, Biotic stress, Insect herbivory, Pathogen infection, Transcriptome analysis

Posted Date: June 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9992724/v1

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

Additional Declarations: No competing interests reported.

Genome-wide identification and integrative analysis of the WRKY gene family reveal candidate regulators involved in biotic stress responses in Hass avocado

Xiafei Wang[1+], Zekun Shi[2+], Bo Yuan[1], Liang Tao[1], Yangmei Zhang[1], Zhonghua Wu[1], Hongchang A[1], Guoyuan Zhu[1], Liyue Xu[1], Yongke Zhang[1*] and Jinqiang Wang[1*]

[+]Xiafei Wang and Zekun Shi contributed equally to this work.

*Correspondence:

Yongke Zhang

zhangyongke99@163.com

Jinqiang Wang

wjinq@126.com

Full list of author information is available at the end of the article

ABSTRACT

Background Avocado (*Persea americana* Mill.) is a significant fruit crop in tropical and subtropical regions. However, its growth, yield, and fruit quality are substantially impacted by biotic stresses, such as insect herbivory and pathogen infections. WRKY transcription factors are crucial in regulating plant growth, development, hormone signaling, and stress responses. Despite their importance, comprehensive genome-wide identification and characterization of the WRKY gene family, particularly in response to various biotic stresses, remain largely uninvestigated in Hass avocado. This study aims to identify the WRKY gene family at the genome-wide level and to explore its evolutionary characteristics, gene structures, tissue-specific expression patterns, and responses to biotic stresses in Hass avocado.

Results A total of 70 *PaWRKY* genes were identified within the Hass avocado genome, exhibiting an uneven distribution across 12 chromosomes. Through analyses of phylogenetic relationships and conserved domain composition, these *PaWRKY* genes were categorized into three principal groups, with Group II further subdivided into five subgroups. Segmental duplication emerged as the primary mechanism driving the expansion of the PaWRKY gene family, with most duplicated gene pairs appearing to have undergone purifying selection. Investigations into gene structure, conserved motifs, *cis*-regulatory elements, and protein interaction networks indicated that PaWRKY members maintained relative conservation within subgroups, while also demonstrating potential for functional diversification. Transcriptomic analyses revealed distinct tissue-specific expression patterns and differential responses of *PaWRKY* genes to insect-infested leaves, diseased leaves, and insect-bored fruits. Compared to insect feeding, diseased leaves elicited more extensive and robust defense responses, which were further linked to tissue damage and senescence-related processes. Through integrative analyses employing K-means clustering, weighted gene co-expression network analysis (WGCNA), differentially expressed genes (DEG) screening, and UpSet plots, several candidate core *PaWRKY* genes implicated in biotic stress responses were identified. The predicted three-dimensional protein structures revealed that these core PaWRKY proteins exhibit localized structural variations while

maintaining highly conserved WRKY domains.

Conclusion This study systematically characterizes the WRKY gene family in Hass avocado, emphasizing its evolutionary diversification and differential responses to biotic stresses. These findings offer valuable candidate genes and theoretical support for future resistance breeding and molecular enhancement of avocado to combat biotic stresses.

Keywords Hass avocado; *Persea americana*; WRKY transcription factor; Gene family; Biotic stress; Insect herbivory; Pathogen infection; Transcriptome analysis.

Backgroud

Among them, biotic stresses caused by pathogenic microorganisms and herbivorous insects can disrupt plant tissue integrity and physiological metabolism, ultimately leading to growth inhibition, quality deterioration, and yield losses [1, 2]. To mitigate these stresses, plants have developed intricate defense mechanisms that preserve structural and functional stability. These defense responses are governed by complex regulatory networks, wherein transcription factors (TFs) play pivotal roles. They achieve this by binding to *cis*-acting elements in the promoters of downstream target genes or by interacting with other TFs to regulate functional gene expression during plant responses to biotic stress [3, 4].

Among the various TF families in plants, the WRKY family is notable for being plant-specific and is among the most extensive in the plant kingdom [5]. In the model organisms *Arabidopsis thaliana* and rice, 74 and 100 WRKY members have been identified, respectively. In contrast, plants with more complex genomes, such as *Brassica juncea*, *Fragaria* × *ananassa*, *Gossypium hirsutum*, and *Camelina sativa*, have been reported to contain over 200 WRKY members [6-9]. The WRKY family is named for its highly conserved WRKY domain, characterized by an N-terminal WRKYGQK heptapeptide and a C-terminal zinc-finger motif, which can be either C2H2 (C–$X_{4–5}$–C–$X_{22–23}$–H–X–H) or C2HC (C–X_{7}–C–X_{23}–H–X–C) [10]. WRKY proteins regulate downstream gene expression by recognizing W-box *cis*-acting elements [TGACC(A/T)] in the promoters of target genes [11]. Based on the number

of WRKY domains and the type of zinc-finger motif, WRKY proteins are generally classified into three groups: I, II, and III. Group I proteins contain two WRKY domains, while Groups II and III contain a single WRKY domain. Additionally, Groups I and II feature C2H2-type zinc-finger motifs, whereas Group III is characterized by C2HC-type motifs [10]. Group II can be further categorized into five subgroups: IIa, IIb, IIc, IId, and IIe [11]. The structural diversity observed in WRKY proteins is likely a consequence of whole-genome duplication events coupled with adaptive pressures arising from diverse ecological conditions [12].

An increasing body of evidence suggests that WRKY proteins play a crucial role in plant growth and development, hormone signal integration, and stress responses. Notably, under conditions of biotic stress, WRKY proteins are often rapidly and robustly induced, facilitating plant responses to fungal, bacterial, and viral pathogens, as well as herbivore predation. This is achieved through the regulation of defense-related gene expression, hormone signaling pathways, and secondary metabolic processes [13-15]. Mechanistically, WRKY proteins function as transcriptional regulators, either activating or repressing gene expression by binding to W-box *cis*-acting elements located in the promoters of target genes, thus influencing downstream gene expression [16]. In cotton, the activation of *GhWRKY40a* facilitates the expression of ethylene-responsive factors and defense-related genes, thereby augmenting resistance to Fusarium wilt [17]. Furthermore, research has demonstrated that a WRKY protein from subgroup IIc can directly interact with and activate the promoter of *GhMKK2*, a pivotal gene in the MAPK signaling pathway, thereby modulating signal cascade regulation and enhancing cotton's defense mechanisms [18]. WRKY proteins also contribute to plant resistance by promoting the biosynthesis of secondary metabolites. In rice, *OsWRKY67* undergoes phosphorylation by *OsMPK6*, subsequently inducing the expression of naringenin 7-O-methyltransferase (*OsNOMT*). This induction results in the accumulation of the flavonoid phytoalexin sakuranetin, significantly mitigating the severity of rice false smut [19]. In grapevine, the overexpression of *VqWRKY56* bolsters resistance to powdery mildew through interaction with *VqbZIPC22* and activation of genes involved in proanthocyanidin (PA)

biosynthesis, thereby increasing PA accumulation [20]. Furthermore, WRKY transcription factors play a pivotal role in mediating plant defenses against pathogens and insect pests by modulating hormone signaling pathways, particularly those involving salicylic acid (SA) and jasmonic acid (JA). In *Betula platyphylla*, *BpWRKY6* directly interacts with the promoters of genes related to JA biosynthesis, such as *BpLOX15*, *BpAOC4*, and *BpAOS1*, leading to increased JA accumulation and consequently enhancing birch resistance to insect herbivory [21]. In *Camellia sinensis*, an LRR-RLK–MPK–WRKY–JA signaling module is activated following insect feeding, underscoring the essential role of WRKY in JA-mediated anti-herbivore signaling [22]. With respect to SA signaling, WRKY in pepper can bind directly to W-box elements in the promoters of *pathogenesis-related* (*PR*) genes, which are key markers of the SA signaling pathway, thereby bolstering resistance to *Phytophthora capsici* [23]. Similarly, in *Nicotiana benthamiana*, *NbWRKY40* interacts with the W-box element in the *NbICS1* promoter, facilitating SA accumulation and upregulating the signaling genes *PR1b* and *PR2* [24].

Avocado (*Persea americana* Mill.), belonging to the genus *Persea* within the family Lauraceae, represents a significant tropical and subtropical fruit crop of economic importance, indigenous to Central America. The fruit of the avocado is abundant in minerals, vitamins, proteins, and unsaturated fatty acids, thereby conferring substantial nutritional value and health-promoting attributes [25-27]. Avocados are typically categorized into three botanical races based on their ecological origins: Mexican, Guatemalan, and West Indian. Among these, the Hass variety, a hybrid of the Guatemalan and Mexican races, currently dominates as the primary commercial cultivar [28]. Recent data from FAOSTAT indicate that global annual avocado production has surpassed 10 million tonnes, positioning avocado among the four major fresh tropical fruits worldwide [29]. Mexico stands as the leading avocado-producing nation, with an annual output of 2.761 million tonnes [30]. In recent years, propelled by rising market demand, the avocado industry in China has experienced rapid growth, with Yunnan emerging as the principal avocado-producing region and Menglian County evolving into the largest avocado cultivation base in the country. In 2024, it was

documented that the avocado cultivation area in Menglian County expanded to 120,700 mu, with 25,700 mu actively under production, yielding an annual fresh fruit output of 19,500 tonnes [31]. Similar to other fruit crops, avocados are nutrient-rich but susceptible to damage from herbivorous insects and pathogens, which can significantly impact fruit yield and quality, leading to considerable economic losses [32, 33]. Our preliminary survey of Hass avocado plantations in Menglian County and adjacent regions revealed frequent occurrences of leaf spot disease, caused by *Pseudocercospora purpurea* (Cooke) Deighton, and malformed fruits due to gall midge infestation (Fig. 1). These findings suggest that local Hass avocado production is currently contending with significant combined pressures from diseases and insect pests.

Presently, research on WRKY transcription factors in avocados is limited. Fick et al. identified 82 *WRKY* genes within the genome of a West Indian avocado accession and further analyzed their expression profiles in roots and lower stems following infection by *Phytophthora cinnamomi* [32]. Nonetheless, due to the significant genetic background variations among different avocado races and cultivars, the current findings may not comprehensively reflect the composition of the WRKY gene family or its response mechanisms to biotic stress within the Hass cultivar. Consequently, this study undertakes a systematic identification of the WRKY gene family utilizing the Hass avocado genome. Subsequently, we conducted an analysis of their structural characteristics, conserved motifs, phylogenetic relationships, and *cis*-acting elements. Furthermore, we explored their expression patterns and potential functions under various biotic stress conditions. This research aims to lay the groundwork for elucidating the molecular mechanisms that underpin disease and insect resistance in Hass avocado, as well as for identifying candidate genes associated with resistance.

Materials and Methods

Plant materials, stress treatments, and sample collection

Avocado samples were procured from the Nuodong orchard located in Menglian County, Yunnan Province, China. The mature experimental trees, aged between 8 and 10 years, were cultivated under natural field conditions. In June 2025, three distinct

types of leaf samples were collected: (1) fully expanded healthy leaves devoid of visible damage or symptoms (CKL, Fig. 1A); (2) leaves exhibiting fresh feeding damage inflicted by common chewing insects present in the orchard (IHL, Fig. 1B), where both the damaged area and adjacent tissues were collected; and (3) leaves displaying characteristic leaf spot symptoms induced by *P. purpurea* infection (DL, Fig. 1C). Leaf samples were exclusively selected from the second to fourth fully expanded leaves below the shoot apex. Additionally, two types of fruit tissues were collected: (1) undamaged healthy fruits (CKF, Fig. 1D); and (2) developmentally arrested malformed fruits affected by gall midge larvae (IHF, Fig. 1E). Furthermore, three biological replicates of newly emerged buds and green stems were collected from different shoot tips. Samples were collected between 10:30 and 11:00 a.m., rinsed thrice with sterile water, rapidly frozen in liquid nitrogen, and subsequently stored at −80 °C. For each tissue type, three biological replicates were prepared, with each replicate comprising pooled tissues from a minimum of three independent trees.

Identification of WRKY genes in *P. americana*

P. americana was conducted utilizing the most recent whole-genome sequence and annotation data for Hass avocado. The WRKY gene family in avocado (*PaWRKY*) was identified through two complementary methodologies. Initially, protein sequences of the *Arabidopsis thaliana* WRKY gene family (*AtWRKY*) were retrieved from the *Arabidopsis* Information Resource (TAIR, https://www.arabidopsis.org/; accessed on 13 June 2025). These sequences served as queries to identify homologous sequences among all avocado protein sequences using BLASTP version 2.16.0, employing an E-value threshold of 1E−5, a sequence identity of at least 40%, and an alignment coverage of at least 50%. Second, the hidden Markov model (HMM) profile corresponding to the WRKY domain (PF03106) was retrieved from the Pfam database (http://pfam-legacy.xfam.org/; accessed on 13 June 2025) and employed as a query to interrogate the annotated avocado protein sequences using HMMER v3.4 [33], applying an E-value threshold of 1E−5 and a domain coverage criterion of ≥ 50%. The candidate PaWRKY protein sequences identified through these two methodologies were consolidated, and

redundant sequences were eliminated using CD-HIT v4.8.1 [34]. The conserved domains within the remaining candidate sequences were subsequently validated through the Conserved Domain Database (CDD, http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi; accessed on 13 June 2025) and the Simple Modular Architecture Research Tool (SMART, http://smart.embl-heidelberg.de/; accessed on 13 June 2025). The isoelectric point (pI) and molecular weight (MW) of each PaWRKY protein were predicted utilizing the R package *Peptides*. Additionally, the subcellular localization of the PaWRKY proteins was predicted using the online WoLF PSORT program (https://wolfpsort.hgc.jp/; accessed on 13 June 2025).

Chromosomal distribution and gene duplication

The chromosomal locations of *PaWRKY* genes were visualized utilizing the *RIdeogram* package in R version 4.4.1, based on the avocado genome annotation file [35]. The PaWRKY family members were designated sequentially according to their physical positions on the chromosomes. Tandem and segmental duplication events among *PaWRKY* genes were identified using MCScanX [36] and subsequently visualized with Circos version 0.69–8 [37]. The synonymous (Ks) and nonsynonymous (Ka) substitution rates of duplicated gene pairs were calculated using TBtools version 2.315 [38]. Furthermore, the Python-based MCScanX (JCVI) pipeline [39] was employed to conduct comparative synteny analyses among avocado, *A. thaliana*, camphor tree (*Cinnamomum camphora*), rice (*Oryza sativa*), and pineapple (*Ananas comosus*), facilitating the visualization of interspecific relationships of *WRKY* genes.

Sequence alignment and phylogenetic analysis

The identified PaWRKY protein sequences underwent multiple sequence alignment using MUSCLE (https://www.ebi.ac.uk/Tools/msa/muscle/) with default settings. The resulting alignment was visualized and examined using Jalview v2.11.4.1 [40]. Subsequently, a neighbor-joining phylogenetic tree was constructed utilizing MEGA11 [41] with 1,000 bootstrap replicates to ensure robustness. The PaWRKY proteins were categorized into distinct groups and subgroups based on the alignment of WRKY

domains and their phylogenetic relationships. Furthermore, *A. comosus* WRKY (AcWRKY) protein sequences were retrieved from a prior study [42], and a comprehensive phylogenetic tree was generated incorporating all PaWRKY, AtWRKY, and AcWRKY protein sequences. The visualization of the phylogenetic tree was optimized by Evolview v3 [43].

Gene structure, *cis*-acting element, and protein interaction network analyses

The conserved motifs of PaWRKY proteins were analyzed utilizing the MEME suite (https://meme-suite.org/meme/tools/meme), with the parameters set to a maximum of 20 motifs, allowing for any number of repetitions, and a motif width range of 6 to 100. The coding sequence (CDS) distribution for each *PaWRKY* gene was visualized using R version 4.4.1, based on the avocado genome annotation file. To analyze *cis*-acting elements, the 1,500 base pair upstream sequences of all PaWRKY coding DNA sequences were submitted to the PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/), and the results were visualized using R version 4.4.1. Protein-protein interaction networks among PaWRKY proteins were predicted using the STRING database (https://string-db.org/), with *A. thaliana* selected as the reference organism. The minimum required interaction score was set at 0.40, while all other parameters were maintained at their default settings.

Transcriptome data processing and differential expression analysis

Total RNA was isolated from the collected avocado tissues utilizing the TRIzol reagent (Tiangen, Beijing, China) in accordance with the manufacturer's guidelines. The concentration and purity of the RNA were determined using a NanoDrop 2000 microvolume spectrophotometer (Thermo Fisher Scientific, MA, USA), while RNA integrity was assessed via an Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA). RNA-seq libraries were prepared employing the Hieff NGS Ultima Dual-mode mRNA Library Prep Kit (Yeasen, Shanghai, China) as per the manufacturer's protocol and subsequently sequenced on the Illumina NovaSeq platform (Illumina, CA, USA). The raw sequencing reads underwent processing with fastp v0.22.0

(https://github.com/OpenGene/fastp) to eliminate adapter sequences, poly-N reads, and low-quality reads. The quality of the filtered reads was further evaluated using FastQC v0.12.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) prior to conducting downstream analyses.

The reference genome sequence and annotation files for avocado were procured from the NCBI database (GCA_051132755.1; accessed on June 13, 2025). High-quality, clean reads were aligned to the reference genome utilizing HISAT2 version 2.1.0. Gene read counts and FPKM (fragments per kilobase of transcript per million mapped reads) values were calculated using StringTie and HTSeq version 2.0.2, with FPKM values serving as normalized expression estimates. Principal component analysis (PCA) was conducted using the *prcomp* function in R to assess sample grouping and biological reproducibility. To elucidate stress-responsive gene expression patterns in leaves and fruits, the optimal number of K-means clusters was initially determined using the *cascadeKM* function from the vegan R package, followed by K-means clustering. Clustering was executed with 1,000 random starts (nstart = 1000) to enhance the stability of the results. Genes within each cluster exhibited similar expression trends. Functional enrichment analyses for each gene cluster were conducted utilizing Gene Ontology (GO; www.geneontology.org) [44] and Kyoto Encyclopedia of Genes and Genomes (KEGG; www.genome.jp/kegg/) [45] terms, employing the *clusterProfiler* R package version 4.6.2 [46]. Differential expression analysis was executed using the *DESeq2* package version 1.38.3 [47] in R v4.2.2. Genes exhibiting an adjusted *p*-value of less than 0.01 and an absolute fold change of at least 2 were classified as differentially expressed genes (DEGs). The expression values of the identified *PaWRKY* genes were extracted to examine their expression profiles across various tissues and to assess their responses to diverse stress conditions. Clustered heatmaps of gene expression were generated using the *circlize* package version 0.4.16 in R.

Weighted gene co-expression network analysis

Previous research has demonstrated that the sequence in which DEG filtering and weighted gene co-expression network analysis (WGCNA) are executed can

significantly influence the precision of functional enrichment outcomes. Notably, conducting WGCNA to identify trait-associated modules prior to DEG filtering has been shown to more effectively capture genes that are significantly correlated with specific traits [48]. Consequently, to enhance the accuracy of stress-responsive gene screening and functional enrichment analysis, and to maintain network topology during WGCNA, co-expression networks were constructed utilizing the expression profiles of all annotated genes across various treatments before implementing DEG-based filtering. The WGCNA was carried out using the *WGCNA* package in R [49]. Initially, sample clustering was performed to identify and exclude outlier samples. Subsequently, a similarity matrix was generated based on Pearson correlation coefficients derived from gene expression values. This similarity matrix was then transformed into an adjacency matrix, and the optimal soft-thresholding power was determined using the *pickSoftThreshold* function. A topological overlap matrix (TOM) was developed utilizing the chosen soft-thresholding power, and genes were categorized into modules through dynamic tree cutting based on 1−TOM dissimilarity. Pearson correlation analysis was employed to assess the associations between module eigengenes and various stress treatments, with results visualized via the *labeledHeatmap* function. Modules exhibiting a *p*-value of less than 0.05 were deemed statistically significant. Genes within these significant modules underwent GO and KEGG enrichment analyses to elucidate their potential biological functions.

Integrated screening of *PaWRKY* genes under biotic stresses

To further elucidate the key *PaWRKY* genes implicated in biotic stress responses, gene sets associated with insect-infested leaves (IHL), diseased leaves (DL), and insect-bored fruits (IHF) were extracted from the aforementioned analyses. These analyses included relevant K-means cluster gene sets, gene sets derived from WGCNA that exhibited significant module correlations, and DEG sets. Subsequently, these gene sets were integrated with the 70 identified *PaWRKY* genes. The intersections among the various gene sets were visualized using the *UpSetR* package version 1.4.0 in R [50] o identify candidate stress-responsive *PaWRKY* genes. *PaWRKY* genes located at the

intersection of K-means clusters, WGCNA modules, and DEG sets were designated as high-confidence core response genes. Conversely, *PaWRKY* genes present in one or two of the relevant gene sets were classified as extended stress-responsive genes. Based on the expression matrix, the expression levels of genes corresponding to each intersection category were extracted and visualized as $\log_{10}(\text{FPKM} + 1)$. To validate the transcriptome-based expression patterns, the core stress-responsive *PaWRKY* genes were further analyzed using quantitative real-time PCR (qRT-PCR) (Table S1). The avocado *β-actin* gene served as an internal reference, and relative expression levels were determined utilizing the $2^{-(\Delta\Delta Ct)}$ method. Each sample comprised three biological replicates, with each biological replicate further analyzed in three technical replicates.

Three-dimensional protein structure prediction

The protein sequences of core stress-responsive *PaWRKY* genes, identified under various biotic stress conditions, were subjected to homology modeling using the SWISS-MODEL platform (https://swissmodel.expasy.org/) to predict their tertiary structures. The quality of the resulting three-dimensional models was assessed through the Global Model Quality Estimate (GMQE) score and sequence identity metrics [51]. The GMQE score, which ranges from 0 to 1, serves as an indicator of model reliability, with higher scores denoting increased reliability. Only models exhibiting a sequence identity of 90% or greater were retained for further analysis.

Statistical analysis

Statistical analyses were conducted utilizing SPSS version 23.0. Mean differences were assessed through Duncan's multiple range test. Results are expressed as the mean ± standard deviation (SD), with statistical significance determined at a threshold of $p < 0.05$.

Results

Identification of WRKY proteins in *P. americana*

Utilizing *A. thaliana* WRKY protein sequences in conjunction with the HMM profile specific to the WRKY domain (PF03106), an initial identification of 73 putative

members of the WRKY gene family was conducted in avocado via BLASTP and HMMER searches. Subsequent analyses focusing on sequence integrity and conserved domain presence refined this number to 70 PaWRKY gene family members, as detailed in Table S2. These genes were systematically designated as *PaWRKY1* through *PaWRKY70*, following their respective chromosomal positions within the avocado genome (Fig. 2A). The CDS lengths of the *PaWRKY* genes varied from 373 base pairs (bp) in *PaWRKY19* to 2,483 bp in *PaWRKY20*. Correspondingly, the resultant proteins ranged from 124 amino acids (PaWRKY19) to 829 amino acids (PaWRKY20), with an average protein length of 371 amino acids. The molecular weights of the PaWRKY proteins spanned from 14.62 kilodaltons (kDa) for PaWRKY19 to 90.58 kDa for PaWRKY20, averaging at 40.93 kDa. The predicted isoelectric points of these proteins ranged from 4.70 (PaWRKY68) to 10.39 (PaWRKY22), with an average isoelectric point of 7.59. The prediction of subcellular localization indicated that all PaWRKY proteins, with the exception of three (PaWRKY14, PaWRKY19, and PaWRKY25) predicted to localize within chloroplasts and one (PaWRKY12) predicted to localize within the cytoskeleton, were anticipated to localize to the nucleus (Table S2).

Chromosomal localization, gene duplication, and comparative genomic analysis

Utilizing the avocado genome annotation, the 70 *PaWRKY* genes were systematically mapped onto 12 chromosomes employing the *RIdeogram* package in R. The distribution of *PaWRKY* genes across the avocado chromosomes was notably uneven (Fig. 2A). Chromosome 1 exhibited the highest concentration of *PaWRKY* genes, encompassing 16 members, followed by chromosome 5 with 11 members. In stark contrast, chromosome 11 contained only a single *PaWRKY* gene, PaWRKY66.

To further elucidate the evolutionary expansion of the PaWRKY gene family, an analysis of gene duplication events was conducted. MCScanX analysis revealed five pairs of tandemly duplicated genes among the 70 *PaWRKY* members: *PaWRKY4*/*PaWRKY5*, *PaWRKY11*/*PaWRKY12*, *PaWRKY23*/*PaWRKY24*, *PaWRKY37*/*PaWRKY38*, and *PaWRKY54*/*PaWRKY55*. These tandemly duplicated pairs were localized on chromosomes 1, 3, 5, and 9. Additionally, 32 pairs of *PaWRKY*

genes were found to be associated with segmental duplication events (Fig. 2B). Chromosome 1 contained the greatest number of segmental duplication events. On chromosome 5, with the exception of *PaWRKY41*, all segmentally duplicated *PaWRKY* genes participated in two segmental duplication events with other genes. The Ka/Ks ratios for the 32 duplicated gene pairs were all below 0.4, with the exception of PaWRKY23/PaWRKY37, which exhibited a value of 0.43. This suggests that the majority of duplicated *PaWRKY* genes have been subject to strong purifying selection throughout their evolutionary history (Table S3).

To further investigate the evolutionary origins and comparative genomic relationships of the PaWRKY gene family, syntenic analyses were conducted among avocado, *A. thaliana*, *C. camphora*, *O. sativa*, and *A. comosus* using JCVI, a Python-based MCScanX pipeline. In these analyses, 18, 13, and 14 orthologous gene pairs were identified between avocado and *A. thaliana*, *O. sativa*, and *A. comosus*, respectively (Fig. 2C).

Phylogenetic analysis of PaWRKY proteins

To enhance the understanding of the subgroup classification within the PaWRKY gene family and to elucidate the phylogenetic relationships among PaWRKYs, *A. thaliana* WRKYs (AtWRKYs), and *A. comosus* WRKYs (AcWRKYs), a phylogenetic tree was constructed using MEGA11. This analysis was based on the protein sequences of 70 PaWRKYs, 72 AtWRKYs, and 54 AcWRKYs, as depicted in Fig. 3. The findings revealed that the 196 WRKY proteins were categorized into three principal groups: I, II, and III. Group II was the most populous, comprising 128 proteins, while Groups I and III contained 39 and 29 members, respectively. Furthermore, Group II was subdivided into five subgroups: IIa, IIb, IIc, IId, and IIe, consisting of 11, 23, 44, 24, and 23 members, respectively. Notably, subgroups IIa, IIb, and IIc demonstrated closer phylogenetic affinity with Group I, whereas subgroups IId and IIe exhibited closer clustering with Group III.

In the analysis of the 70 PaWRKY proteins, 13 were categorized into Group I, while 6, 8, 16, 10, and 9 were assigned to subgroups IIa through IIe, respectively, and 8 were

classified under Group III (Fig. 4). The proteins in Group I uniformly exhibited two WRKY domains along with a C2H2-type zinc-finger motif characterized by the sequence C–X_4–C–X_{23}–H–X–H. Notably, the domains and zinc-finger motifs within Group I demonstrated greater conservation compared to those in Groups II and III, with no significant variations observed. Members of Group II were characterized by the presence of a single WRKY domain and a C2H2-type zinc-finger motif, denoted as C–$X_{4–5}$–C–$X_{23–24}$–H–X–H. Within Group II, two distinct variants of the WRKY heptapeptide were identified: WRKYGKK in PaWRKY17 from subgroup IIc and WKKYGQK in PaWRKY34 from subgroup IId. Furthermore, variations in the C2H2-type zinc-finger motif were identified in PaWRKY14 from subgroup IIc and PaWRKY52 from subgroup IId. Group III proteins were characterized by a single WRKY domain, with the heptapeptides in PaWRKY8 and PaWRKY39 being replaced by WIKYGLK and WKKYGQK, respectively. Among the Group III members, seven contained a C2HC-type zinc-finger motif (C–X_7–C–X_{23}–H–X–C), while one member exhibited a C2H2-type zinc-finger motif (C–X_7–C–X_{23}–H–X–H) (Fig. 4). In summary, PaWRKY proteins exhibit the characteristic classification features and structural conservation typical of WRKY transcription factors. However, variations observed in the conserved motifs of certain members may suggest potential functional divergence in their roles in transcriptional regulation.

Gene structure, *cis*-acting element, and protein interaction network analyses

The diversity in gene structures can significantly influence the evolutionary trajectory of gene families. In order to elucidate the structural characteristics and potential functional divergence within the PaWRKY family, a comprehensive analysis was conducted focusing on conserved domains, motif composition, and gene structures (Fig. 5). It was observed that all PaWRKY proteins possessed at least one WRKY domain, with Group I members uniquely containing two WRKY domains. Notably, all members of subgroup IId exhibited a Pfam Plant zinc cluster domain in conjunction with the WRKY domain. Furthermore, the majority of PaWRKY proteins within subgroups IIa and IIb were found to possess a coiled-coil region, which may be indicative of specific

functional roles (Fig. 5A). The MEME motif analysis revealed that PaWRKY proteins sharing similar motif compositions were systematically grouped, with all PaWRKY proteins exhibiting WRKY-associated motifs, specifically motif 1 and motif 3. Additionally, motif 7 was exclusively identified in all members of subgroups IIa and IIb (Fig. 5B). The CDS regions within *PaWRKY* genes varied from 2 to 9, with *PaWRKY56* possessing the highest number of CDS regions, totaling 9 (Fig. 5C). The findings suggest that the PaWRKY family exhibits significant conservation within its subfamilies concerning domain composition, conserved motifs, and gene structure. This conservation is likely intricately linked to functional divergence and phylogenetic relationships.

Following structural characterization, an in-depth analysis of *cis*-acting elements was conducted to predict the potential regulatory functions of *PaWRKY* genes. The promoter regions were delineated as the 1,500-bp sequences upstream of the transcription start sites (TSSs) of *PaWRKY* genes and were subsequently analyzed using the PlantCARE database (Fig. 6 and Fig. S2). This analysis identified 53 distinct types of *cis*-acting elements, categorized into four principal groups: light-responsive elements (30 types), phytohormone-responsive elements (8 types), stress-responsive elements (6 types), and elements related to plant growth and development (9 types). Notably, light-responsive elements were the most prevalent, with G-box, I-box, and Box 4 elements identified in the promoters of 42, 36, and 35 *PaWRKY* genes, respectively. Additionally, elements associated with responses to both biotic and abiotic stresses were extensively distributed across *PaWRKY* promoters. Within the category of phytohormone-responsive elements, the promoters of 35 *PaWRKY* genes were found to contain abscisic acid-responsive ABRE elements, as well as MeJA-responsive elements, including the TGACG and CGTCA motifs. Furthermore, the analysis revealed the presence of at least one gibberellin-responsive *cis*-acting element, either the GARE motif or the P-box, within the promoters of 35 *PaWRKY* genes. Promoters of 19 genes contained salicylic acid-responsive TGA elements. Among the abiotic stress-related elements, the ARE motif was the most prevalent, identified in 60 *PaWRKY* genes. MYB-binding sites associated with drought inducibility (MBS) and

low-temperature-responsive elements (LTR) were found in 33 and 25 *PaWRKY* genes, respectively. Regarding elements related to plant growth and development, the CAT-box elements, associated with meristem expression, and the O2-site elements, linked to zein metabolism regulation, were the most common, appearing in 24 and 34 genes, respectively. Collectively, the presence of a diverse array of *cis*-acting elements associated with light signaling, hormone signaling, and stress responses in the promoters of *PaWRKY* genes indicates that these genes are likely to play significant roles in the growth, development, and stress-responsive biological processes of avocado.

To further investigate the potential cooperative relationships among PaWRKY proteins, an interaction network for these proteins was constructed utilizing the STRING online database, with reference to WRKY protein interaction data from *A. thaliana* (Fig. 7). The analysis revealed that 26 proteins were involved in the network, forming a total of 164 predicted interactions. Notably, PaWRKY40, PaWRKY70, and PaWRKY6 exhibited significantly higher connectivity than the network's average, occupying central positions within the interaction framework. These findings imply that these three proteins may serve as critical hubs within the PaWRKY regulatory network, playing pivotal roles in coordinating various biological processes and stress responses.

Transcriptome overview and expression patterns of *PaWRKY* genes across tissues and under biotic stresses

To elucidate the transcriptomic responses of avocado to biotic stresses and to establish a foundation for deducing the potential functions of *PaWRKY* genes under such conditions, RNA-seq was conducted on samples collected from various tissues subjected to different biotic stress treatments. Following quality control and data filtering, a total of 134.84 GB of high-quality clean data was acquired, with mean Q30 values ranging from 94.98% to 95.59%, indicative of superior sequencing quality. PCA demonstrated that biological replicates from identical treatments clustered cohesively, while distinct groups were distinctly separated. The first and second principal components (PC1 and PC2) accounted for 22.75% and 21.51% of the total variance, respectively, underscoring the stability and reproducibility of the transcriptomic data

generated in this study (Fig. S2).

To elucidate gene expression patterns across various tissues and stress conditions, a K-means clustering analysis was conducted on the expressed genes. The optimal number of clusters was determined to be five, as identified using the *cascadeKM* function in the *vegan* package (Fig. S3). These five clusters comprised 7,492, 4,193, 5,747, 6,926, and 5,366 genes, respectively (Fig. 8A). Notably, clusters 2, 4, and 5 exhibited a strong association with biotic stress responses. Genes within cluster 2 were significantly upregulated in insect-bored fruits (IHF). Enrichment analysis revealed that these genes were predominantly linked to plant defense and hormone signaling pathways, including the diterpenoid biosynthetic process (GO:0016102), plant hormone signal transduction (ko04075), and cytochrome P450 (ko00199) (Fig. 8B and Fig. S4). Conversely, genes in cluster 4 were induced in insect-infested leaves (IHL). Functional analysis indicated that these genes were primarily associated with photosynthetic processes and defense responses, encompassing photosynthesis (GO:0015979), chloroplast organization (GO:0009658), photosynthesis, light reaction (GO:0019684), photosynthesis proteins (ko00194), and carbon fixation in photosynthetic organisms (ko00710). Pathways related to hormone signaling and defense mechanisms, such as plant hormone signal transduction (ko04075) and cytochrome P450 (ko00199), exhibited significant enrichment (Fig. 8B and Fig. S4). Within cluster 5, gene activation was observed in diseased leaves (DL), primarily linked to defense mechanisms, hormone responses, and senescence. This includes responses to fungal infection (GO:0009620), regulation of defense responses (GO:0031347), responses to salicylic acid (GO:0009751), responses to jasmonic acid (GO:0009753), plant organ senescence (GO:0090693), abscission (GO:0009838), leaf senescence (GO:0010150), the MAPK signaling pathway in plants (ko04016), and plant hormone signal transduction (ko04075) (Fig. 8B and Fig. S4). Collectively, biotic stresses across various tissues initiated transcriptional responses predominantly centered on hormone signal transduction and defense mechanisms. In diseased leaves, these responses were further associated with processes related to senescence and abscission, indicating that pathogen infection may elicit more pronounced tissue damage responses compared to

insect herbivory.

To further elucidate the transcriptional responses elicited by various biotic stresses, DEGs were identified utilizing the *DESeq2* R package. The analysis revealed a total of 6,743, 7,878, and 6,059 DEGs in IHL, DL, and IHF, respectively. Both insect herbivory and pathogen infection resulted in substantial down-regulation of gene expression in the leaves, with diseased leaves exhibiting the highest number of down-regulated genes (5,034) (Fig. S5). GO and KEGG enrichment analyses of these DEGs indicated significant enrichment in hormone- and defense-related biological processes and pathways (Fig. 9). These include the response to salicylic acid (GO:0009751), regulation of hormone levels (GO:0010817), response to jasmonic acid (GO:0009753), regulation of defense response (GO:0031347), plant hormone signal transduction (ko04075), phenylpropanoid biosynthesis (ko00940), flavonoid biosynthesis (ko00941), cytochrome P450 (ko00199), and alpha-linolenic acid metabolism (ko00592). The findings align with the clustering analysis, suggesting that biotic stresses elicit extensive gene expression responses in avocado, encompassing genes associated with hormone signaling and defense mechanisms.

The expression levels of *PaWRKY* genes were analyzed to assess their expression profiles across various tissues, including flower buds, stems, leaves, and fruits (Fig. 10A). In general, *PaWRKY* genes exhibited distinct tissue-specific expression patterns. A majority of *PaWRKY* genes demonstrated high expression in leaves, whereas only two genes, *PaWRKY18* and *PaWRKY54*, exhibited elevated expression specifically in fruits, implying that other members may predominantly be expressed under stress conditions. Notable expression differences were observed among different subfamilies. Members of Groups I, IId, IIe, and III were highly expressed across multiple tissues, indicating their potential involvement in the regulation of avocado growth and development. Conversely, most members of Groups IIa, IIb, and IIc were predominantly expressed in leaves, suggesting possible functional divergence among the subfamilies. Collectively, these findings indicate that *PaWRKY* genes demonstrate tissue-specific expression patterns, potentially reflecting the diverse functions of various subfamily members during avocado growth and development.

The expression profiles of *PaWRKY* genes under biotic stress conditions were further examined (Fig. 10B). In fruit tissues, IHF significantly upregulated the expression of most *PaWRKY* genes, although a few genes were downregulated. In leaf tissues, DL exhibited a considerably higher number of highly expressed *PaWRKY* genes compared to IHL. Numerous members from Groups I, IIa, IIb, IIc, and III were prominently expressed in DL. Conversely, *PaWRKY* genes induced by insect feeding in leaves were predominantly found in Groups I and IIe, suggesting that these two subfamilies may primarily respond to insect feeding in leaves. Furthermore, the expression levels of *PaWRKY1*, *PaWRKY61*, and *PaWRKY14* in fruits remained unaffected by insect boring. These differential response patterns suggest that members of distinct PaWRKY subfamilies may play specific regulatory roles in avocado responses to disease infection and insect feeding.

WGCNA identifies stress-associated co-expression modules

To elucidate gene co-expression patterns under various biotic stress conditions, WGCNA was conducted utilizing the expression profiles of 46,892 annotated avocado genes. The clustering of samples revealed that all samples grouped appropriately according to their respective categories, with no outliers detected (Fig. S6A). To achieve an approximately scale-free network topology, a soft-thresholding power analysis was executed using the *WGCNA* package in R, resulting in the selection of $\beta = 16$ for subsequent network construction (Fig. S6B, C). Ultimately, 28 color-coded modules were identified (Fig. 11A, B).

The analysis of module–trait correlations indicated that distinct tissues and treatments were associated with specific characteristic modules. In CKL, the modules that exhibited significant correlations included black, darkolivegreen, lightcyan, lightgreen, lightyellow, orange, and salmon2 ($p < 0.05$), encompassing a total of 5,800 genes. The genes were predominantly associated with processes related to growth, development, and hormonal regulation, including the regulation of post-embryonic development (GO:0048580), plant epidermis morphogenesis (GO:0090626), photoperiodism (GO:0009648), response to jasmonic acid (GO:0009753), and response

to ethylene (GO:0009723) (Fig. 11C, D). In IHL, the modules that exhibited significant correlations included black, brown4, coral1, darkgrey, darkmagenta, plum2, and skyblue3 ($p < 0.05$), encompassing 5,917 genes. These genes were primarily enriched in processes related to photosynthesis and defense, such as photosynthesis (GO:0015979), carbohydrate biosynthetic process (GO:0016051), chloroplast organization (GO:0009658), and photosynthesis proteins (ko00194). Compared to CKL, the IHL-associated modules contained a greater number of genes enriched in photosynthesis-related terms, while the number of genes associated with growth and development-related terms was significantly reduced (Fig. 11C, D). In DL, the modules that showed significant associations were brown4, skyblue3, cyan, lavenderblush3, lightpink4, maroon, and saddlebrown ($p < 0.05$), comprising 3,827 genes. The genes were predominantly enriched in several biological processes and pathways, including response to fungus (GO:0009620), defense response to fungus (GO:0050832), plant organ senescence (GO:0090693), leaf senescence (GO:0010150), MAPK signaling pathway in plants (ko04016), and the metabolism of xenobiotics by cytochrome P450 (ko00980) (Fig. 11C, D). These findings were consistent with the results obtained from K-means clustering analysis. In CKF, the modules that exhibited significant correlations were darkolivegreen, skyblue3, antiquewhite4, coral2, lightsteelblue1, darkgreen, and skyblue2 ($p < 0.05$), encompassing a total of 5,781 genes. These genes were primarily enriched in processes such as meristem development (GO:0048507), meristem maintenance (GO:0010073), regulation of developmental growth (GO:0048638), and ribosome (ko03011). Conversely, the modules associated with IHF, including blue2, darkolivegreen, darkseagreen4, and skyblue1 ($p < 0.05$), comprised only 1,365 genes. The majority of enriched developmental processes in IHF closely resembled those observed in CKF; however, there was a notable reduction in the number of associated genes. Furthermore, stress-related processes, such as the cellular response to acidic chemicals and the response to wounding, were enriched within the IHF-associated modules (Fig. 11C, D). In general, the genes within modules significantly associated with healthy leaves and fruits predominantly pertained to growth and development, hormone responses, and developmental regulation.

Following insect herbivory and pathogen infection, modules linked to defense responses, hormone signal transduction, and stress regulation were activated, with disease infection further prompting the senescence of plant organs.

Integrative analysis reveals key *PaWRKY* genes involved in biotic stress responses

To further elucidate the key *PaWRKY* genes responsive to various biotic stresses, an UpSet integrative analysis was conducted utilizing stress-associated K-means clusters, significantly correlated WGCNA modules, DEG sets, and the *PaWRKY* gene set in IHL, DL, and IHF (Fig. 12).

In the context of IHL, 30 *PaWRKY* genes were exclusively present within the WRKY set, while the remaining 40 *PaWRKY* genes exhibited overlap with at least one stress-associated gene set. Notably, 25 *PaWRKY* genes coincided with the IHL_DOWN set, implying that insect feeding may transcriptionally repress the expression of certain *PaWRKY* genes. Specifically, *PaWRKY14*, *PaWRKY40*, *PaWRKY50*, and *PaWRKY56* were concurrently located at the intersection of the DEG_UP, WGCNA, and K-means stress-associated sets, suggesting that these genes may serve as critical core regulators in the response of avocado leaves to insect feeding (Fig. 12A).

In the context of DL, a total of 20 *PaWRKY* genes were identified exclusively within the WRKY set, while the remaining 50 *PaWRKY* genes exhibited overlap with disease-associated gene sets. This observation suggests that pathogen infection triggers the activation of a broader array of PaWRKY family members. Notably, sixteen *PaWRKY* genes, namely *PaWRKY4*, *PaWRKY5*, *PaWRKY9*, *PaWRKY12*, *PaWRKY17*, *PaWRKY24*, *PaWRKY25*, *PaWRKY34*, *PaWRKY35*, *PaWRKY38*, *PaWRKY44*, *PaWRKY46*, *PaWRKY57*, *PaWRKY64*, *PaWRKY65*, and *PaWRKY69*, were concurrently located at the intersection of the DEG_UP, WGCNA, and K-means stress-associated sets. These genes potentially serve as core regulatory factors in the avocado leaf's response to disease stress (Fig. 12B).

Conversely, in the IHF condition, 35 *PaWRKY* genes were exclusively found in the WRKY set, with an additional 35 *PaWRKY* genes overlapping with other stress-associated sets. Although 28 *PaWRKY* genes coincided with the IHF_UP set, suggesting

that insect boring may induce the up-regulation of certain *PaWRKY* genes, only *PaWRKY42* was concurrently located at the intersection of the DEG_UP, WGCNA, and K-means stress-associated sets. Furthermore, *PaWRKY26* and *PaWRKY48* were simultaneously present in both the DEG_UP and K-means sets (Fig. 12C). The findings suggest that insect-bored fruits possess a reduced number of core candidate *PaWRKY* genes that meet multiple screening criteria when compared to leaf stresses. Several *PaWRKY* genes, identified as having potential core regulatory roles under various biotic stresses, were selected for validation through qRT-PCR. The expression trends observed were largely consistent with the RNA-seq results (Fig. 13).

Collectively, these findings demonstrate that *PaWRKY* genes exhibit distinct response patterns under varying biotic stress conditions. Notably, diseased leaves contained the highest number of *PaWRKY* genes commonly identified across multiple stress-associated analyses, indicating that the *PaWRKY* gene family exhibits differential responses to diverse types of biotic stress.

Prediction of three-dimensional structures of core stress-responsive PaWRKY proteins

Protein function is intricately linked to its three-dimensional conformation [52]. To further elucidate the structural characteristics of the 21 core stress-responsive PaWRKY proteins identified under various biotic stress conditions, homology modeling was conducted using SWISS-MODEL based on their amino acid sequences. Notably, PaWRKY25 was excluded from this analysis due to a maximum sequence identity of only 66.67% among all predicted models, which did not satisfy the established screening criterion. Consequently, 20 structural models with sequence identities $\geq 90\%$ and high GMQE scores were retained for subsequent examination (Fig. 14). The predicted structures revealed that these core stress-responsive PaWRKY proteins exhibited some variation in their overall spatial conformations; however, all retained a distinct conserved WRKY domain. Most candidate proteins contained the canonical WRKYGQK heptapeptide, whereas PaWRKY17 and PaWRKY34 possessed the variant motifs WRKYGKK and WKKYGQK, respectively. Beyond the conserved

WRKY core domain, one or two α-helical secondary structures were predicted in the peripheral regions of most PaWRKY proteins, potentially indicating local structural variations associated with functional divergence. In summary, the fundamental stress-responsive PaWRKY proteins exhibited structural variability in their three-dimensional conformations, while the spatial configuration of the conserved WRKY domain remained remarkably consistent. Localized structural variations, particularly within α-helical regions, may offer insights into the functional diversification of PaWRKY proteins in the context of avocado responses to biotic stresses.

Discussion

The WRKY family constitutes a substantial group of plant-specific transcription factors extensively implicated in plant growth and development, stress responses, and hormone regulation. Notably, under biotic stress conditions, WRKY transcription factors frequently serve as critical regulatory nodes that connect pathogen- or insect-derived signals to the activation of downstream defense mechanisms. These factors are involved in modulating hormone signaling, defense responses, and secondary metabolic reprogramming, thereby affecting plant resistance to pathogens and herbivorous insects [12]. Avocado, a crop of significant economic value in tropical and subtropical regions, is often challenged by a variety of diseases and insect pests. Nonetheless, research on the WRKY family in avocado, particularly in the predominant cultivar Hass, and its response mechanisms under various biotic stresses, remains sparse. Consequently, this study systematically identified the WRKY gene family in Hass avocado, examined its expression profiles under different biotic stress conditions, and identified key *WRKY* genes with potential regulatory functions. These findings lay the groundwork for the subsequent identification of candidate genes linked to disease and insect resistance, as well as for the functional characterization of their roles in avocado stress responses.

The number of WRKY gene family members exhibits significant variability across different crop species, ranging from 34 to 238 [53], suggesting that the WRKY family has likely experienced frequent gene loss and duplication events throughout evolution. In this study, 70 *PaWRKY* genes were identified in Hass avocado, a figure comparable

to those reported in other Lauraceae species, such as *Litsea cubeba* (64) [54], *Phoebe bournei* (60) [55], and *C. camphora* (72) [56]. Notably, another study reported only 60 WRKY members in *C. camphora* [57]. Similarly, 82 *WRKY* genes were previously identified based on the genome of a West Indian avocado accession [32]. These discrepancies suggest that cultivar background and genome assembly versions can significantly impact the number of gene family members identified. Furthermore, PaWRKY proteins exhibited considerable variation in protein length, molecular weight, and isoelectric point. The majority of PaWRKY proteins are predicted to localize within the nucleus, aligning with observations from numerous other plant species [7, 42, 58] and underscoring the conventional role of WRKY transcription factors as nuclear transcriptional regulators. Nonetheless, a minority of PaWRKY proteins are anticipated to localize to chloroplasts or the cytoskeleton, indicating a degree of functional divergence within the PaWRKY family.

The chromosomal distribution of *PaWRKY* genes is notably uneven, with the highest concentration on chromosome 1, while relatively few members are dispersed across several other chromosomes (Fig. 2A). Certain *PaWRKY* genes are found in clusters on chromosomes. Comparable patterns of uneven distribution have been documented in the WRKY families of various plant species, such as soybean [59], barley [60], and *Mikania micrantha* [61], potentially linked to gene duplication events and chromosomal structural rearrangements [62, 63]. Further intraspecific collinearity analysis identified multiple homologous gene pairs among *PaWRKY* genes, comprising five tandemly duplicated pairs and 32 segmentally duplicated pairs. Chromosome 1 exhibited the highest number of segmental duplication events, suggesting significant gene duplication activity on this chromosome (Fig. 2B). Gene expansion events play a crucial role in driving genomic diversity during the evolution of complex plant genomes. Segmental duplication is widely recognized as a significant mechanism contributing to the expansion of plant gene families, particularly in species that have undergone whole-genome duplication [64, 65]. The Ka/Ks ratios for all 32 duplicated gene pairs were less than 1 and below 0.5 (Table S3), indicating that these duplicated genes have been subject to purifying selection throughout evolution. Similarly, the Ka/Ks ratios of

duplicated *WRKY* gene pairs in various plant species are generally less than 1, suggesting that the WRKY gene family has remained relatively conserved over evolutionary time, with minimal nonsynonymous substitutions, thereby preserving protein functional stability [66]. A comparative synteny analysis across different species was also conducted to investigate the evolutionary characteristics of PaWRKY genes. A comparative analysis revealed the identification of only 18, 13, and 14 orthologous *WRKY* gene pairs between avocado and *Arabidopsis*, rice, and pineapple, respectively (Fig. 2C). These numbers are significantly lower than those reported for other Lauraceae species and related taxa [54-56]. It is noteworthy that despite the evident chromosome-level synteny between avocado and *C. camphora*, both members of the Lauraceae family, no orthologous *WRKY* gene pairs were detected between these two species. This observation suggests that *PaWRKY* genes may have undergone a relatively conserved yet lineage-specific evolutionary trajectory, potentially forming or diverging subsequent to the divergence of avocado and *C. camphora* from their last common ancestor.

Phylogenetic analysis revealed that all PaWRKY proteins can be categorized into the classical Groups I, II, and III, with Group II further subdivided into five subgroups: IIa, IIb, IIc, IId, and IIe. Among these, Group II encompasses the largest number of members (Fig. 3), aligning with the classification observed in West Indian avocado and other plant species [32]. Conserved domain analysis indicated that WRKY proteins generally possess one or two conserved WRKYGQK motifs and zinc-finger structures (Fig. 4). All Group I PaWRKY proteins maintained two complete WRKY domains and a C2H2 zinc-finger motif, whereas several members of Groups IIc, IId, and III exhibited variations in the WRKY heptapeptide or zinc-finger motif. Three types of WRKY heptapeptide variants were identified in PaWRKY proteins: the WKKYGQK variant was identified in Group IId (PaWRKY34) and Group III (PaWRKY39), the WRKYGKK variant was detected in Group IIc (PaWRKY17), and the WIKYGLK variant was present in Group III (PaWRKY8). It is noteworthy that the WRKYGKK variant appears to be relatively prevalent among Group IIc WRKY members across different plant species. For instance, four WRKYGKK variants have been identified in

Group IIc of *Caragana korshinskii* [67], two in Group IIc of *Helianthus annuus* L. [68], and two in Group IIc of *M. micrantha* [61]. Consequently, while the conserved WRKYGQK motif facilitates the recognition and binding of WRKY proteins to W-box cis-acting elements, its variation may influence the binding specificity and regulatory activity of WRKY proteins with respect to downstream target genes [16]. Further functional characterization of WRKY proteins with these variant motifs will enhance our understanding of the regulatory mechanisms underlying WRKY-mediated stress responses [69, 70].

Conserved protein domains, motif composition, and gene structure are not only fundamental for classification and evolutionary analysis but are also intricately linked to potential protein functions [71]. In this study, members of different PaWRKY subfamilies exhibited similar domain architectures, motif compositions, and CDS organization patterns (Fig. 5). Members of Group I typically exhibit two WRKY domains, whereas most members of Groups IIa and IIb are characterized by the presence of coiled-coil regions. In contrast, all members of Group IId possess a Plant zinc cluster domain in conjunction with the WRKY domain. Prior research has demonstrated that coiled-coil or leucine-zipper domains within WRKY proteins are implicated in protein–protein interactions and may play a role in the formation of transcriptional complexes. Meanwhile, the unique structural attributes of Group IId members may affect target gene selection and transcriptional regulation [72, 73]. Furthermore, members within the same subgroup often display similar motif compositions and exon number ranges. Notably, high structural conservation among WRKY subfamilies has been observed in pineapple, *Verbena bonariensis*, soybean, and various other plant species [42, 74, 75]. These observations imply that distinct PaWRKY subfamilies may have experienced unique evolutionary pathways, while preserving high structural conservation and showing potential for functional divergence [76].

Plant promoters serve as crucial *cis*-regulatory regions governing gene expression. These regions encompass numerous cis-acting elements that are recognized by transcription factors, thereby influencing gene expression patterns across various

tissues, developmental stages, and environmental conditions. Consequently, promoter elements are frequently utilized to predict potential gene functions and regulatory patterns [77, 78]. In this study, we identified a widespread distribution of light-responsive, hormone-responsive, and stress-responsive elements within PaWRKY promoters. These include a substantial number of cis-acting elements associated with gibberellic acid (GA), abscisic acid (ABA), jasmonic acid (JA), and salicylic acid (SA) signaling pathways, as well as responses to drought, low temperature, and hypoxia (Fig. 6 and Fig. S1). These findings imply that the PaWRKY family may play a significant role in biological processes related to growth and development, hormone signaling, and environmental responses in Hass avocado. Further insights into the potential regulatory functions of PaWRKY proteins were obtained by constructing a protein interaction network using the STRING database, which predicted that 26 PaWRKY proteins are involved in 164 interaction relationships (Fig. 7). In plants, WRKY proteins are capable of interacting not only with other WRKY proteins but also with VQ proteins, MAPK cascade components, and various other regulatory proteins. This interaction contributes to signal amplification, interaction specificity, and finely-tuned transcriptional regulation [79-81]. Within the predicted PaWRKY interaction network, PaWRKY40, PaWRKY70, and PaWRKY6 serve as central hubs, suggesting that these proteins may play pivotal roles in coordinating hormone signaling, growth and development, and stress responses in Hass avocado.

The availability of high-quality RNA-seq data and well-clustered biological replicates provided a robust foundation for subsequent analyses of stress responses in this study. The combined outcomes of K-means clustering (Fig. 8 and Fig. S4) and differential expression gene (DEG) analysis (Fig. S5 and Fig. 9) revealed that Hass avocado exhibits distinct defense responses to various biotic stresses, characterized by clear tissue- and stress-specific patterns. Similar observations have been documented in comparative transcriptomic analyses of barley, where DEGs induced by fungal infection and aphid feeding displayed significant differences. A mere 1.1% of DEGs were common between datasets associated with fungal and aphid interactions, while only 0.1% of DEGs were shared across all datasets. Furthermore, fungal infection

elicited a significantly greater number of DEGs compared to aphid feeding, corroborating our previous findings [82]. In this study, biotic stressors in various tissues initiated transcriptional responses primarily focused on hormone signal transduction and defense mechanisms. In diseased leaves, these responses were further linked to processes related to senescence and abscission, implying that pathogen infection may provoke more pronounced tissue damage responses than insect feeding. The regulation of specific tissue or organ developmental processes, such as senescence, growth arrest, and abscission, induced by pathogens, may either reflect a pathogenic strategy or an active plant defense mechanism aimed at limiting pathogen spread through the shedding of infected leaves [83]. For instance, severe infection by *Pseudomonas syringae* can trigger leaf abscission in *Arabidopsis*, and similar abscission has been noted in tomato following infection with *P. syringae* and powdery mildew [84]. Consequently, pathogen infection in Hass avocado leaves may elicit a more robust defense mechanism compared to insect feeding, transitioning from defense activation to tissue damage and ultimately leading to abscission.

The *PaWRKY* genes demonstrated distinct tissue-specific expression patterns (Fig. 10A) and exhibited significant variations in induction under different biotic stresses (Fig. 10B). The majority of highly expressed *PaWRKY* genes were predominantly found in leaves, aligning with the tissue-specific expression patterns of *WRKY* genes observed in other plant species, such as *H. annuus* L. [68], cotton [85], and mulberry [86]. Leaves serve as the primary organs for photosynthesis, material exchange, and environmental perception, possessing the largest interface with the external environment. Consequently, they are crucial sites for the initial detection of environmental changes and stress responses [87]. Therefore, the elevated expression of PaWRKY genes in leaves is likely closely linked to their functions in leaf development, senescence, and stress responses. In contrast, in fruits, only *PaWRKY18* and *PaWRKY54* exhibited high expression specific to fruit tissue, indicating that these two genes may play significant roles in fruit development (Fig. 10A). Furthermore, the majority of *PaWRKY* genes were up-regulated in Hass avocado fruits infested by gall midges, suggesting that these genes are actively involved in the response to insect

feeding stress. Conversely, a subset of *PaWRKY* genes exhibited down-regulation following insect infestation. In other plant species, certain WRKY transcription factors have been identified as negative regulators of insect resistance. For instance, the knockout of *OsWRKY36* in rice has been shown to significantly enhance resistance to planthoppers [88], whereas the knockout of *AtWRKY22* in *Arabidopsis* has been associated with a marked reduction in aphid feeding performance [89]. This implies that gall midge infestation may not only activate stress-responsive *PaWRKY* genes but also suppress those *PaWRKY* members that negatively regulate insect resistance. Consistent with the findings from K-means clustering and differential expression analyses, pathogen infection in leaves (DL) resulted in the up-regulation of a greater number of *PaWRKY* genes compared to insect feeding. In the context of plant defense against fungal pathogens, WRKY transcription factors play a crucial role by enhancing pattern- and effector-triggered immunity. This is achieved through the activation of *pathogenesis-related* (*PR*) gene expression, reinforcement of the cell wall, RNA interference, and the induction of programmed cell death to thwart pathogen invasion [12]. Therefore, pathogen infection in Hass avocado may trigger the activation of a broader array of *WRKY* genes, thereby eliciting more robust immune responses against pathogens. A relatively large number of members from Groups I, IIa, IIb, IIc, and III exhibited high expression levels in DL (Fig. 10B). In contrast, *PaWRKY* genes that were significantly upregulated by insect feeding in leaves (IHL) were predominantly found in Groups I and IIe. Similarly, Group I WRKY members, such as *NaWRKY3*/*NaWRKY6* in *Nicotiana attenuata* and *OsWRKY70* in rice, have been implicated in JA-mediated anti-herbivore defense [90, 91]. These observations suggest that Groups I and IIe of the PaWRKY family may play crucial roles in the defense response of Hass avocado to insect predation.

To further elucidate the transcriptional regulatory patterns of Hass avocado across various tissues and under different stress conditions at the co-expression network level, and to identify key *PaWRKY* genes, WGCNA was conducted utilizing all annotated avocado genes (Fig. 11). In this study, modules that demonstrated significant associations with healthy leaves and fruits were predominantly enriched in pathways

related to growth and development, hormone responses, and developmental regulation. Conversely, modules significantly associated with insect feeding and pathogen infection exhibited enrichment in defense response pathways and hormone signal transduction pathways (Fig. 11C, D). This observed transition from growth to defense aligns with the classical growth-defense trade-off observed in plants [92, 93]. Furthermore, in comparison to modules associated with insect feeding, those associated with disease-related stress (DL) were significantly enriched in pathways pertaining to fungal response, MAPK signaling, cytochrome P450 metabolism, leaf senescence, and plant organ senescence. This pattern aligns with the K-means clustering and DEG analyses previously discussed, further indicating that pathogen infection may more strongly induce a shift from defense activation to tissue damage and organ senescence compared to insect feeding [83]. Building on these analyses, an UpSet analysis was conducted by integrating K-means clusters, WGCNA-associated modules, DEG sets, and the *PaWRKY* gene set (Fig. 12). In total, 16, 4, and 1 key *PaWRKY* genes were identified in DL, IHL, and IHF, respectively, which were concurrently present in DEG_UP, WGCNA-associated modules, and the corresponding K-means clusters. The significant variation in the number of core candidate *PaWRKY* genes supported by multiple analyses under different biotic stresses underscores the critical role of *PaWRKY* genes in avocado leaf spot defense. In *Malus domestica*, the *MdVQ37–MdWRKY100* complex influences resistance to Glomerella leaf spot by modulating SA accumulation and *MdRPM1* expression [94]. In *Populus nigra*, *PsnWRKY70* enhances resistance to leaf blight caused by *Alternaria alternata* by activating MAPK cascade genes, calcium signaling-related genes, and other immune-related genes [95]. The up-regulation of WRKY transcription factors has been documented in mung bean (*Vigna radiata*) following infection with Cercospora leaf spot disease [96]. This suggests that WRKY transcription factors may have a broadly conserved regulatory function in plant defense mechanisms against leaf spot diseases. The qRT-PCR results generally aligned with the RNA-seq expression patterns (Fig. 13), thereby corroborating the reliability of the key *PaWRKY* genes identified in this research. These candidate genes represent significant targets for future functional validation and the development of regulatory

networks related to disease and insect resistance in Hass avocado.

Homology modeling was conducted for the 21 key candidate PaWRKY proteins identified in this study, resulting in the retention of 20 structural models after filtering (Fig. 14). Beyond the central WRKY domain, one or two α-helical secondary structures were predicted in the peripheral regions of most PaWRKY proteins. Prior research has demonstrated that the functionality of WRKY proteins is not exclusively dictated by the core WRKY domain. Instead, peripheral regions or distinct structural elements can also engage in regulatory interactions, exemplified by the interaction between WRKY proteins and their VQ protein partner, SIB1 [97]. Consequently, these α-helical secondary structures may offer a crucial structural foundation for the functional differentiation of various PaWRKY proteins in response to disease or insect stress..

Conclusion

The Hass avocado is currently the leading commercial cultivar of avocado. In this study, 70 members of the WRKY gene family were systematically identified in Hass avocado and categorized into three primary groups and five Group II subgroups based on phylogenetic relationships and structural characteristics. Segmental duplication was determined to be the principal mechanism driving the expansion of the PaWRKY gene family, with most duplicated gene pairs appearing to have undergone purifying selection throughout evolution. Members of different PaWRKY subfamilies exhibited similar gene structures, conserved motifs, and patterns of domain composition and distribution. Transcriptome analysis further revealed that *PaWRKY* genes demonstrated distinct tissue-specific expression patterns and were differentially expressed in response to various biotic stresses. In leaves, pathogen infection elicited more extensive and robust responses than insect feeding, particularly in relation to tissue damage and organ senescence. Through the integration of WGCNA, K-means clustering, DEG analysis, and expression profiling, several core candidate *PaWRKY* genes potentially involved in biotic stress responses were identified. This study comprehensively elucidates the family composition, evolutionary characteristics, and expression responses of *PaWRKY* genes across various tissues and under biotic stress conditions. It offers valuable

candidate resources for future research on the molecular mechanisms underlying disease and insect resistance in Hass avocado.

List of abbreviations

ABA: Abscisic acid

ABRE: ABA-responsive element

AcWRKY: *Ananas comosus* WRKY

ARE: Anaerobic-responsive element

AtWRKY: *Arabidopsis thaliana* WRKY

BLASTP: Basic Local Alignment Search Tool for proteins

CDD: Conserved Domain Database

CDS: Coding sequence

CKB: Healthy bud

CKF: Healthy fruit

CKL: Healthy leaf

CKS: Healthy stem

DEG: Differentially expressed gene

DL: Diseased leaf

FPKM: Fragments per kilobase of transcript per million mapped reads

GA: Gibberellin

GMQE: Global Model Quality Estimate

GO: Gene Ontology

HMM: Hidden Markov model

IHF: Insect-bored fruit

IHL: Insect-infested leaf

JA: Jasmonic acid

KEGG: Kyoto Encyclopedia of Genes and Genomes

KO: KEGG Orthology

Ka: Nonsynonymous substitution rate

Ks: Synonymous substitution rate

LTR: Low-temperature-responsive element

MAPK: Mitogen-activated protein kinase

MBS: MYB-binding site involved in drought inducibility

MeJA: Methyl jasmonate

MW: Molecular weight

NJ: Neighbor-joining

PaWRKY: *Persea americana* WRKY

PCA: Principal component analysis

pI: Isoelectric point

PR: Pathogenesis-related

qRT-PCR: Quantitative real-time PCR

RNA-seq: RNA sequencing

SA: Salicylic acid

SMART: Simple Modular Architecture Research Tool

STRING: Search Tool for the Retrieval of Interacting Genes/Proteins

TAIR: The *Arabidopsis* Information Resource

TF: Transcription factor

TOM: Topological overlap matrix

TSS: Transcription start site

WGCNA: Weighted gene co-expression network analysis

Supplementary Information

Supplementary Material 1: **Fig. S1** Distribution of *cis*-acting elements identified in the promoter regions of Hass avocado *PaWRKY* genes. **Fig. S2** Principal component analysis of transcriptome samples from different tissues and biotic stress treatments. **Fig. S3** Determination of the optimal number of clusters for K-means clustering analysis. **Fig. S4** KEGG enrichment analysis of genes in different K-means clusters. **Fig. S5** Numbers of differentially expressed genes under different biotic stresses. **Fig. S6** Sample clustering and soft-thresholding power selection for WGCNA.

Supplementary Material 2: **Table S1** Primers used for qRT-PCR.

Supplementary Material 3: **Table S2** Summary information of *PaWRKY* genes in Hass avocado.

Supplementary Material 4: **Table S3** Ka, Ks, and Ka/Ks values of duplicated *PaWRKY* gene pairs.

Authors' contributions

XW and ZS contributed equally to this work. XW conceived and designed the study, performed bioinformatics analyses, prepared figures, interpreted the data, and was a major contributor in writing the manuscript. ZS contributed to sample collection, experimental validation, data curation, formal analysis, visualization, and manuscript revision. BY, LT, ZW, HA, and GZ participated in field investigation, sample collection, and resource preparation. YZ contributed to experimental validation and data curation. LX contributed to formal analysis, software, and visualization. YZ and JW supervised the study, provided resources, acquired funding, administered the project, and revised the manuscript. JW contributed to the conceptualization of the study. All authors read and approved the final manuscript. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

We thank Miss Yanfang Mo and Mr. Fazhong Yang from the Avocado Science and Technology Research Centre of Menglian County, Pu'er, for their assistance with sample collection.

Funding

This research was supported by the Funds of Sci-Tech Innovation System Construction for Tropical Crops of Yunnan Province (ZY7 and 727-8) and the Key Research and Development Projects of Yunnan Province (202603AS090002).

Data availability

Upon publication, all raw sequencing data described in this study will be available via the NCBI Sequencing Read Archive (SRA) through BioProject ID PRJNA1476890 and

BioSample accessions SAMN60735718-SAMN60735738.

Declarations

Ethics approval and consent to participate

Not applicable. All methods were carried out according to relevant guidelines and regulations.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Yunnan Institute of Tropical Crops, Jinghong 666100, China

[2]Avocado Science and Technology Research Centre of Menglian County, Puer 665899, China

References

1. Jones J, Dangl J. The plant immune system. Nature 2006, 444(7117):323-9. https://doi.org/10.1038/nature05286.
2. Feng M, Augstein F, Kareem A, Melnyk CW. Plant grafting: Molecular mechanisms and applications. Mol. Plant 2024, 17(1):75-91. https://doi.org/10.1016/j.molp.2023.12.006.
3. Cheng M, Liao P, Kuo W, Lin TP. The *Arabidopsis ETHYLENE RESPONSE FACTOR1* regulates abiotic stress-responsive gene expression by binding to different *cis*-acting elements in response to different stress signals. Plant Physiol. 2013, 162(3):1566-82. https://doi.org/10.1104/pp.113.221911.
4. Fujita M, Fujita Y, Noutoshi Y, Takahashi F, Narusaka Y, Yamaguchi S, Kazuko, Shinozaki K. Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. Curr. Opin. Plant Biol. 2006, 9(4):436-42. https://doi.org/10.1016/j.pbi.2006.05.014.
5. Chen X, Zhang T, Wang H, Zhao W, Guo Z. Transcription factor WRKY complexes in plant

signaling pathways. Phytopathol. Res. 2025, 7(1):54. https://doi.org/10.1186/s42483-025-00349-x.

6. Liu G, Zhang D, Zhao T, Yang H, Jiang J, Li J, Zhang H, Xu X. Genome-wide analysis of the WRKY gene family unveil evolutionary history and expression characteristics in tomato and its wild relatives. Front. Genet. 2022, 13:962975. https://doi.org/10.3389/fgene.2022.962975.

7. Li S, Ji Q, An X, Chen C, Luo X, Liu T, Zou L. Genome-wide analysis of WRKY gene family and the dynamic responses of key WRKY genes involved in cadmium stress in *Brassica juncea*. Front. Plant Sci. 2024, 15:1465905. https://doi.org/10.3389/fpls.2024.1465905.

8. Dong T, Su J, Li H, Du Y, Wang Y, Chen P, Duan H. Genome-wide identification of the WRKY gene family in four cotton varieties and the positive role of G*hWRKY31* in response to salt and drought stress. Plants 2024, 13(13):1814. https://doi.org/10.3389/fpls.2024.1465905.

9. Song Y, Cui H, Shi Y, Xue J, Ji C, Zhang C, Yuan L, Li R. Genome-wide identification and functional characterization of the *Camelina sativa* WRKY gene family in response to abiotic stress. BMC Genomics 2020, 21(1):786. https://doi.org/10.1186/s12864-020-07189-3.

10. Eulgem T, Rushton PJ, Robatzek S, Somssich IE. The WRKY superfamily of plant transcription factors. Trends Plant Sci. 2000, 5(5):199-206. https://doi.org/10.1016/S1360-1385(00)01600-9.

11. Ma Z, Hu L. WRKY transcription factor responses and tolerance to abiotic stresses in plants. Int. J. Mol. Sci. 2024, 25(13):6845. https://doi.org/10.3390/ijms25136845.

12. Wang D, Zhang R, Zou W, Zhang Y, Zhao W, Sun T, Wu Q, Fu Z, Que Y. WRKY transcription factors: integral regulators of defence responses to biotic stress in crops. Plant Biotechnol. J 2026, 24(5):1-17. https://doi.org/10.1111/pbi.70542.

13. Phukan UJ, Jeena GS, Shukla RK. WRKY transcription factors: molecular regulation and stress responses in plants. Front. Plant Sci. 2016, 7:760. https://doi.org/10.3389/fpls.2016.00760.

14. Wani SH, Anand S, Singh B, Bohra A, Joshi R. WRKY transcription factors and plant

defense responses: latest discoveries and future prospects. Plant Cell Rep. 2021, 40(7):1071-85. https://doi.org/10.1007/s00299-021-02691-8.

15. Finatto T, Viana VE, Woyann LG, Busanello C, Maia LCD, Oliveira AC. Can WRKY transcription factors help plants to overcome environmental challenges? Genet. Mol. Biol. 2018, 41(3):533-44. http://dx.doi.org/10.1590/1678-4685-GMB-2017-0232.

16. Wang H, Chen W, Xu Z, Chen M, Yu D. Functions of WRKYs in plant growth and development. Trends Plant Sci. 2023, 28(6):630-45. https://doi.org/10.1016/j.tplants.2022.12.012.

17. Mi X, Li W, Chen C, Xu H, Wang G, Jin X, Zhang D, Guo W. GhMPK9-GhRAF39_1-GhWRKY40a regulates the GhERF1b- and GhABF2-mediated pathways to increase cotton disease resistance. Adv. Sci. 2024, 11(29):e2404400. https://doi.org/10.1002/advs.202404400.

18. Wang L, Guo D, Zhao G, Wang J, Zhang S, Wang C, Guo X. Group IIc WRKY transcription factors regulate cotton resistance to *Fusarium oxysporum* by promoting GhMKK2-mediated flavonoid biosynthesis. New Phytol. 2022, 236(1):249-65. https://doi.org/10.1111/nph.18329.

19. Ma J, Wei L, Huang K, Wang D, Gao J, Chen X, Guo H, Gao S, Zhang M, Li S et al. Biosynthesis of sakuranetin regulated by OsMPK6-OsWRKY67-OsNOMT cascade enhances resistance to false smut disease. New Phytol. 2025, 245(3):1216-31. https://doi.org/10.1111/nph.20308.

20. Wang Y, Wang X, Fang J, Yin W, Yan X, Tu M, Liu H, Zhang Z, Li Z, Gao M et al. VqWRKY56 interacts with VqbZIPC22 in grapevine to promote proanthocyanidin biosynthesis and increase resistance to powdery mildew. New Phytol. 2023, 237(5):1856-75. https://doi.org/10.1111/nph.18688.

21. Xie Q, Dong W, Wang M, Wang J, Sun L, Liu Z, Gao C, Cao C. BpWRKY6 regulates insect resistance by affecting jasmonic acid and terpenoid synthesis in *Betula platyphylla*. Plant Biotechnol. J 2025, 23(9):3682-96. https://doi.org/10.1111/pbi.70169.

22. Jiang Q, Ding C, Feng L, Wu Z, Liu Y, He L, Liu C, Wang L, Zeng J, Huang J et al. Two leucine-rich repeat receptor-like kinases initiate herbivory defense responses in tea plants. Hortic. Res. 2024, 12(1):uhae281. https://doi.org/10.1093/hr/uhae281.

23. Cheng W, Wang N, Li Y, Zhou X, Bai X, Liu L, Ma X, Wang S, Li X, Gong B et al. CaWRKY01-10 and CaWRKY08-4 confer pepper's resistance to *Phytophthora capsici* infection by directly activating a cluster of defense-related genes. J. Agr. Food. Chem. 2024, 72(20):11682-93. https://doi.org/10.1021/acs.jafc.4c01024.

24. Jiang Y, Zheng W, Li J, Liu P, Zhong K, Jin P, Xu M, Yang J, Chen J. NbWRKY40 positively regulates the response of *Nicotiana benthamiana* to tomato mosaic virus via salicylic acid signaling. Front. Plant Sci. 2021, 11:603518. https://doi.org/10.3389/fpls.2020.603518.

25. Tesfaye T, Ayele M, Gibril M, Ferede E, Limeneh DY, Kong F. Beneficiation of avocado processing industry by-product: A review on future prospect. Curr. Res. Green Sustain. Chem. 2022, 5:100253. https://doi.org/10.1016/j.crgsc.2021.100253.

26. da Silva GG, Pimenta LPS, Melo JOF, Mendonça HOP, Augusti R, Takahashi JA. Phytochemicals of avocado residues as potential acetylcholinesterase inhibitors, antioxidants, and neuroprotective agents. Molecules 2022, 27(6):1892. https://doi.org/10.3390/molecules27061892.

27. Dantas D, Pasquali MA, Cavalcanti-Mata M, Duarte ME, Lisboa HM. Influence of spray drying conditions on the properties of avocado powder drink. Food Chem. 2018, 266:284-91. https://doi.org/10.1016/j.foodchem.2018.06.016.

28. Schaffer B, Wolstenholme BN, Whiley AW. Introduction. In: Schaffer B, Wolstenholme BN, Whiley AW, editors. The avocado: botany, production and uses. 2nd ed. Wallingford: CABI; 2013. p. 1-9.

29. Araújo RG, Rodriguez-Jasso RM, Ruiz HA, Pintado MME, Aguilar CN. Avocado by-products: Nutritional and functional properties. Trends Food Sci. Technol. 2018, 80:51-60. https://doi.org/10.1016/j.tifs.2018.07.027.

30. Food and Agriculture Organization of the United Nations. FAOSTAT. In. Food and Agriculture Organization of the United Nations. 2026. https://www.fao.org/faostat/en/#home. Accessed 20 Apr 2026.

31. People's Daily Online. "Foreign fruits" have become "local specialties". In: Yunnan Channel. People's Daily Online. 2025. http://yn.people.com.cn/n2/2025/0621/c378439-41267148.html. Accessed 20 April 2026.

32. Fick A, Swart V, Bombarely A, van den Berg N. Comparative transcriptional analysis of *Persea americana* MYB, WRKY and AP2/ERF transcription factors following *Phytophthora cinnamomi* infection. Mol. Plant Pathol. 2024, 25(4):e13453. https://doi.org/10.1111/mpp.13453.

33. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011, 39:W29-W37. https://doi.org/10.1093/nar/gkr367.

34. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 2012, 28(23):3150-2. https://doi.org/10.1093/bioinformatics/bts565.

35. Hao Z, Lv D, Ge Y, Shi J, Weijers D, Yu G, Chen J. RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. PeerJ Comput. Sci. 2020:e251. https://doi.org/10.7717/peerj-cs.251.

36. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T, Jin H, Marler B, Guo H et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012, 40(7):49. https://doi.org/10.1093/nar/gkr1293.

37. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: An information aesthetic for comparative genomics. Genome Res. 2009, 19(9):1639-45. https://doi.org/10.1101/gr.092759.109.

38. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol. Plant 2020, 13(8):1194-202. https://doi.org/10.1016/j.molp.2020.06.009.

39. Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. Synteny and collinearity in plant genomes. Science 2008, 320(5875):486-8. https://doi.org/10.1126/science.1153917.

40. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 2009, 25(9):1189-91. https://doi.org/10.1093/bioinformatics/btp033.

41. Tamura K, Stecher G, Kumar S. MEGA11: Molecular evolutionary genetics analysis version 11. Mol. Biol. Evol. 2021, 38(7):3022-7. https://doi.org/10.1093/molbev/msab120.

42. 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(1):490. https://doi.org/10.1186/s12864-018-4880-x.

43. Subramanian B, Gao S, Lercher MJ, Hu S, Chen W. Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees. Nucleic Acids Res. 2019, 47(W1):W270-5. https://doi.org/10.1093/nar/gkz357.

44. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT et al. Gene Ontology: tool for the unification of biology. Nat. Genet. 2000, 25(1):25-9. https://doi.org/10.1038/75556.

45. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000, 28(1):27-30. https://doi.org/10.1093/nar/28.1.27.

46. Yu G, Wang L, Han Y, He Q. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics 2012, 16(5):284-7. https://doi.org/10.1089/omi.2011.0118.

47. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014, 15(12):550. https://doi.org/10.1186/s13059-014-0550-8.

48. Sánchez-Baizán N, Ribas L, Piferrer F. Improved biomarker discovery through a plot twist in transcriptomic data analysis. BMC Biol. 2022, 20(1):208. https://doi.org/10.1186/s12915-022-01398-w.

49. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. 2005, 4:Article17. https://doi.org/10.2202/1544-6115.1128.

50. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics 2017, 33(18):2938-40. https://doi.org/10.1093/bioinformatics/btx364.

51. Benkert P, Tosatto SCE, Schomburg D. QMEAN: A comprehensive scoring function for model quality assessment. Proteins 2008, 71(1):261-77. https://doi.org/10.1002/prot.21715.

52. Tang H, Zou P, Zhang C, Chen R, Chen W, Lin H. Identification of apolipoprotein using feature selection technique. Sci. Rep. 2016, 6:30441. https://doi.org/10.1038/srep30441.

53. Wu W, Yang J, Yu N, Li R, Yuan Z, Shi J, Chen J. Evolution of the WRKY family in angiosperms and functional diversity under environmental stress. Int. J. Mol. Sci. 2024, 25(6):3551. https://doi.org/10.3390/ijms25063551.

54. Gao J, Chen Y, Gao M, Wu L, Zhao Y, Wang Y. *LcWRKY17*, a WRKY transcription factor from litsea cubeba, effectively promotes monoterpene synthesis. Int. J. Mol. Sci. 2023, 24(8):7210. https://doi.org/10.3390/ijms24087210.

55. Wang Z, You L, Gong N, Li C, Li Z, Shen J, Wan L, Luo K, Su X, Feng L et al. Comprehensive expression analysis of the WRKY gene family in *Phoebe bournei* under drought and waterlogging stresses. Int. J. Mol. Sci. 2024, 25(13):7280. https://doi.org/10.3390/ijms24087210.

56. Long S, Jiang J, Lin H. Whole-genome identification and expression profiles of *WRKY* genes related to the leaf expansion period in the camphor tree. Forests 2025, 16(2):266. https://doi.org/10.3390/f16020266.

57. Hui N, Weihong S, Le D, Weiwei Z, Shuangquan Z. Identification and analysis of the WRKY gene family in whole genome of *Cinnamomum camphora* (L.) Presl. Plant Sci. J. 2022, 40(4):512-23. https://doi.org/10.11913/PSJ.2095-0837.2022.40513.

58. Song X, Hou X, Zeng Y, Jia D, Li Q, Gu Y, Miao H. Genome-wide identification and comprehensive analysis of WRKY transcription factor family in safflower during drought stress. Sci. Rep. 2023, 13(1):16955. https://doi.org/10.1038/s41598-023-44340-y.

59. Yang Y, Zhou Y, Chi Y, Fan B, Chen Z. Characterization of soybean WRKY gene family and identification of soybean *WRKY* genes that promote resistance to soybean cyst nematode. Sci. Rep. 2017, 7(1):17804. https://doi.org/10.1038/s41598-017-18235-8.

60. Mao C, Wang L, Kong D, Wang H, Zhang H, Shen Y, Xu J. Bioinformatics analysis of WRKY gene family in barley genome. Mol. Plant Breed. 2019, 17(23):7623-33. https://doi.org/10.13271/j.mpb.017.007623.

61. Zhang Z, Ji M, Ze S, Song W, Yang B, Zhao N. Genome-wide identification and expression analysis of the WRKY gene family in *Mikania micrantha*. BMC Genomics 2025, 26(1):2. https://doi.org/10.1186/s12864-024-11187-0.

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

63. Rushton PJ, Somssich IE, Ringler P, Shen QJ. WRKY transcription factors. Trends Plant Sci. 2010, 15(5):247-58. https://doi.org/10.1016/j.tplants.2010.02.006.

64. Freeling M. Bias in plant gene content following different sorts of duplication: Tandem, whole-genome, segmental, or by transposition. Annu. Rev. Plant Biol. 2009, 60(60):433-53. https://doi.org/10.1146/annurev.arplant.043008.092122.

65. Xie M, Zhao C, Song M, Xiang Y, Tong C. Genome-wide identification and comparative analysis of CLE family in rapeseed and its diploid progenitors. Front. Plant Sci. 2022, 13:998082. https://doi.org/10.3389/fpls.2022.998082.

66. Li J, Wang X, Lu J, Song H, Lei H, Niu T, Liu A. Genome-wide identification and expression analysis of the WRKY gene family in *Sophora flavescens* during tissue development and salt stress. Front. Plant Sci. 2024, 15:1520786. https://doi.org/10.3389/fpls.2024.1520786.

67. Liu J, Li G, Wang R, Wang G, Wan Y. Genome-wide analysis of WRKY transcription factors involved in abiotic stress and ABA response in Caragana korshinskii. Int. J. Mol. Sci. 2023, 24(11):9519. https://doi.org/10.3390/ijms24119519.

68. Li J, Islam F, Huang Q, Wang J, Zhou W, Xu L, Yang C. Genome-wide characterization of WRKY gene family in *Helianthus annuus* L. and their expression profiles under biotic and abiotic stresses. 2020, 15(12):e0241965. https://doi.org/10.1371/journal.pone.0241965.

69. van Verk MC, Pappaioannou D, Neeleman L, Bol JF, Linthorst HJM. A Novel WRKY transcription factor is required for induction of *PR-1a* gene expression by salicylic acid and bacterial elicitors. Plant Physiol. 2008, 146(4):1983-95. https://doi.org/10.1104/pp.107.112789.

70. 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. Crit. Rev. Plant Sci. 2017, 36:311-35. https://doi.org/10.1080/07352689.2018.1441103.

71. Li Y, Song Q, Zhang Y, Li Z, Guo J, Chen X, Zhang G. Genome-wide identification, characterization, and expression patterns analysis of the SBP-box gene family in wheat (*Triticum aestivum* L.). Sci Rep. 2020, 10(1):17250. https://doi.org/10.1038/s41598-020-74417-x.

72. Du P, Wang Q, Yuan DY, Chen SS, Su YN, Li L, Chen S, He XJ. WRKY transcription factors and OBERON histone-binding proteins form complexes to balance plant growth and stress tolerance. EMBO J. 2023, 42(19):e113639.

https://doi.org/10.15252/embj.2023113639.

73. Phukan UJ, Jeena GS, Shukla RK. WRKY transcription factors: Molecular regulation and stress responses in plants. Front. Plant Sci. 2016, 7:760. https://doi.org/10.3389/fpls.2016.00760.

74. Yuan D, Cai J, Zhang T, Wang S, Yang X, Li Y. Identification and analysis of the WRKY transcription factor gene family in *Verbena bonariensis*. Phyton Int. J. Exp. Bot. 2024, 93(8):1875-96. https://doi.org/10.32604/phyton.2024.052190.

75. Zhao Z, Wang R, Su W, Sun T, Qi M, Zhang X, Wei F, Yu Z, Xiao F, Yan L et al. A comprehensive analysis of the WRKY family in soybean and functional analysis of GmWRKY164-GmGSL7c in resistance to soybean mosaic virus. BMC Genomics 2024, 25(1):620. https://doi.org/10.1186/s12864-024-10523-8.

76. Qiao X, Yin H, Li L, Wang R, Wu J, Wu J, Zhang S. Different modes of gene duplication show divergent evolutionary patterns and contribute differently to the expansion of gene families involved in important fruit traits in pear (*Pyrus bretschneideri*). Front. Plant Sci. 2018, 9:161. https://doi.org/10.3389/fpls.2018.00161.

77. Marand AP, Eveland AL, Kaufmann K, Springer NM. *cis*-regulatory elements in plant development, adaptation, and evolution. Annu. Rev. Plant Biol. 2023, 74:111-37. https://doi.org/10.1146/annurev-arplant-070122-030236.

78. Xu X, Mo Q, Cai Z, Jiang Q, Zhou D, Yi J. Promoters, key cis-regulatory elements, and their potential applications in regulation of cadmium (Cd) in rice. Int. J. Mol. Sci. 2024, 25(24):13237. https://doi.org/10.3390/ijms252413237.

79. Xu X, Chen C, Fan B, Chen Z. Physical and functional interactions between pathogen-induced *Arabidopsis* WRKY18, WRKY40, AND WRKY60 transcription factors. Plant Cell 2006, 18(5):1310-26. https://doi.org/10.1105/tpc.105.037523.

80. Cheng Y, Zhou Y, Yang Y, Chi Y, Zhou J, Chen J, Wang F, Fan B, Shi K, Zhou Y et al. Structural and functional analysis of VQ motif-containing proteins in *Arabidopsis* as interacting proteins of WRKY transcription factors Plant Physiol. 2012, 159(2):810-25. https://doi.org/10.1104/pp.112.196816.

81. Chen X, Li C, Wang H, Guo Z. WRKY transcription factors: evolution, binding, and action. Phytopathol. Res. 2019, 1(1):13. https://doi.org/10.1186/s42483-019-0022-x.

82. Soltani Z, Moghadam A, Shamekh M. Comparative meta-analysis of barely transcriptome: Pathogen type determines host preference. PLoS One 2025, 20(6):e0320708. https://doi.org/10.1371/journal.pone.0320708.

83. Kong F, Yang L. Pathogen-triggered changes in plant development: Virulence strategies or host defense mechanism? Front. Microbiol. 2023, 14:1122947. https://doi.org/10.3389/fmicb.2023.1122947.

84. Patharkar OR, Gassmann W, Walker JC. Leaf shedding as an anti-bacterial defense in *Arabidopsis* cauline leaves. PLoS Genet. 2017, 13(12):e1007132. https://doi.org/10.1371/journal.pgen.1007132.

85. Wang N, Xu S, Sun Y, Liu D, Zhou L, Li Y, Li X. The cotton WRKY transcription factor (GhWRKY33) reduces transgenic *Arabidopsis* resistance to drought stress. Sci. Rep. 2019, 9(1):724. https://doi.org/10.1038/s41598-018-37035-2.

86. Su X, Zhao M, Zhou R, Xu C, Zhang R, Li R, Wang T. The mulberry *WRKY* transcription factor *MAWRKYIIc7* participates in regulating plant drought stress tolerance. Int. J. Mol. Sci. 2025, 26(4):1714. https://doi.org/10.3390/ijms26041714.

87. Yang D, Zhang J, Zhou D, Qian M, Zheng Y, Jin L. Leaf and twig functional traits of woody plants and their relationships with environmental change; A review. Chin. J. Ecol. 2012, 31:702-13.

88. Zhao L, Yu J, Chen S, Liu L, Li X, Chen W, Xu Z, Pan X, Zhou W, Chen L et al. The transcription factor *OsWRKY36* regulates leaf angle in rice by modulating brassinosteroid signalling. Plant Cell 2026, 38:118. https://doi.org/10.1093/plcell/koag118.

89. Kloth KJ, Wiegers GL, Busscher-Lange J, van Haarst JC, Kruijer W, Bouwmeester HJ, Dicke M, Jongsma MA. AtWRKY22 promotes susceptibility to aphids and modulates salicylic acid and jasmonic acid signalling. J. Exp. Bot. 2016, 67(11):3383-96. https://doi.org/10.1093/jxb/erw159.

90. Skibbe M, Qu N, Galis I, Baldwin IT. Induced plant defenses in the natural environment: *Nicotiana attenuata* WRKY3 and WRKY6 coordinate responses to herbivory. Plant Cell 2008, 20(7):1984-2000. https://doi.org/10.1105/tpc.108.058594.

91. Li R, Zhang J, Li J, Zhou G, Wang Q, Bian W, Erb M, Lou Y. Prioritizing plant defence over growth through WRKY regulation facilitates infestation by non-target herbivores.

eLife 2015, 4:e04805. https://doi.org/10.7554/eLife.04805.

92. Hou M, Xu G. Growth-defense trade-off in plants: From hypothesis to principle to paradigm. Cell Host Microbe 2025, 33(8):1222-6. https://doi.org/10.1016/j.chom.2025.05.022.

93. Zrimec J, Correa S, Zagorščak M, Petek M, Bleker C, Stare K, Schuy C, Sonnewald S, Gruden K, Nikoloski Z. Evaluating plant growth–defense trade-offs by modeling the interaction between primary and secondary metabolism. PNAS 2025, 122(32):e2502160122. https://doi.org/10.1073/pnas.2502160122.

94. Dong Q, Duan D, Wang F, Yang K, Song Y, Wang Y, Wang D, Ji Z, Xu C, Jia P et al. The MdVQ37-MdWRKY100 complex regulates salicylic acid content and *MdRPM1* expression to modulate resistance to Glomerella leaf spot in apples. Plant Biotechnol. J 2024, 22(8):2364-76. https://doi.org/10.1111/pbi.14351.

95. Wang W, Bai X, Chen K, Gu C, Yu Q, Jiang J, Liu G. Role of PsnWRKY70 in regulatory network response to infection with *Alternaria alternata* (Fr.) Keissl in *Populus*. Int. J. Mol. Sci. 2022, 23(14):7537. https://doi.org/10.3390/ijms23147537.

96. Inthaisong S, Boonchuen P, Jaichopsanthia T, Songwattana P, Khairum A, Chueakhunthod W, Tharapreuksapong A, Tittabutr P, Teaumroong N, Tantasawat PA. Insights into mungbean defense response to Cercospora leaf spot based on transcriptome analysis. Sci. Rep. 2025, 15(1):1334. https://doi.org/10.1038/s41598-024-84787-1.

97. Dong X, Yu L, Zhang Q, Yang J, Gong Z, Niu X, Li H, Zhang X, Liu M, Jin C et al. Structural basis for the regulation of plant transcription factor WRKY33 by the VQ protein SIB1. Commun. Biol. 2024, 7(1):561. https://doi.org/10.1038/s42003-024-06258-7.

Figure legend

Fig. 1 Representative phenotypes of healthy and biotically stressed tissues in Hass avocado. (**A**) Healthy fruit (CKF). (**B**) Developmentally arrested fruit caused by gall midge infestation (IHF) and representative longitudinal sections showing gall midge larvae and feeding sites. (**C**) Healthy leaf (CKL). (**D**) Diseased leaf (DL) showing typical leaf spot symptoms caused by *Pseudocercospora purpurea* (Cooke) Deighton.

(**E**) Leaf damaged by chewing insects. Bar = 2 cm.

Fig. 2 Chromosomal distribution and synteny analysis of *PaWRKY* genes. (**A**) Distribution of *PaWRKY* genes on the 12 avocado chromosomes. The scale on the left indicates chromosome length in megabases (Mb). Gene density along each chromosome is represented by a color gradient from blue (low density) to red (high density). (**B**) Intraspecific synteny analysis of *PaWRKY* genes. Colored links indicate homologous *PaWRKY* gene pairs, and red gene names indicate tandemly duplicated genes. (**C**) Collinearity analysis of *WRKY* genes among avocado, *Arabidopsis thaliana*, *Cinnamomum camphora*, *Oryza sativa*, and *Ananas comosus*. Colored links indicate interspecific homologous *WRKY* gene pairs, and gray lines indicate collinear gene pairs between species.

Fig. 3 Phylogenetic analysis of WRKY proteins from avocado, *A. thaliana*, and *A. comosus*. The phylogenetic tree was constructed using full-length WRKY protein sequences from avocado (PaWRKYs), *A. thaliana* (AtWRKYs), and *A. comosus* (AcWRKYs) based on the neighbor-joining method with 1,000 bootstrap replicates. WRKY proteins were classified into three major groups, I, II, and III, with Group II further divided into five subgroups: IIa, IIb, IIc, IId, and IIe. Different branch colors represent different groups or subgroups. Red stars, green triangles, and blue circles indicate WRKY proteins from avocado, *A. thaliana*, and *A. comosus*, respectively.

Fig. 4 Alignment of conserved WRKY domains and zinc-finger motifs in PaWRKY proteins. PaWRKY proteins were classified according to their phylogenetic relationships. The WRKY heptapeptide and zinc-finger motifs are shown. Group I and II members contain C2H2-type motifs, whereas Group III members contain C2HC-type motifs. PaWRKY proteins with variant heptapeptides are highlighted in red.

Fig. 5 Conserved domains, motif composition, and gene structures of PaWRKY. (**A**) Conserved domain composition of PaWRKY proteins predicted using the Pfam and SMART databases. (**B**) Distribution of conserved motifs in PaWRKY proteins predicted by MEME. Different colored boxes represent different motifs. (**C**) Gene structures of *PaWRKY* genes. Green boxes indicate CDS regions, and black lines indicate introns.

Fig. 6 Distribution of *cis*-acting elements in the promoter regions of *PaWRKY* genes. *Cis*-acting regulatory elements in the 1.5 kb upstream promoter regions of *PaWRKY* genes were predicted using the PlantCARE database. The number in each box represents the copy number of the corresponding *cis*-acting element in the promoter region of each gene.

Fig. 7 Protein interaction network of PaWRKY proteins. The interaction network of PaWRKY proteins was constructed using the STRING database based on orthologous relationships with *Arabidopsis* WRKY proteins. Node size and color indicate the degree of connectivity of each PaWRKY protein, with larger nodes and red color representing higher connectivity. PaWRKY6, PaWRKY40, and PaWRKY70 are highlighted in red.

Fig. 8 K-means clustering and Gene Ontology (GO) enrichment analysis of expressed genes across different tissues and biotic stress conditions. (**A**) K-means clustering of standardized expression profiles (Z-score normalized) of expressed genes across different tissues and biotic stress conditions (K = 5). The y-axis represents Z-score-normalized expression levels, and the x-axis represents different tissues and biotic stress conditions (CKL, IHL, DL, CKF, and IHF). Each thin line represents the expression pattern of one gene, and the thick red line represents the overall expression trend within each cluster. (**B**) GO enrichment analysis of genes in each K-means cluster ($p < 0.05$). Different colors indicate different clusters, and circle size represents the number of genes.

Fig. 9 GO and KEGG enrichment analyses of differentially expressed genes (DEGs) under different biotic stresses. (**A**) and (**B**) GO enrichment analysis of up- and down-regulated genes in IHL, DL, and IHF, respectively. Bar length represents the number of genes, and bar color indicates enrichment significance [$-\log_{10}(p$ value)]. (**C**) and (**D**) KEGG enrichment analysis of up- and down-regulated genes in IHL, DL, and IHF, respectively. Bubble size represents the number of genes, and bubble color indicates enrichment significance [$-\log_{10}(p$ value)]. Pathways labeled in red and yellow indicate defense- and hormone signaling-related pathways enriched among up- and down-regulated genes, respectively. Only significantly enriched pathways in each treatment are shown ($p < 0.05$).

Fig. 10 Expression heatmaps of *PaWRKY* genes across different tissues and under biotic stress conditions. (**A**) Expression profiles of *PaWRKY* genes in flower buds, stems, leaves, and fruits. (**B**) Expression profiles of *PaWRKY* genes in CKL, IHL, DL, CKF, and IHF. The outer ring indicates different WRKY subgroups. The color gradient represents standardized expression levels, with pink and red indicating relatively high expression and green and blue indicating relatively low expression.

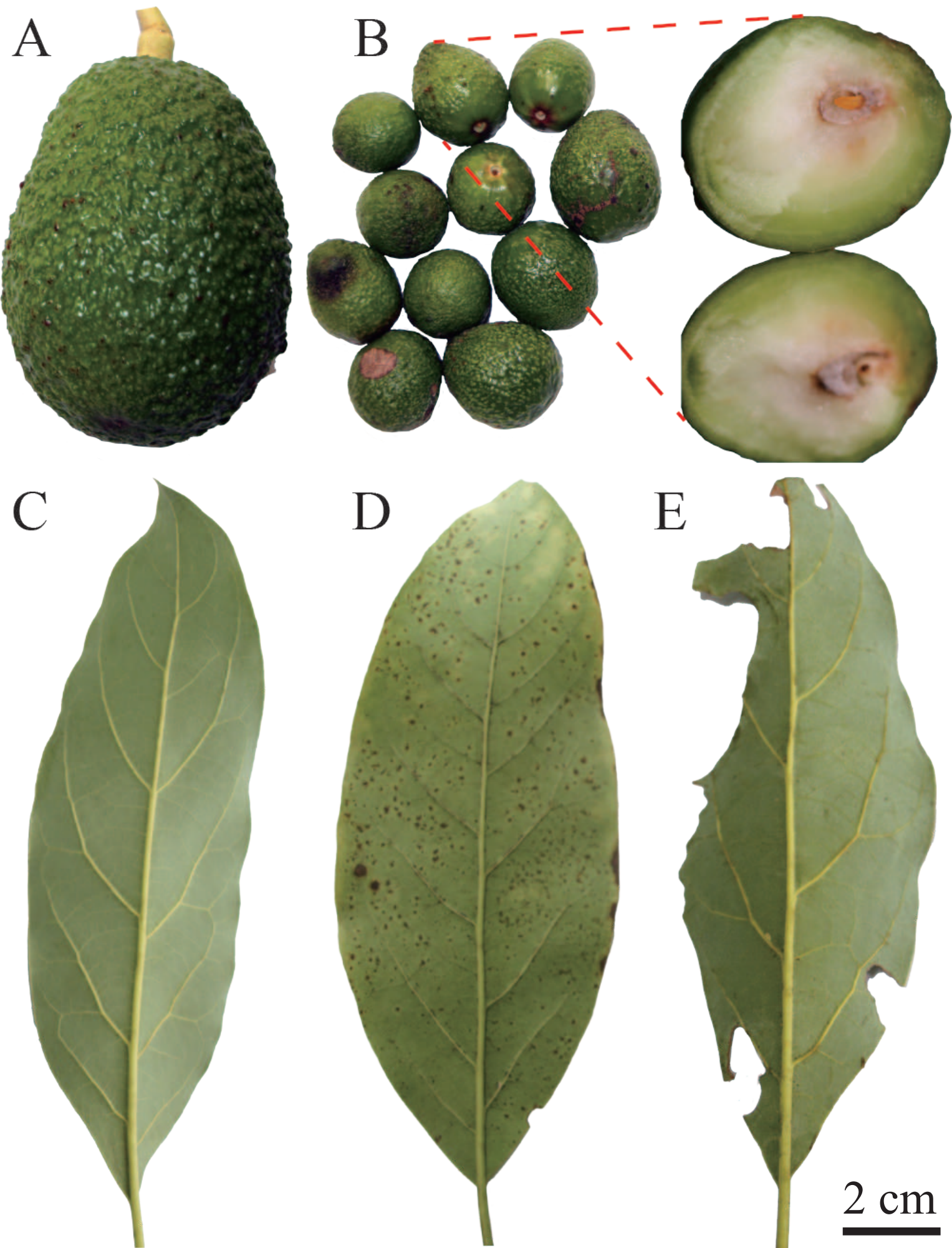
Fig. 11 Weighted gene co-expression network analysis (WGCNA) under different biotic stress conditions. (**A**) Gene dendrogram of 46,892 annotated genes constructed by hierarchical clustering based on topological overlap. Each branch represents one gene, and different colors indicate different co-expression modules. (**B**) Heatmap of module–trait relationships showing the correlations between module eigengenes and different treatments or tissues (CKL, IHL, DL, CKF, and IHF). Each cell shows the Pearson correlation coefficient and corresponding p value. Red indicates positive correlations, and blue indicates negative correlations. (**C**) GO enrichment analysis of genes in modules significantly associated with different treatments. Bubble size represents the number of enriched genes, and bubble color indicates enrichment significance [$-\log_{10}(p$ value)]. Only significantly enriched terms in module genes associated with each treatment are shown ($p < 0.05$). (**D**) KEGG enrichment analysis of genes in modules significantly associated with different treatments. From the outer to inner rings: (**1**) the first ring shows KEGG Orthology (KO) accession numbers, colored according to treatment; (**2**) the second ring represents the total number of genes enriched in each KO term; (**3**) the third ring represents enrichment significance based on $-\log_{10}(p$ value), with both bar height and color indicating the significance level; and (**4**) the fourth ring represents the enrichment factor, calculated as the ratio of enriched genes to background genes in each pathway.

Fig. 12 Integrative analysis of *PaWRKY* genes under different biotic stresses. (**A**) UpSet analysis of IHL based on intersections among K-means cluster 4, WGCNA_IHL, IHL_UP, IHL_DOWN, and WRKY gene sets. (**B**) UpSet analysis of DL based on intersections among K-means cluster 5, WGCNA_DL, DL_UP, DL_DOWN, and WRKY gene sets. (**C**) UpSet analysis of IHF based on intersections among K-means

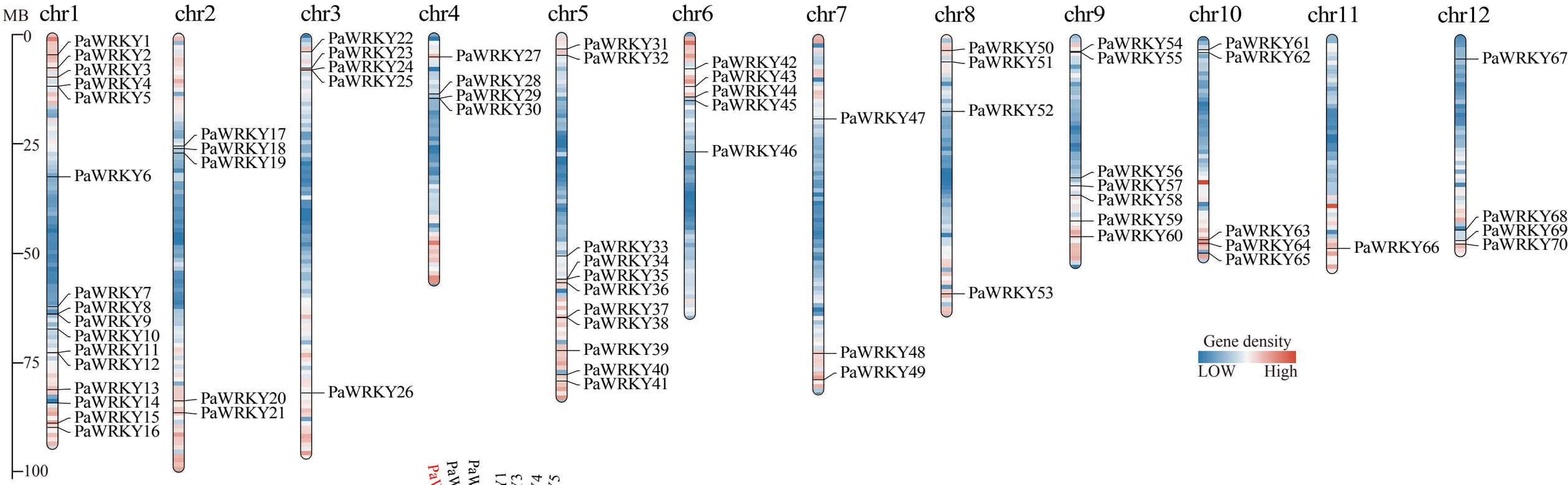
cluster 2, WGCNA_IHF, IHF_UP, IHF_DOWN, and WRKY gene sets. Horizontal bars on the left indicate the number of genes in each gene set, connected dots at the bottom indicate exact intersection combinations, and upper bars indicate the number of genes in each intersection. Scatter points represent the expression levels of genes in each intersection [$\log_{10}$(FPKM)]. Different colors represent different gene categories: red, up-regulated genes; green, down-regulated genes; and blue, *WRKY* genes.

Fig. 13. qRT-PCR validation of 10 core WRKY genes responsive to biotic stresses in avocado. The expression patterns of 10 selected core stress-responsive *PaWRKY* genes were validated by qRT-PCR and compared with RNA-seq expression levels across CKL, IHL, DL, CKF, and IHF samples. Blue bars represent qRT-PCR relative expression levels, and orange bars represent RNA-seq expression levels based on FPKM values. The left y-axis indicates qRT-PCR relative expression, and the right y-axis indicates RNA-seq expression level. qRT-PCR data are presented as the mean ± SD of nine measurements from three biological replicates with three technical replicates each (n = 9), whereas RNA-seq data are presented as the mean ± SD of three biological replicates (n = 3). Different letters indicate significant differences among samples within the same dataset ($p < 0.05$).

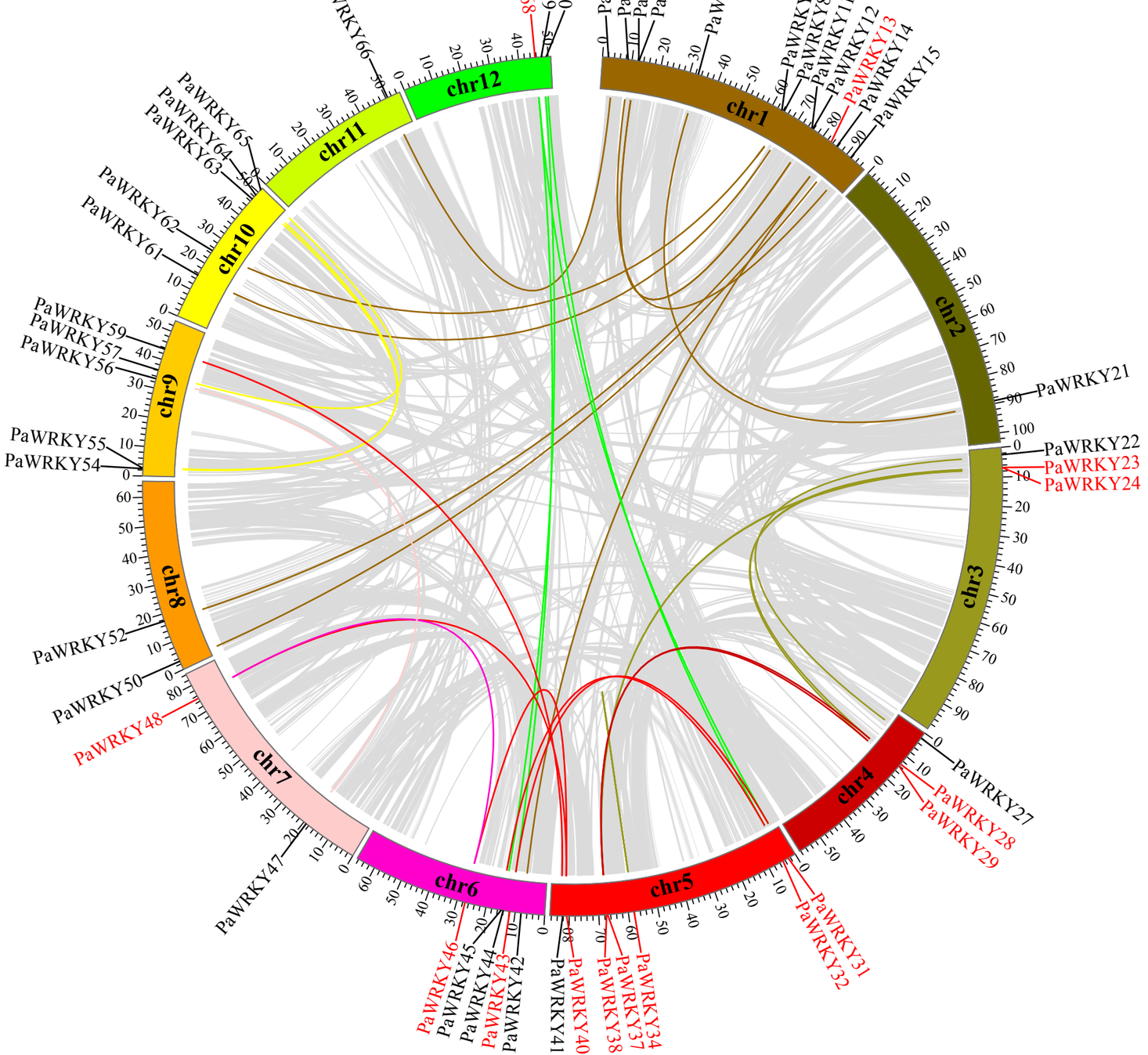
Fig. 14 Predicted three-dimensional structures of core biotic stress-responsive PaWRKY proteins based on homology modelling using SWISS-MODEL. The predicted three-dimensional structures of core PaWRKY proteins responsive to biotic stresses are shown. Each model is labeled with its corresponding WRKY heptapeptide motif and structural features. The blue helices in the peripheral regions represent α-helical structures.


A
B
C
D
E
2 cm

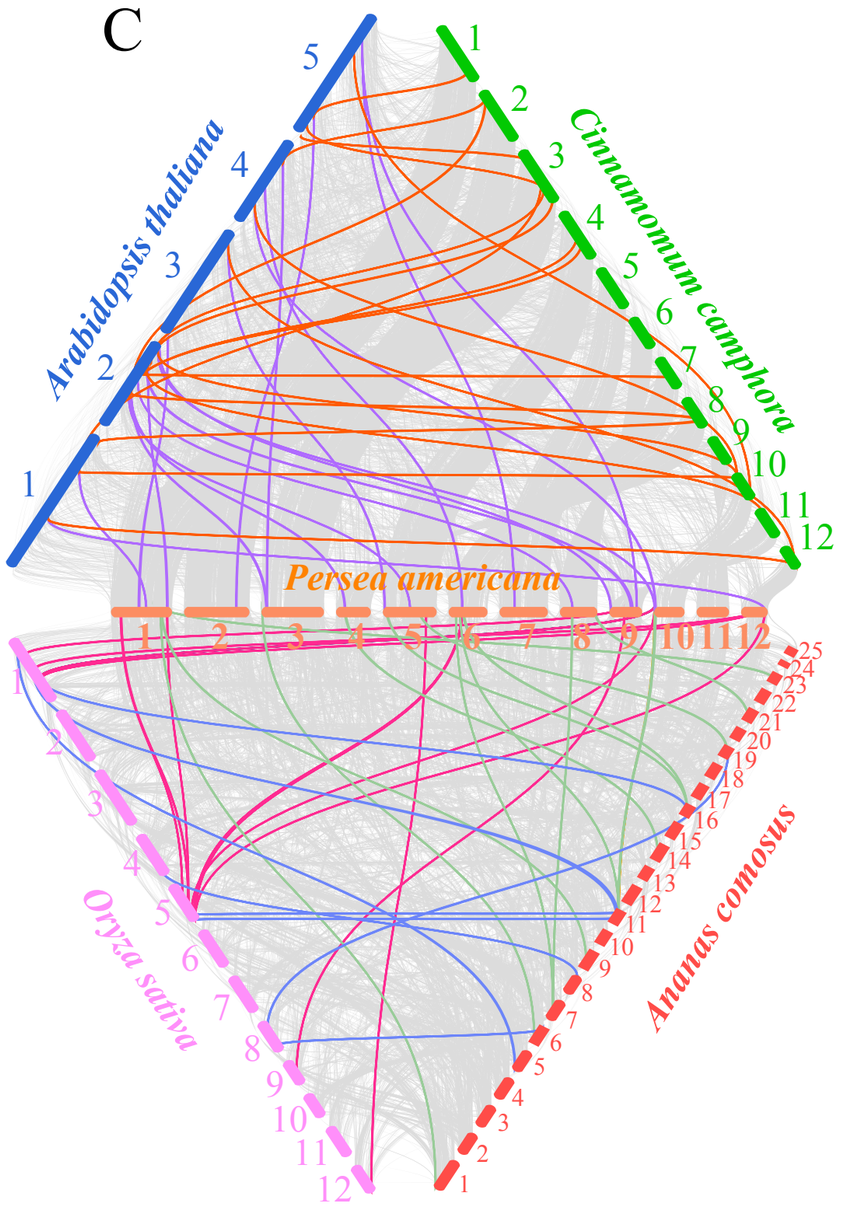
A

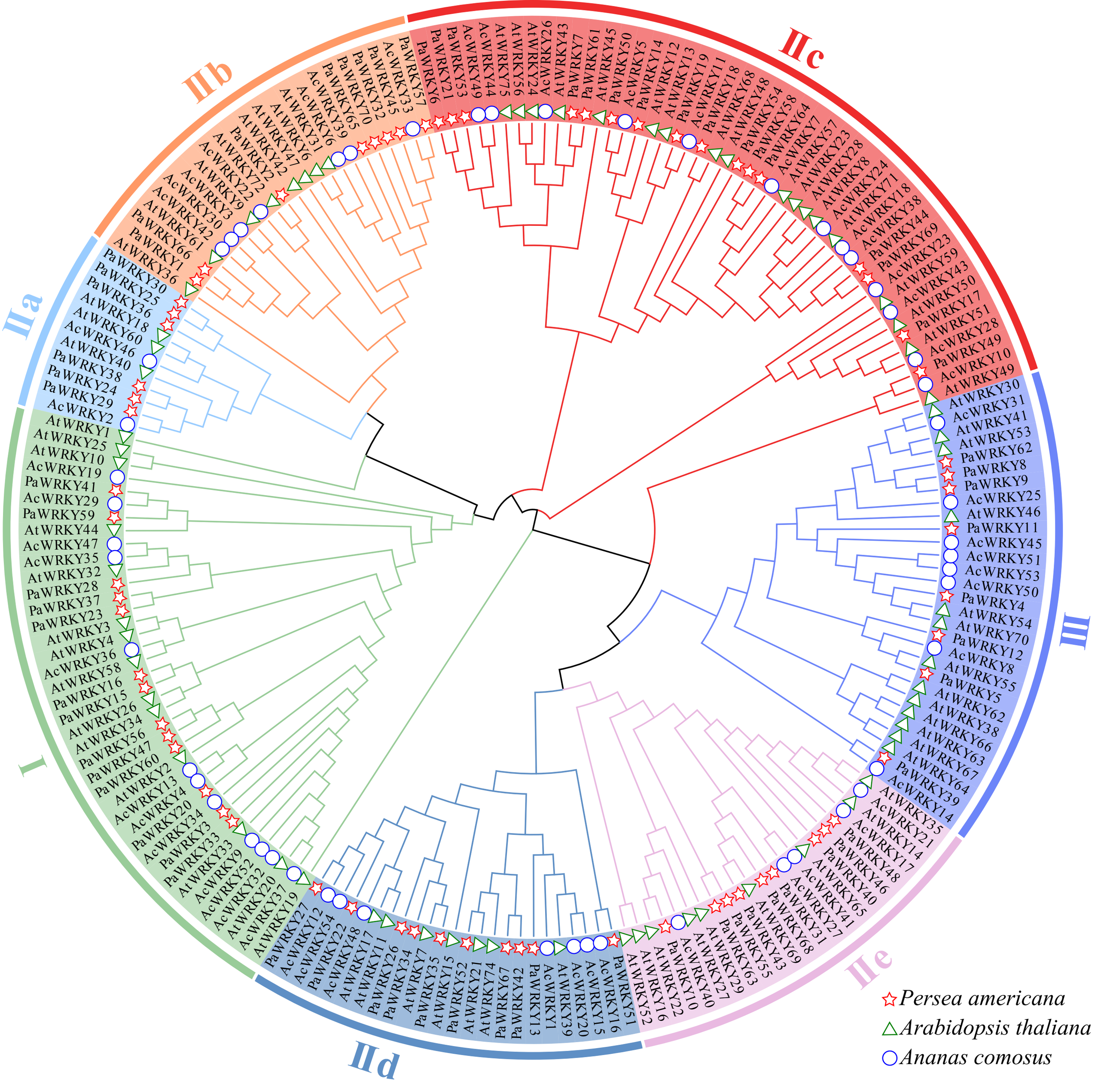

MB
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
0
25
50
75
100
PaWRKY1
PaWRKY2
PaWRKY3
PaWRKY4
PaWRKY5
PaWRKY6
PaWRKY7
PaWRKY8
PaWRKY9
PaWRKY10
PaWRKY11
PaWRKY12
PaWRKY13
PaWRKY14
PaWRKY15
PaWRKY16
PaWRKY17
PaWRKY18
PaWRKY19
PaWRKY20
PaWRKY21
PaWRKY22
PaWRKY23
PaWRKY24
PaWRKY25
PaWRKY26
PaWRKY27
PaWRKY28
PaWRKY29
PaWRKY30
PaWRKY31
PaWRKY32
PaWRKY33
PaWRKY34
PaWRKY35
PaWRKY36
PaWRKY37
PaWRKY38
PaWRKY39
PaWRKY40
PaWRKY41
PaWRKY42
PaWRKY43
PaWRKY44
PaWRKY45
PaWRKY46
PaWRKY47
PaWRKY48
PaWRKY49
PaWRKY50
PaWRKY51
PaWRKY52
PaWRKY53
PaWRKY54
PaWRKY55
PaWRKY56
PaWRKY57
PaWRKY58
PaWRKY59
PaWRKY60
PaWRKY61
PaWRKY62
PaWRKY63
PaWRKY64
PaWRKY65
PaWRKY66
PaWRKY67
PaWRKY68
PaWRKY69
PaWRKY70
Gene density
LOW
High

B


chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12

C


Arabidopsis thaliana
Cinnamomum camphora
Persea americana
Oryza sativa
Ananas comosus


I
IIa
IIb
IIc
IId
IIe
III
PaWRKY6
PaWRKY21
PaWRKY53
AcWRKY49
AcWRKY44
AtWRKY75
AtWRKY59
AtWRKY24
AcWRKY26
AtWRKY43
PaWRKY7
PaWRKY61
AtWRKY45
PaWRKY50
AcWRKY5
PaWRKY14
AtWRKY12
PaWRKY13
AcWRKY19
PaWRKY11
AtWRKY18
AtWRKY68
PaWRKY48
PaWRKY54
PaWRKY58
PaWRKY64
AcWRKY7
AtWRKY57
AtWRKY23
AtWRKY28
AtWRKY8
AcWRKY24
AtWRKY71
AtWRKY18
AcWRKY38
PaWRKY44
PaWRKY69
AcWRKY23
AtWRKY59
AcWRKY43
AtWRKY50
AcWRKY17
PaWRKY51
AtWRKY28
AcWRKY49
PaWRKY10
AtWRKY49
PaWRKY57
AcWRKY33
PaWRKY45
PaWRKY32
PaWRKY70
PaWRKY65
AcWRKY39
AcWRKY6
AtWRKY31
AtWRKY6
AtWRKY47
AtWRKY42
PaWRKY2
AtWRKY72
AcWRKY22
AtWRKY9
AcWRKY3
AcWRKY30
AtWRKY42
AcWRKY61
PaWRKY66
PaWRKY1
AtWRKY36
PaWRKY30
PaWRKY25
PaWRKY36
AtWRKY18
AtWRKY60
AcWRKY46
AtWRKY40
PaWRKY38
PaWRKY24
PaWRKY29
AcWRKY2
AtWRKY1
AtWRKY25
AtWRKY10
AcWRKY19
PaWRKY41
AcWRKY29
PaWRKY59
AtWRKY44
AcWRKY47
AcWRKY35
AtWRKY32
PaWRKY28
PaWRKY37
PaWRKY23
AtWRKY3
AtWRKY4
AcWRKY36
AtWRKY58
PaWRKY16
PaWRKY15
AtWRKY26
AtWRKY34
PaWRKY56
PaWRKY47
PaWRKY60
AtWRKY2
AcWRKY13
AcWRKY4
PaWRKY20
AcWRKY34
PaWRKY3
PaWRKY33
AtWRKY33
AcWRKY9
AcWRKY52
AcWRKY32
AtWRKY20
AcWRKY37
AtWRKY19
PaWRKY27
AcWRKY12
PaWRKY54
AcWRKY22
PaWRKY48
AcWRKY17
AtWRKY11
PaWRKY26
PaWRKY34
AtWRKY7
PaWRKY35
AtWRKY15
PaWRKY52
AtWRKY21
AtWRKY74
PaWRKY67
PaWRKY42
PaWRKY13
AcWRKY1
AtWRKY39
AcWRKY20
AcWRKY15
AcWRKY16
PaWRKY51
AtWRKY52
AtWRKY16
PaWRKY22
AcWRKY10
AtWRKY40
AtWRKY27
PaWRKY29
PaWRKY63
PaWRKY55
AtWRKY43
PaWRKY69
PaWRKY68
AtWRKY31
AcWRKY27
AcWRKY41
PaWRKY65
PaWRKY40
PaWRKY46
PaWRKY48
AtWRKY17
AcWRKY14
AcWRKY21
AtWRKY35
AcWRKY39
PaWRKY64
AtWRKY67
AtWRKY63
AtWRKY66
AtWRKY38
AtWRKY62
PaWRKY5
AtWRKY55
PaWRKY12
AcWRKY8
AtWRKY70
AtWRKY54
PaWRKY4
AcWRKY50
AcWRKY53
AcWRKY51
AcWRKY45
PaWRKY11
AtWRKY46
AcWRKY25
PaWRKY9
PaWRKY8
PaWRKY62
AtWRKY53
AtWRKY41
AtWRKY30
Persea americana
Arabidopsis thaliana
Ananas comosus

Subfamily I

WRKY C X_4 C C2H2 X_{23} H X H

PaWRKY3 WRKYGQK VVKGNPNPRSYYKCTSVGCPVRKHVERASHDLRAVITTYEGKHNH
PaWRKY15 WRKYGQK VVKGNANPRSYYKCTQVGCTVRKHVERASNDLRDVITTYEGKHNH
PaWRKY16 WRKYGQK VVKGNSNPRSYYKCTNSGCTVRKHVERASADPKAVITTYEGKHNH
PaWRKY20 WRKYGQK VVKGNPNPRSYYKCTNAGCMVRKHVERASHDLKSVITTYEGKHNH
PaWRKY23 WRKYGQK IVKGNPNPRSYYKCTYDRCQVRKHVERASDDAKALVITYEGKHNH
PaWRKY28 WRKYGQK IVKGNPNPRSYYKCTHVGCSVRKHVERASDDENALIITYEGKHNH
PaWRKY33 WRKYGQK VVKGNPNPRSYYKCTNTGCPVRKHVERASHDPKAVITTYEGKHNH
PaWRKY37 WRKYGQK MVKGNPNPRSYYKCTYDECPVRRHVERASDDSKTLVITYDGKHNH
PaWRKY41 WRKYGQK VVKGNPYPRSYYRCTSHKCNVRKHVERASDDPRAFITTYEGKHNH
PaWRKY47 WRKYGQK VVKGNPNPRSYYKCTNAGCTVRKHVERASHDLKSVITTYEGKHNH
PaWRKY56 WRKYGQK VVKGNPNPRSYYKCTNPGCTVRKHVERASHDLKSVITTYEGKHNH
PaWRKY59 WRKYGQK VVKGNQYPRSYYRCTNLKCNVRKHVERTSDDPRAFITTYEGKHNH
PaWRKY60 WRKYGQK VVKGNPNPRSYYKCTNAGCTVRKHVERASHDLKSVITTYEGKHNH

10 20 30 40 50

Subfamily IIa

WRKY C X_5 C C2H2 X_{23} H X H

PaWRKY24 WRKYGQK VTRDNPSPRAYFKCSFAPACPVKKKVQRSAKDRSILIATYEGEHIH
PaWRKY25 WRKYGQK NTKDNPSPRAYFRCSTAPGCPVKKKVQRCIDDKSVLMATYEGEHTH
PaWRKY29 WRKYGQK VTRDNPCPRAYFKCSFAPTCPVKKKVQRRAEDRSILVATYEGEHNH
PaWRKY30 WRKYGQK ITKDNPWPRAYFRCSTASGCPVKKKVQRCIEDKSILMVMYEGEHTH
PaWRKY36 WRKYGQK ITKDNPSPRAYFRCSTSSDCPVRKKVQRCAGDKSILVVTYEGEHAH
PaWRKY38 WRKYGQK VTRDNPSPRAYFKCSFAPACPVKKKVQRSAEDRSILIATYEGVHNH

10 20 30 40 50

Subfamily IIb

WRKY C X_5 C C2H2 X_{23} H X H

PaWRKY1 WRKYGQK IAKANPCPRAYYRCTVAPGCPVRKQVQRCMEDMSILITTYEGTHNH
PaWRKY2 WRKYGQK IAKGNPCPRAYYRCTVAPGCPVRKQVQRCADDMSILITTYEGNHNH
PaWRKY32 WRKYGQK MAKGNPCPRAYYRCTMAAGCPVRKQVQRCAEDRSILITTYEGNHNH
PaWRKY45 WRKYGQK MAKGNPCPRAYYRCTMAAGCPVRKQVQRCAEDRTILITTYEGNHNH
PaWRKY57 WRKYGQK MAKGNPCPRAYYRCTMAIGCPVRKQVQRCAEDKTILITTYEGNHNH
PaWRKY65 WRKYGQK MAKGNPCPRAYYRCTMAVGCPVRKQVQRCAEDKTILITTYEGNHNH
PaWRKY66 WRKYGQK IAKGNPCPRAYYRCTVAPGCPVRKQVQRCMEDMSILITTYEGTHNH
PaWRKY70 WRKYGQK MAKGNPCPRAYYRCTMAAGCPVRKQVQRCMEDRSILITTYEGNHNH

10 20 30 40 50

Subfamily IIc

WRKY C X_4 C C2H2 X_{23} H X H

PaWRKY6 WRKYGQK AVKNNKYPRSYYRCTNQDCNVKKQVQRLSKDEGVVVTTYEGTHTH
PaWRKY7 WRKYGQK AVKNSIYPRSYYRCTHHTCNVKKQVQRLSKDTSIVVTTYEGIHNH
PaWRKY14 WRKYGQK VVKNSLHPRSYYRCTHTNCRVKKRVERLSEDCRMVITTYEVYKGF
PaWRKY17 WRKYGKK TVKNSPNPRNYYRCSNGGCSVKKRIERDREDRNYVITTYEGVHNH
PaWRKY18 WRKYGQK GVKNSPFPRSYYRCTNSKCIVKKRVERCSKDPSIVITTYEGQHCH
PaWRKY19 WRKYGQK VVKNTQHPRSYYRCTQDNCRVKKRVERLAEDPRMVITTYEGRHIH
PaWRKY21 WRKYGQK AVKNNKYPRSYYRCTNQGCNVKKQVQRLSKDEGVVVTTYEGIHTH
PaWRKY44 WRKYGQK AVKNSPYPRSYYRCTTQKCSVKKRVERSFQDPSIVITTYEGQHTH
PaWRKY49 WRKYGQK SIKNSPNPRSYYRCTNPRCSAKKQVERSTEDPDTLIINYEGLHLH
PaWRKY50 WRKYGQK VVKNSLHPRSYYRCTHNNCRVKKRVERLSEDCRMVITTYEGRHTH
PaWRKY53 WRKYGQK AVKNNIHPRSYYRCTHQGCNVKKQVQRLSKDEGIVVTTYEGIHTH
PaWRKY54 WRKYGQK AVKNSPYPRSYYRCTSAMCNVKKRVERSSDDPAIVVTTYEGKHTH
PaWRKY58 WRKYGQK AVKDSPFPRSYYRCTSHMCGVKKRVERSSDDPAIVVTTYEGQHKH
PaWRKY61 WRKYGQK AVKNSNYPRSYYRCTHHTCNVKKQVQRLSKDTSIVVTTYEGVHNH
PaWRKY64 WRKYGQK AVKNSPFPRSYYRCTSAMCGVKKRVERSSEDPAIVVTTYEGQHTH
PaWRKY69 WRKYGQK AVKNSPYPRSYYRCTSQKCSVKKRVERSFQDPTTVITTYEGVHTH

10 20 30 40 50

Subfamily IId

WRKY C X_5 C C2H2 X_{24} H X H

PaWRKY13 WRKYGQK PIKGSPHPRGYYKCSSMRGCPARKHVERCLQDPSMLIVTYEGE-HNH
PaWRKY22 WRKYGQK PIKGSPYPRGYYKCSSVRGCPARKHVERAPDEPTMLIVTYEGE-HRH
PaWRKY26 WRKYGQK PIKGSPHPRGYYKCSSVRGCPARKHVERALDDASMLIVTYEGE-HNH
PaWRKY27 WRKYGQK PIKGSPYPRGYYKCSTMRGCPARKHVERAPDDPTMLIVTYEGE-HRH
PaWRKY34 WKKYGQK PIKGSPEPRGHYKCSSLRGCPARKHVDRALDDPSMLIITYEGD-HNH
PaWRKY35 WRKYGQK PIKGSPHPRGYYKCSSVRGCPARKHVERALDDPSMLIITYEGD-HNH
PaWRKY42 WRKYGQK PIKGSPHPRGYYKCSSMRGCPARKHVERCLEEPSMLIITYEGE-HNH
PaWRKY51 WRKYGQK PIKGSPYPRSYYKCSSMRGCPARKHVERCFGDPTMLIVTYEGD-HNH
PaWRKY52 WRKYGQK PIKGSPHPRGYYKCSSMRGCPARKHVERCLEDPSMLIVTYEAPVHNT
PaWRKY67 WRKYGQK PIKGSPHPRGYYKCSSMRGCPARKHVERCLEEPSMLIITYEGE-HNH

10 20 30 40 50

Subfamily IIe

WRKY C X_5 C C2H2 X_{23} H X H

PaWRKY10 WRKYGQK PIKGSPYPRGYYRCSSSKGCLARKQVERSRADPTMFIITYTAEHNH
PaWRKY31 WRKYGQK PIKGSPYPRGYYRCSSSKGCPARKQVERSRMDPTTLVITYAWDHNH
PaWRKY40 WRKYGQK PIKGSPYPRGYYRCSSSKGCPARKQVERSRNDPNMLVITYSSEHNH
PaWRKY43 WRKYGQK PIKGSPYPRGYYRCSSSKGCPARKQVERSRADPTMLVITYACDHNH
PaWRKY46 WRKYGQK PIKGSPYPRGYYRCSSSKGCSARKQVERSRNDPKMLVITYTSEHNH
PaWRKY48 WRKYGQK PIKGSPYPRGYYRCSSSKGCSARKQVERSRHDPNMLVITYTSEHNH
PaWRKY55 WRKYGQK PIKGSPYPRGYYRCSTSKGCSAKKQVERCRTDASMLIITYTCSHNH
PaWRKY63 WRKYGQK PIKGSPYPRGYYRCSTSKGCSAKKQVERCRTDASMLIITYTSSHNH
PaWRKY68 WRKYGQK PIKGSPYPRGYYRCSSWKGCPARKQVERSSVDPTMLVITYACDHNH

10 20 30 40 50

Subfamily III

WRKY C X_7 C C2HC X_{23} H X C

PaWRKY4 WRKYGQK EILGAKHPRSYFRCTHRNDQGCQATKQVQKMEEDPPNFLITYMGLHTC
PaWRKY5 WRKYGQK EILKSKFPRSYYRCTHKSFYGCDAKKQVQRLDDDPYTFEVMYRGHHTC
PaWRKY8 WIKYGLK DILGAKYPRAYYRCTHSNVQGCAAKKQVQRSEEDPSIFQVMYQGHHTC
PaWRKY9 WRKYGQK DILGAKYPRAYYRCTHRNVQGCAAKKQVQRSEEDPSIFQVMYQGHHTC
PaWRKY11 WRKYGQK EILGATYPRGYFRCTHKHDQGCQATKHVQKSEDDPSLYVVTYMEQHTC
PaWRKY12 WRKYGQK TILNSKFPRSYYHCAHKNL-GCVAKKKVQRMDEVPYMMEITYHFHHTC
PaWRKY39 WKKYGQK FIRSIQKIRCYFKCQRA---TCIAKKKVEWPPTDPSNLIVTYEGHHNH
PaWRKY62 WRKYGQK DILGAKYPRAYYRCTHKHVQGCSARKQVQRSEEDPSIFNIMYQGTHTC

10 20 30 40 50

PaWRKY9
PaWRKY8
PaWRKY62
PaWRKY11
PaWRKY4
PaWRKY39
PaWRKY12
PaWRKY5
PaWRKY10
PaWRKY46
PaWRKY40
PaWRKY48
PaWRKY31
PaWRKY43
PaWRKY68
PaWRKY26
PaWRKY35
PaWRKY34
PaWRKY27
PaWRKY22
PaWRKY13
PaWRKY52
PaWRKY42
PaWRKY67
PaWRKY51
PaWRKY55
PaWRKY63
PaWRKY24
PaWRKY38
PaWRKY29
PaWRKY25
PaWRKY36
PaWRKY30
PaWRKY1
PaWRKY66
PaWRKY2
PaWRKY57
PaWRKY65
PaWRKY70
PaWRKY32
PaWRKY45
PaWRKY44
PaWRKY69
PaWRKY54
PaWRKY64
PaWRKY18
PaWRKY58
PaWRKY17
PaWRKY49
PaWRKY14
PaWRKY50
PaWRKY19
PaWRKY6
PaWRKY21
PaWRKY53
PaWRKY7
PaWRKY61
PaWRKY23
PaWRKY37
PaWRKY28
PaWRKY41
PaWRKY59
PaWRKY3
PaWRKY15
PaWRKY16
PaWRKY33
PaWRKY20
PaWRKY60
PaWRKY47
PaWRKY56

A

Ⅲ
Ⅱe
Ⅱd
Ⅱa
Ⅱb
Ⅱc
Ⅰ

5'
3'
0
200
400
600
800

Domain
Low complexity region
WRKY
Coiled coil region
Pfam:Plant zn clust
Signal peptide
Internal repeat 1
Transmembrane domain
SCOP:d1bg1a1

B

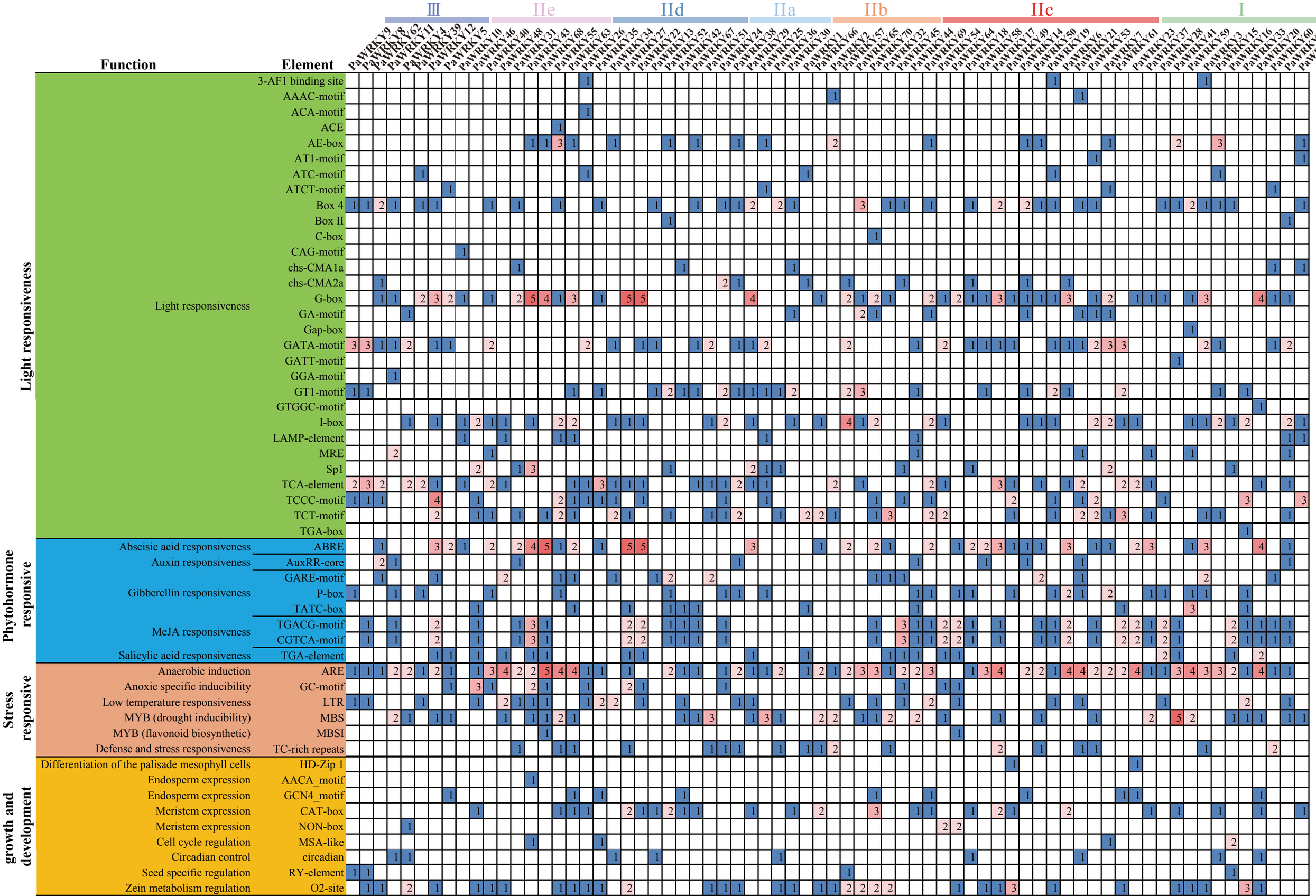
5'
3'
0
200
400
600
800

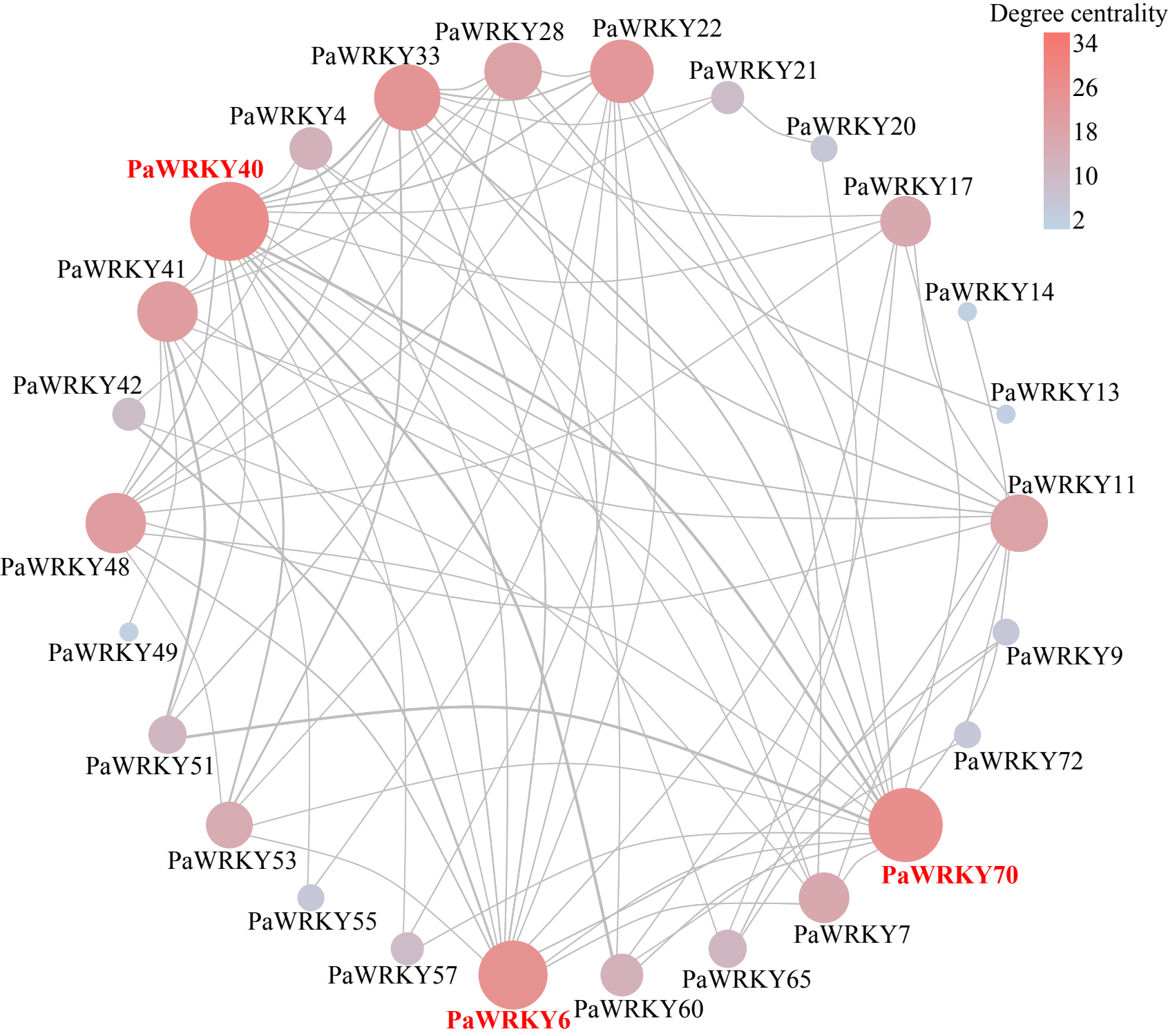
Motif
Motif 1
Motif 2
Motif 3
Motif 4
Motif 5
Motif 6
Motif 7
Motif 8
Motif 9
Motif 10
Motif 11
Motif 12
Motif 13
Motif 14
Motif 15
Motif 16
Motif 17
Motif 18
Motif 19
Motif 20

C

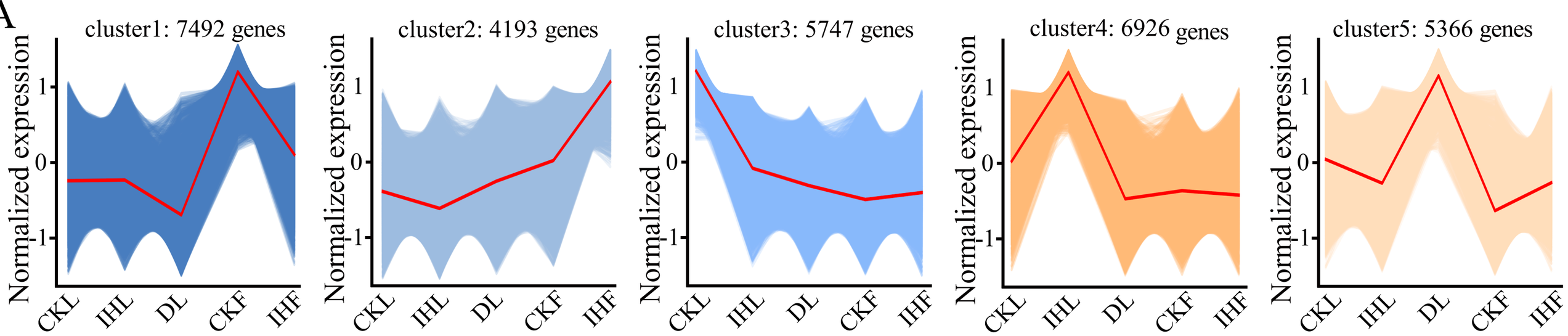
CDS

5'
3'
0
10000
20000

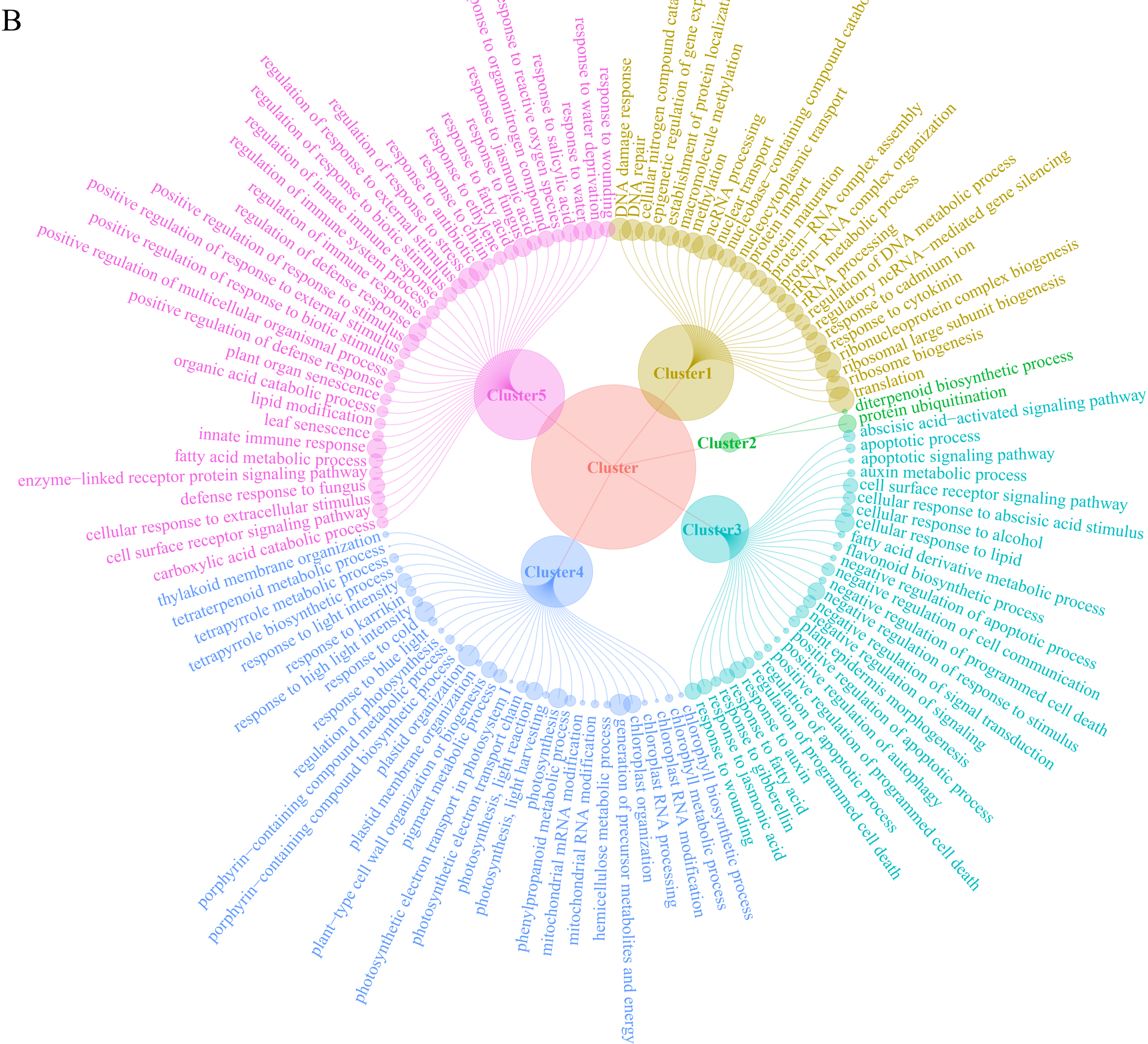

III
IIe
IId
IIa
IIb
IIc
I
Function
Element
PaWRKY9
PaWRKY8
PaWRKY62
PaWRKY11
PaWRKY4
PaWRKY39
PaWRKY12
PaWRKY5
PaWRKY10
PaWRKY46
PaWRKY40
PaWRKY48
PaWRKY31
PaWRKY43
PaWRKY68
PaWRKY55
PaWRKY63
PaWRKY26
PaWRKY35
PaWRKY34
PaWRKY27
PaWRKY22
PaWRKY13
PaWRKY52
PaWRKY42
PaWRKY67
PaWRKY51
PaWRKY24
PaWRKY38
PaWRKY29
PaWRKY25
PaWRKY36
PaWRKY30
PaWRKY71
PaWRKY66
PaWRKY52
PaWRKY57
PaWRKY65
PaWRKY70
PaWRKY32
PaWRKY45
PaWRKY44
PaWRKY69
PaWRKY54
PaWRKY64
PaWRKY18
PaWRKY58
PaWRKY17
PaWRKY49
PaWRKY14
PaWRKY50
PaWRKY19
PaWRKY6
PaWRKY21
PaWRKY53
PaWRKY7
PaWRKY61
PaWRKY23
PaWRKY37
PaWRKY28
PaWRKY41
PaWRKY59
PaWRKY3
PaWRKY15
PaWRKY16
PaWRKY33
PaWRKY20
PaWRKY60
PaWRKY47
PaWRKY56
Light responsiveness
3-AF1 binding site
AAAC-motif
ACA-motif
ACE
AE-box
AT1-motif
ATC-motif
ATCT-motif
Box 4
Box II
C-box
CAG-motif
chs-CMA1a
chs-CMA2a
G-box
GA-motif
Gap-box
GATA-motif
GATT-motif
GGA-motif
GT1-motif
GTGGC-motif
I-box
LAMP-element
MRE
Sp1
TCA-element
TCCC-motif
TCT-motif
TGA-box
Phytohormone responsive
Abscisic acid responsiveness
ABRE
Auxin responsiveness
AuxRR-core
Gibberellin responsiveness
GARE-motif
P-box
TATC-box
MeJA responsiveness
TGACG-motif
CGTCA-motif
Salicylic acid responsiveness
TGA-element
Stress responsive
Anaerobic induction
ARE
Anoxic specific inducibility
GC-motif
Low temperature responsiveness
LTR
MYB (drought inducibility)
MBS
MYB (flavonoid biosynthetic)
MBSI
Defense and stress responsiveness
TC-rich repeats
Plant growth and development
Differentiation of the palisade mesophyll cells
HD-Zip 1
Endosperm expression
AACA_motif
Endosperm expression
GCN4_motif
Meristem expression
CAT-box
Meristem expression
NON-box
Cell cycle regulation
MSA-like
Circadian control
circadian
Seed specific regulation
RY-element
Zein metabolism regulation
O2-site


Degree centrality
34
26
18
10
2
PaWRKY28
PaWRKY22
PaWRKY33
PaWRKY21
PaWRKY4
PaWRKY20
PaWRKY40
PaWRKY17
PaWRKY41
PaWRKY14
PaWRKY42
PaWRKY13
PaWRKY11
PaWRKY48
PaWRKY49
PaWRKY9
PaWRKY51
PaWRKY72
PaWRKY53
PaWRKY70
PaWRKY55
PaWRKY7
PaWRKY57
PaWRKY65
PaWRKY6
PaWRKY60

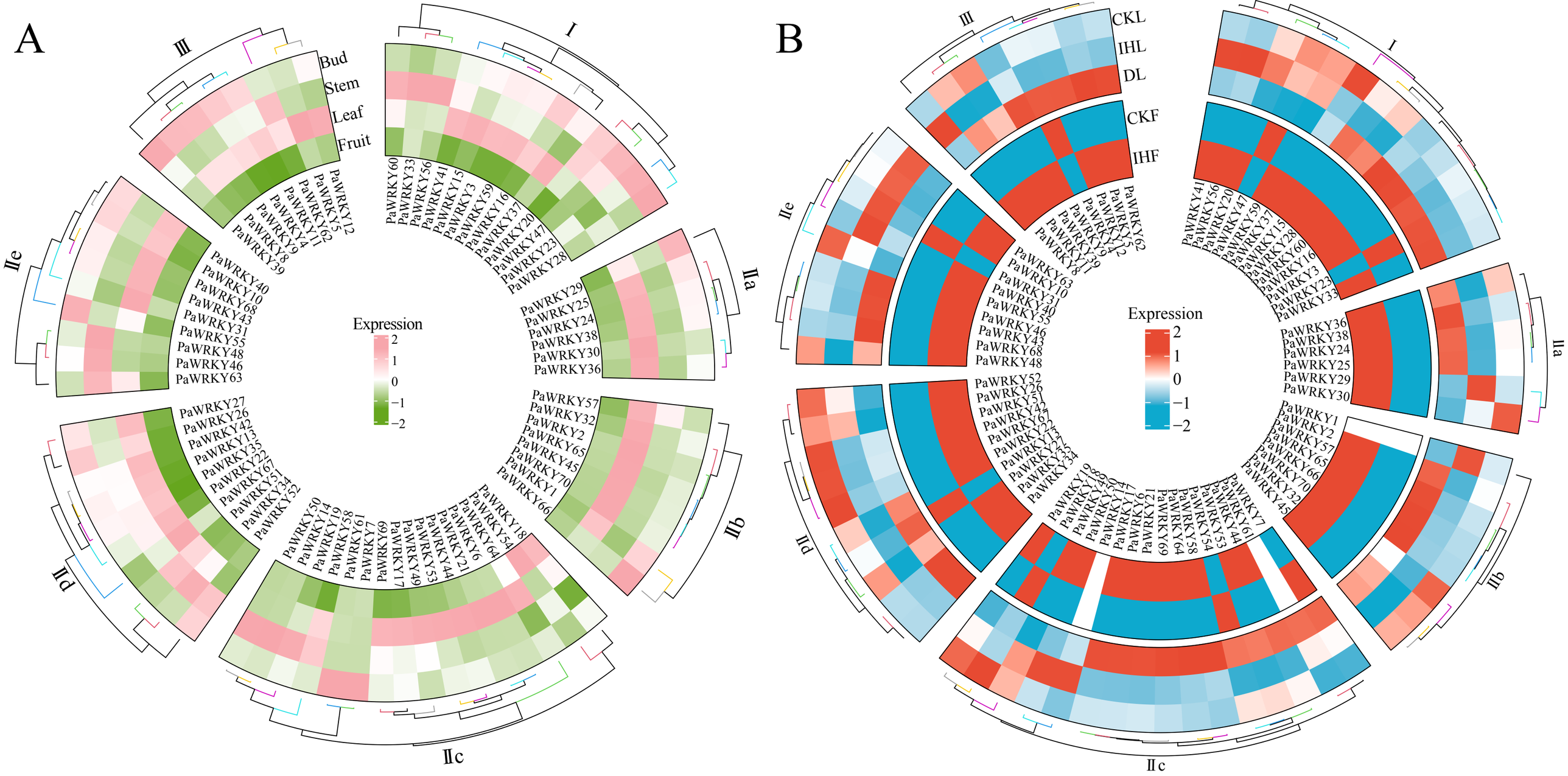
A


cluster1: 7492 genes
cluster2: 4193 genes
cluster3: 5747 genes
cluster4: 6926 genes
cluster5: 5366 genes
Normalized expression
1
0
−1
CKL
IHL
DL
CKF
IHF

B


Cluster
Cluster1
Cluster2
Cluster3
Cluster4
Cluster5
DNA damage response
DNA repair
cellular nitrogen compound catabolic process
epigenetic regulation of gene expression
establishment of protein localization to organelle
macromolecule methylation
methylation
ncRNA processing
nuclear transport
nucleobase−containing compound catabolic process
nucleocytoplasmic transport
protein import
protein maturation
protein−RNA complex assembly
protein−RNA complex organization
rRNA metabolic process
rRNA processing
regulation of DNA metabolic process
regulatory ncRNA−mediated gene silencing
response to cadmium ion
response to cytokinin
ribonucleoprotein complex biogenesis
ribosomal large subunit biogenesis
ribosome biogenesis
translation
diterpenoid biosynthetic process
protein ubiquitination
abscisic acid−activated signaling pathway
apoptotic process
apoptotic signaling pathway
auxin metabolic process
cell surface receptor signaling pathway
cellular response to abscisic acid stimulus
cellular response to alcohol
cellular response to lipid
fatty acid derivative metabolic process
flavonoid biosynthetic process
negative regulation of apoptotic process
negative regulation of cell communication
negative regulation of programmed cell death
negative regulation of response to stimulus
negative regulation of signal transduction
negative regulation of signaling
plant epidermis morphogenesis
positive regulation of apoptotic process
positive regulation of autophagy
positive regulation of programmed cell death
regulation of apoptotic process
regulation of programmed cell death
response to auxin
response to fatty acid
response to gibberellin
response to jasmonic acid
response to wounding
chlorophyll biosynthetic process
chlorophyll metabolic process
chloroplast RNA modification
chloroplast RNA processing
chloroplast organization
generation of precursor metabolites and energy
hemicellulose metabolic process
mitochondrial RNA modification
mitochondrial mRNA modification
phenylpropanoid metabolic process
photosynthesis
photosynthesis, light harvesting
photosynthesis, light reaction
photosynthetic electron transport chain
photosynthetic electron transport in photosystem I
pigment metabolic process
plant−type cell wall organization or biogenesis
plastid membrane organization
plastid organization
regulation of photosynthesis
regulation of response to stimulus
response to light intensity
response to blue light
response to cold
response to high light intensity
response to karrikin
tetrapyrrole biosynthetic process
tetrapyrrole metabolic process
tetraterpenoid metabolic process
thylakoid membrane organization
carboxylic acid catabolic process
cell surface receptor signaling pathway
cellular response to extracellular stimulus
defense response to fungus
enzyme−linked receptor protein signaling pathway
fatty acid metabolic process
innate immune response
leaf senescence
lipid modification
organic acid catabolic process
plant organ senescence
positive regulation of defense response
positive regulation of multicellular organismal process
positive regulation of response to biotic stimulus
positive regulation of response to external stimulus
regulation of defense response
regulation of immune response
regulation of immune system process
regulation of response to biotic stimulus
regulation of response to external stimulus
response to antibiotic
response to chitin
response to ethylene
response to fatty acid
response to jasmonic acid
response to organonitrogen compound
response to reactive oxygen species
response to salicylic acid
response to water
response to water deprivation
response to wounding


A
generation of precursor metabolites and energy
photosynthesis
plastid organization
photosynthesis, light reaction
response to cold
polysaccharide metabolic process
response to water deprivation
plant−type cell wall organization or biogenesis
carbohydrate biosynthetic process
chloroplast organization
cell wall biogenesis
response to cytokinin
polysaccharide biosynthetic process
response to heat
response to toxic substance
−log10(p-value)
40
30
20
10
CKLvsIHL
secondary metabolic process
regulation of response to stress
response to fungus
response to antibiotic
response to auxin
secondary metabolite biosynthetic process
defense response to fungus
regulation of defense response
response to salicylic acid
glucan metabolic process
plant organ senescence
leaf senescence
positive regulation of immune response
positive regulation of innate immune response
response to molecule of bacterial origin
−log10(p-value)
5
4
CKLvsDL
inorganic ion transmembrane transport
response to water
response to water deprivation
innate immune response
regulation of response to stress
plant organ morphogenesis
response to auxin
response to wounding
response to antibiotic
response to organonitrogen compound
response to jasmonic acid
cellular response to abscisic acid stimulus
response to salicylic acid
abscisic acid−activated signaling pathway
terpenoid metabolic process
−log10(p-value)
11
9
7
5
CKFvsIHF
0
25
50
75
100
Gene Count
GO Term
B
regulation of hormone levels
response to wounding
regulation of response to stress
plant organ morphogenesis
response to auxin
response to jasmonic acid
hormone metabolic process
regulation of defense response
response to salicylic acid
plant−type cell wall organization or biogenesis
response to reactive oxygen species
post−embryonic plant morphogenesis
cellular response to abscisic acid stimulus
multicellular organismal reproductive process
jasmonic acid mediated signaling pathway
−log10(p-value)
8
6
4
CKLvsIHL
regulation of hormone levels
plant−type cell wall organization or biogenesis
response to auxin
response to wounding
cell wall biogenesis
hormone metabolic process
response to jasmonic acid
carbohydrate biosynthetic process
flavonoid metabolic process
phenylpropanoid metabolic process
hormone transport
phenylpropanoid biosynthetic process
flavonoid biosynthetic process
fatty acid biosynthetic process
auxin transport
−log10(p-value)
12
9
6
3
CKLvsDL
regulation of hormone levels
monocarboxylic acid biosynthetic process
microtubule−based process
fatty acid metabolic process
DNA replication
DNA−templated DNA replication
fatty acid biosynthetic process
microtubule−based movement
terpenoid biosynthetic process
pyridine−containing compound metabolic process
nicotinamide nucleotide metabolic process
pyruvate metabolic process
nucleoside phosphate catabolic process
nucleotide catabolic process
purine nucleotide catabolic process
−log10(p-value)
8
7
6
5
4
CKFvsIHF
0
25
50
75
100
Gene Count
GO Term
C
Photosynthesis proteins
Photosynthesis
Starch and sucrose metabolism
Carbon fixation in photosynthetic organisms
Tryptophan metabolism
Glycolysis / Gluconeogenesis
Metabolism of xenobiotics by cytochrome P450
Glyoxylate and dicarboxylate metabolism
Steroid hormone biosynthesis
Photosynthesis − antenna proteins
Biosynthesis of various plant secondary metabolites
Drug metabolism − cytochrome P450
Fructose and mannose metabolism
Methane metabolism
Arachidonic acid metabolism
Gene Count
20
30
40
50
60
−log10(p-value)
40
30
20
10
CKLvsIHL
Cytochrome P450
Plant hormone signal transduction
MAPK signaling pathway − plant
Starch and sucrose metabolism
Phenylpropanoid biosynthesis
Metabolism of xenobiotics by cytochrome P450
Tryptophan metabolism
Biosynthesis of various plant secondary metabolites
Monoterpenoid biosynthesis
Drug metabolism − other enzymes
Glutathione metabolism
Glycine, serine and threonine metabolism
Cyanoamino acid metabolism
Zeatin biosynthesis
Drug metabolism − cytochrome P450
Gene Count
20
30
40
50
−log10(p-value)
7
6
5
4
3
2
CKLvsDL
Glycosyltransferases
Plant hormone signal transduction
Cytochrome P450
MAPK signaling pathway − plant
Phenylpropanoid biosynthesis
Starch and sucrose metabolism
Metabolism of xenobiotics by cytochrome P450
Zeatin biosynthesis
Monoterpenoid biosynthesis
Biosynthesis of various plant secondary metabolites
Galactose metabolism
alpha−Linolenic acid metabolism
Retinol metabolism
Arachidonic acid metabolism
Cutin, suberine and wax biosynthesis
Calcium signaling pathway
Gene Count
20
40
60
−log10(p-value)
8
6
4
CKFvsIHF
10
20
30
40
50
60
70
80
Gene Count
KEGG Pathway
D
Cytochrome P450
Phenylpropanoid biosynthesis
Starch and sucrose metabolism
Lipid biosynthesis proteins
Amino sugar and nucleotide sugar metabolism
MAPK signaling pathway − plant
Metabolism of xenobiotics by cytochrome P450
Glycerophospholipid metabolism
Biosynthesis of unsaturated fatty acids
Biosynthesis of various plant secondary metabolites
Drug metabolism − cytochrome P450
alpha−Linolenic acid metabolism
Monoterpenoid biosynthesis
Glycerolipid metabolism
Retinol metabolism
Gene Count
30
40
50
60
−log10(p-value)
9
7
5
3
CKLvsIHL
Glycosyltransferases
Plant hormone signal transduction
Cytochrome P450
Phenylpropanoid biosynthesis
Starch and sucrose metabolism
Lipid biosynthesis proteins
Photosynthesis proteins
Photosynthesis
Flavonoid biosynthesis
Biosynthesis of unsaturated fatty acids
Fatty acid elongation
Retinol metabolism
Cutin, suberine and wax biosynthesis
Fatty acid biosynthesis
Sesquiterpenoid and triterpenoid biosynthesis
Gene Count
20
40
60
−log10(p-value)
10
8
6
4
CKLvsDL
DNA replication proteins
Lipid biosynthesis proteins
Glycolysis / Gluconeogenesis
Terpenoid backbone biosynthesis
Fatty acid biosynthesis
Pyruvate metabolism
Monoterpenoid biosynthesis
Carbon fixation in photosynthetic organisms
Flavonoid biosynthesis
Sesquiterpenoid and triterpenoid biosynthesis
Fructose and mannose metabolism
Biosynthesis of unsaturated fatty acids
Phenylalanine metabolism
DNA replication
Biotin metabolism
Gene Count
20
30
40
−log10(p-value)
12.5
7.5
2.5
CKFvsIHF
20
40
60
80
Gene Count


A
Bud
Stem
Leaf
Fruit
Expression
2
1
0
−1
−2
Ⅰ
Ⅱa
Ⅱb
Ⅱc
Ⅱd
Ⅱe
Ⅲ
PaWRKY60
PaWRKY33
PaWRKY56
PaWRKY41
PaWRKY15
PaWRKY3
PaWRKY59
PaWRKY16
PaWRKY37
PaWRKY20
PaWRKY47
PaWRKY23
PaWRKY28
PaWRKY29
PaWRKY25
PaWRKY24
PaWRKY38
PaWRKY30
PaWRKY36
PaWRKY57
PaWRKY32
PaWRKY2
PaWRKY65
PaWRKY45
PaWRKY70
PaWRKY1
PaWRKY66
PaWRKY18
PaWRKY54
PaWRKY64
PaWRKY6
PaWRKY21
PaWRKY44
PaWRKY53
PaWRKY49
PaWRKY17
PaWRKY69
PaWRKY7
PaWRKY61
PaWRKY58
PaWRKY19
PaWRKY14
PaWRKY50
PaWRKY27
PaWRKY26
PaWRKY42
PaWRKY13
PaWRKY35
PaWRKY22
PaWRKY67
PaWRKY51
PaWRKY34
PaWRKY52
PaWRKY40
PaWRKY10
PaWRKY68
PaWRKY43
PaWRKY31
PaWRKY55
PaWRKY48
PaWRKY46
PaWRKY63
PaWRKY62
PaWRKY5
PaWRKY12
PaWRKY11
PaWRKY4
PaWRKY9
PaWRKY8
PaWRKY39
B
CKL
IHL
DL
CKF
IHF
Expression
2
1
0
−1
−2
Ⅰ
Ⅱa
Ⅱb
Ⅱc
Ⅱd
Ⅱe
Ⅲ
PaWRKY41
PaWRKY56
PaWRKY20
PaWRKY47
PaWRKY59
PaWRKY37
PaWRKY15
PaWRKY28
PaWRKY60
PaWRKY16
PaWRKY3
PaWRKY23
PaWRKY33
PaWRKY36
PaWRKY38
PaWRKY24
PaWRKY25
PaWRKY29
PaWRKY30
PaWRKY1
PaWRKY2
PaWRKY57
PaWRKY65
PaWRKY66
PaWRKY70
PaWRKY32
PaWRKY45
PaWRKY7
PaWRKY61
PaWRKY44
PaWRKY53
PaWRKY54
PaWRKY58
PaWRKY64
PaWRKY69
PaWRKY6
PaWRKY21
PaWRKY17
PaWRKY14
PaWRKY50
PaWRKY49
PaWRKY18
PaWRKY19
PaWRKY52
PaWRKY26
PaWRKY51
PaWRKY42
PaWRKY67
PaWRKY22
PaWRKY13
PaWRKY35
PaWRKY34
PaWRKY63
PaWRKY10
PaWRKY31
PaWRKY40
PaWRKY55
PaWRKY46
PaWRKY43
PaWRKY68
PaWRKY48
PaWRKY62
PaWRKY5
PaWRKY12
PaWRKY4
PaWRKY11
PaWRKY9
PaWRKY39
PaWRKY8

A

Cluster dendrogram

1.0 0.8 0.6 0.4 0.2 0.0

Height

Module colors

B

Module−trait relationships

	CKL	IHL	DL	CKF	IHF
MEmaroon	−0.097(0.7)	−0.12(0.7)	0.56(0.03)	−0.19(0.5)	−0.16(0.6)
MElightpink4	−0.11(0.7)	−0.11(0.7)	0.57(0.03)	−0.17(0.5)	−0.17(0.5)
MElavenderblush3	−0.13(0.6)	−0.11(0.7)	0.58(0.02)	−0.17(0.5)	−0.17(0.5)
MEcyan	−0.29(0.3)	−0.47(0.08)	0.72(0.003)	−0.44(0.1)	0.48(0.07)
MEsaddlebrown	−0.12(0.7)	−0.19(0.5)	0.98(5e−11)	−0.38(0.2)	−0.29(0.3)
MEthistle3	−0.16(0.6)	−0.12(0.7)	−0.15(0.6)	−0.086(0.8)	0.52(0.05)
MEskyblue1	−0.12(0.7)	−0.12(0.7)	−0.15(0.6)	−0.16(0.6)	0.55(0.03)
MEblue2	−0.13(0.7)	−0.14(0.6)	−0.16(0.6)	−0.13(0.6)	0.56(0.03)
MEdarkseagreen4	−0.3(0.3)	−0.34(0.2)	−0.15(0.6)	−0.21(0.5)	0.99(2e−12)
MEantiquewhite4	−0.12(0.7)	−0.13(0.6)	−0.2(0.5)	0.6(0.02)	−0.15(0.6)
MEcoral2	−0.11(0.7)	−0.14(0.6)	−0.18(0.5)	0.56(0.03)	−0.13(0.6)
MEskyblue2	−0.14(0.6)	−0.14(0.6)	−0.18(0.5)	0.56(0.03)	−0.093(0.7)
MEdarkgreen	0.52(0.05)	−0.35(0.2)	−0.51(0.05)	0.69(0.004)	−0.34(0.2)
MElightsteelblue1	−0.3(0.3)	−0.32(0.2)	−0.5(0.06)	0.86(4e−05)	0.25(0.4)
MEplum2	−0.061(0.8)	0.53(0.04)	−0.15(0.6)	−0.17(0.5)	−0.15(0.6)
MEdarkmagenta	−0.13(0.6)	0.56(0.03)	−0.14(0.6)	−0.16(0.6)	−0.13(0.6)
MEbrown4	−0.29(0.3)	0.67(0.006)	−0.61(0.02)	0.51(0.05)	−0.29(0.3)
MEdarkgrey	−0.11(0.7)	0.57(0.03)	−0.2(0.5)	−0.16(0.6)	−0.1(0.7)
MEskyblue3	0.3(0.3)	0.55(0.03)	−0.56(0.03)	−0.64(0.01)	0.35(0.2)
MEblack	0.58(0.02)	0.64(0.01)	−0.4(0.1)	−0.42(0.1)	−0.41(0.1)
MEcoral1	−0.15(0.6)	0.99(3e−13)	−0.21(0.4)	−0.31(0.3)	−0.32(0.2)
MElightyellow	0.62(0.01)	−0.12(0.7)	−0.16(0.6)	−0.17(0.5)	−0.16(0.6)
MElightgreen	0.55(0.03)	−0.11(0.7)	−0.14(0.6)	−0.16(0.6)	−0.15(0.6)
MEdarkolivegreen	0.63(0.01)	−0.32(0.3)	−0.21(0.4)	−0.64(0.01)	0.54(0.04)
MEorange	0.59(0.02)	−0.12(0.7)	−0.15(0.6)	−0.16(0.6)	−0.16(0.6)
MElightcyan	0.91(2e−06)	−0.24(0.4)	0.14(0.6)	−0.46(0.08)	−0.36(0.2)
MEsalmon2	0.79(4e−04)	−0.19(0.5)	−0.2(0.5)	−0.19(0.5)	−0.21(0.5)
MEgrey	0.089(0.8)	0.092(0.7)	−0.29(0.3)	−0.35(0.2)	0.46(0.09)

1 0.5 0 −0.5 −1

C

regulation of reproductive process
regulation of developmental growth
meristem maintenance
meristem development
response to virus
leaf senescence
plant organ senescence
defense response to fungus
response to fungus
auxin−activated signaling pathway
plant−type cell wall organization
pigment metabolic process
photosynthesis, light reaction
response to cytokinin
cell wall biogenesis
chloroplast organization
carbohydrate biosynthetic process
photosynthesis
flavonoid biosynthetic process
jasmonic acid mediated signaling pathway
regulation of cellular response to alcohol
regulation of response to alcohol
regulation of abscisic acid−activated signaling pathway
cellular response to jasmonic acid stimulus
plant−type cell wall biogenesis
photoperiodism
plant epidermis morphogenesis
phenylpropanoid biosynthetic process
phenylpropanoid metabolic process
response to gibberellin
response to ethylene
response to salicylic acid
plant−type cell wall organization or biogenesis
response to jasmonic acid
response to wounding
cellular response to acid chemical
regulation of post−embryonic development
cellular response to lipid

Description

CKL IHL DL CKF IHF

Group

Gene Count
25
50
75
100

−log10(*p*−value)
10
5

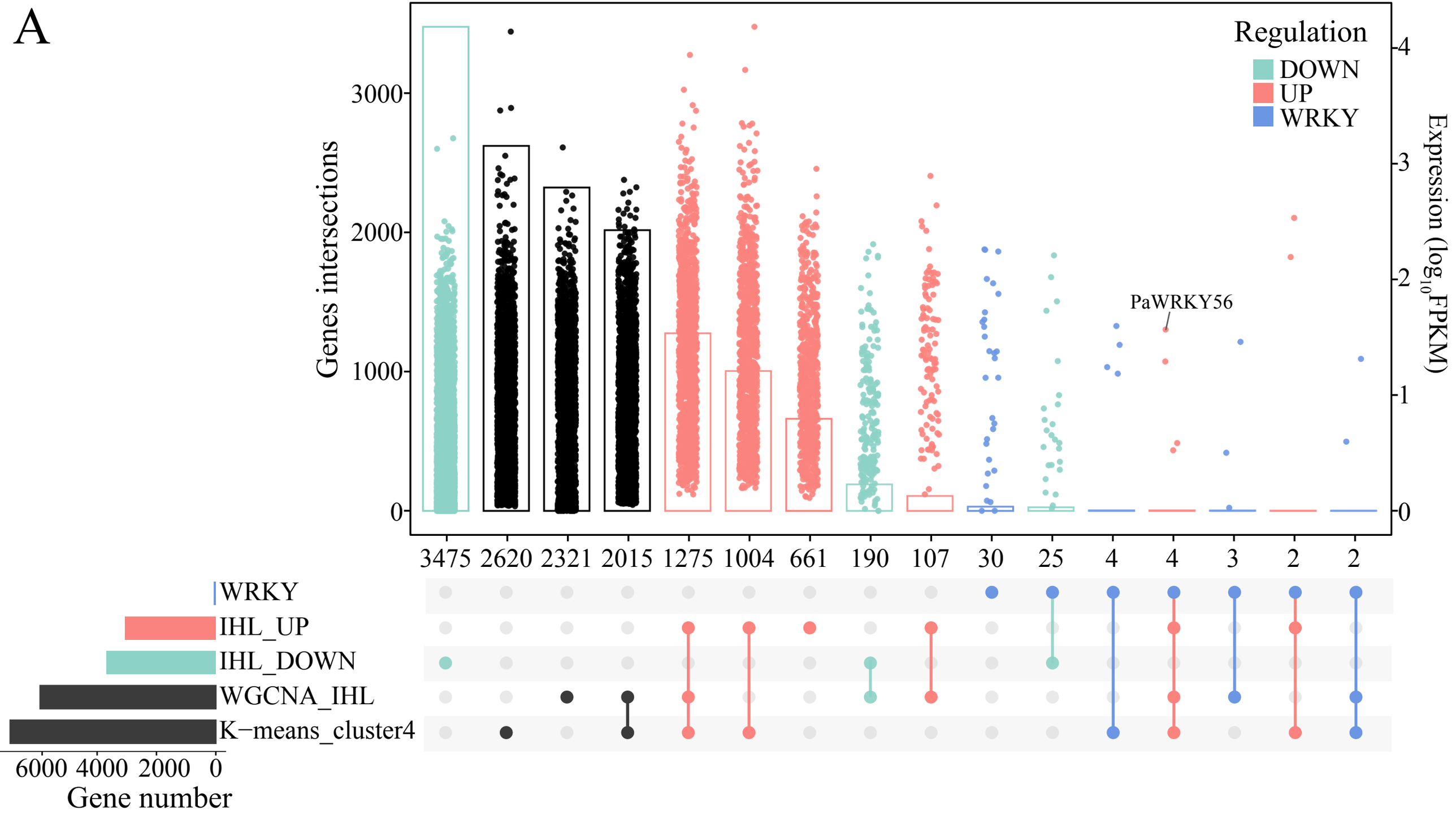
D

ko00940 ko01004 ko00982 ko00073 ko00941 ko01040 ko03110 ko00194 ko01007 ko00630 ko00195 ko00710 ko00520 ko00380 ko04016 ko00980 ko00983 ko00330 ko03011 ko03032 ko03012 ko00230 ko03082 ko03420 ko01009 ko00562 ko04020 ko03083 ko04710 ko04744

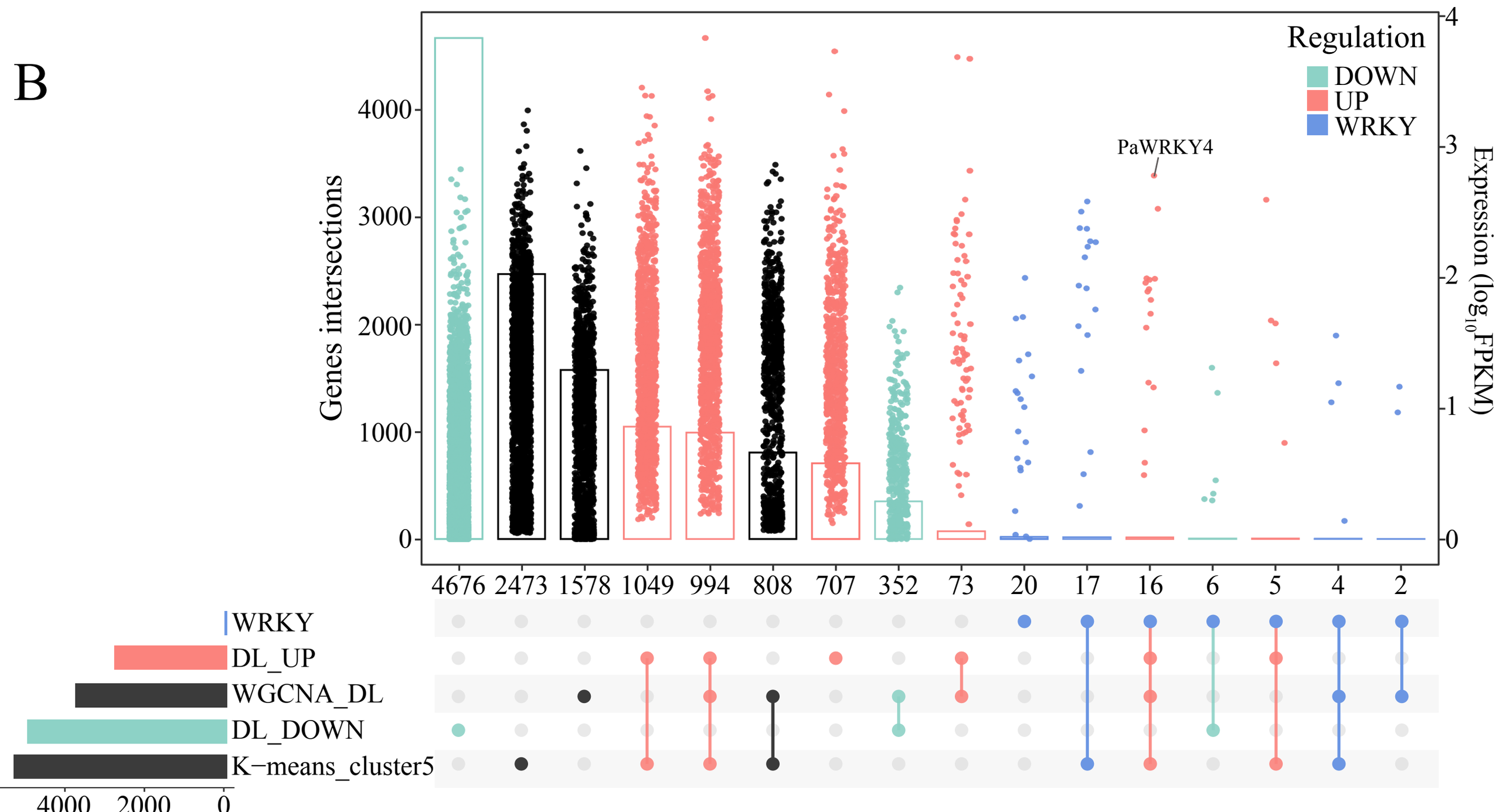
Group
CKL
IHL
DL
CKF
IHF

−log10(*p*−value)
15
10
5
0

A


Regulation
DOWN
UP
WRKY
Genes intersections
Expression (log10FPKM)
PaWRKY56
3475 2620 2321 2015 1275 1004 661 190 107 30 25 4 4 3 2 2
WRKY
IHL_UP
IHL_DOWN
WGCNA_IHL
K−means_cluster4
6000 4000 2000 0
Gene number

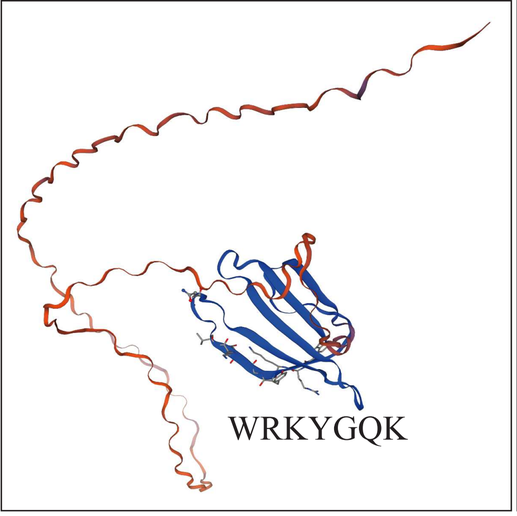
B


Regulation
DOWN
UP
WRKY
Genes intersections
Expression (log10FPKM)
PaWRKY4
4676 2473 1578 1049 994 808 707 352 73 20 17 16 6 5 4 2
WRKY
DL_UP
WGCNA_DL
DL_DOWN
K−means_cluster5
4000 2000 0
Gene number

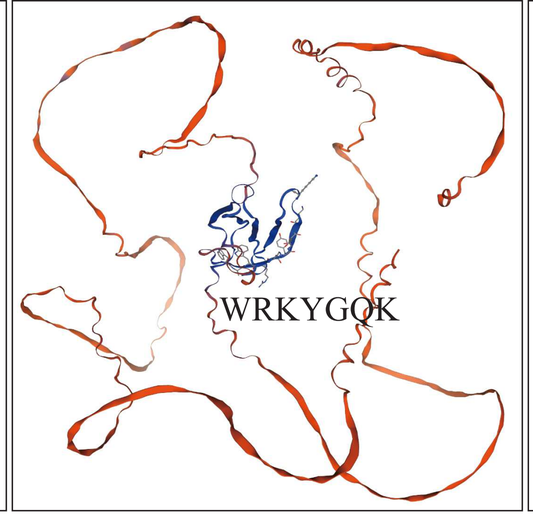
C

Regulation
DOWN
UP
WRKY
Genes intersections
Expression (log10FPKM)
PaWRKY26
2830 2560 2541 496 435 377 374 71 45 35 28 2 2 1 1 1
WRKY
WGCNA_IHF
IHF_DOWN
IHF_UP
K−means_cluster2
4000 2000 0
Gene number

PaWRKY4
PaWRKY12
PaWRKY14
PaWRKY24
PaWRKY35
PaWRKY40
PaWRKY42
PaWRKY56
PaWRKY57
PaWRKY69
qRT−PCR relative expression
RNA−seq expression level
Sample
CKL
IHL
DL
CKF
IHF
qRT−PCR
RNA−seq



PaWRKY14



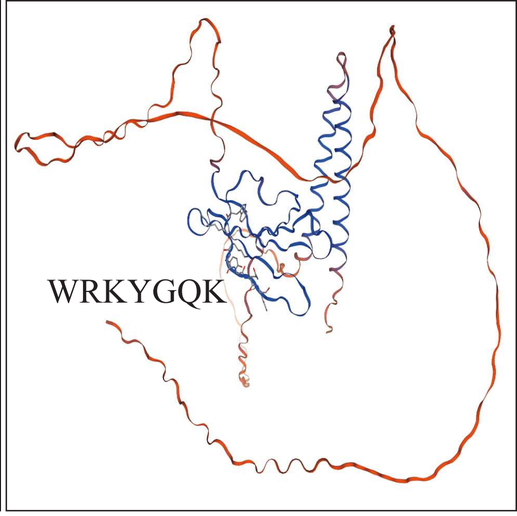
PaWRKY40



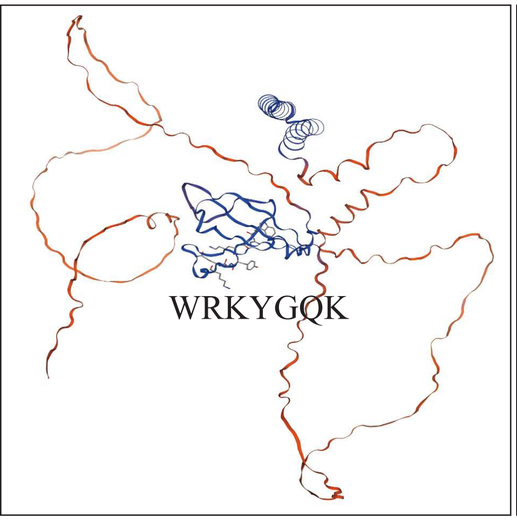
PaWRKY50



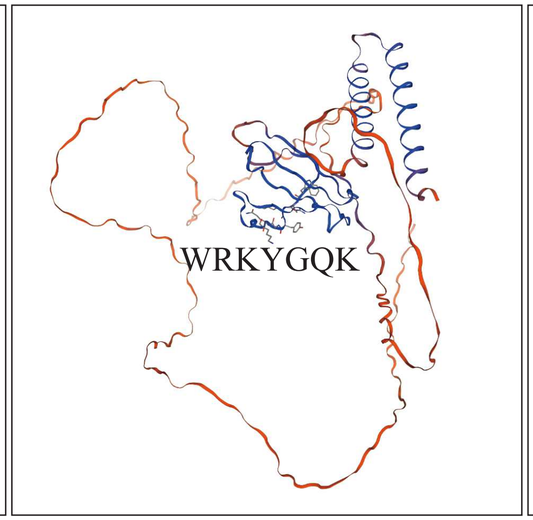
PaWRKY56



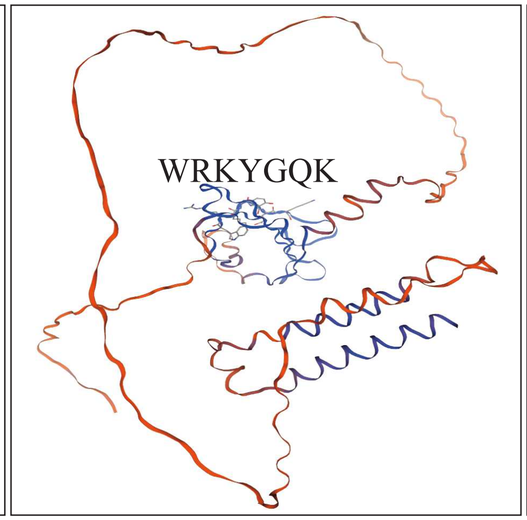
PaWRKY4



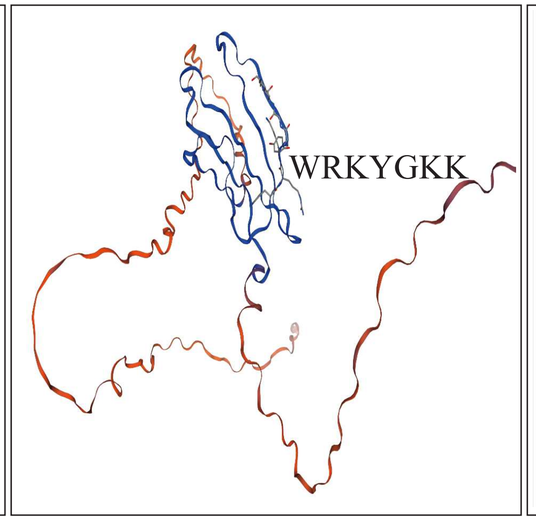
PaWRKY5



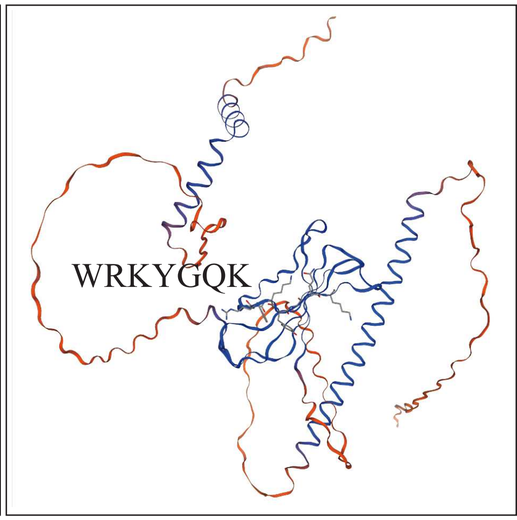
PaWRKY9



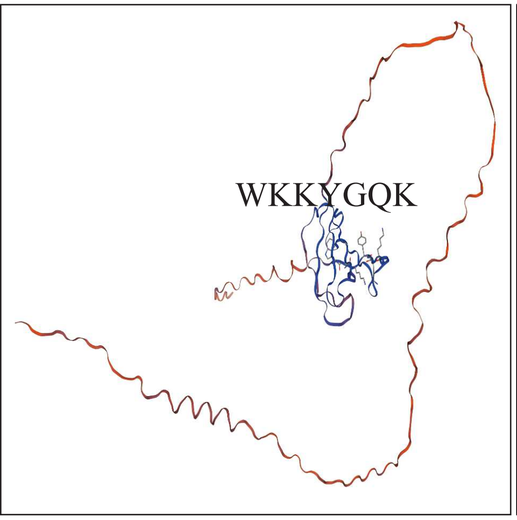
PaWRKY12



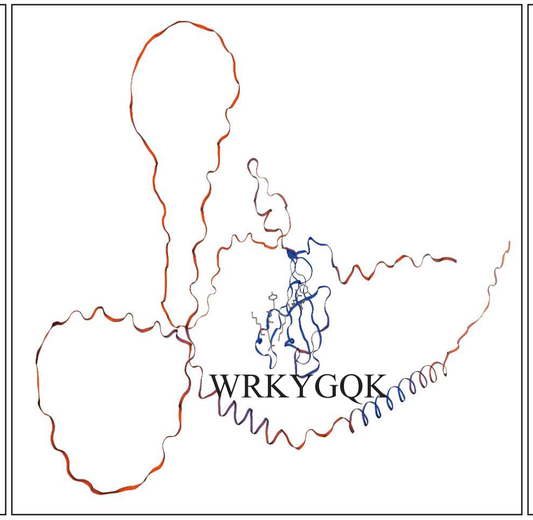
PaWRKY17



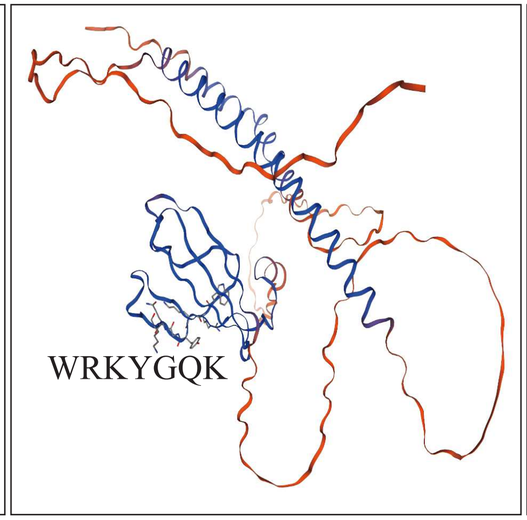
PaWRKY24



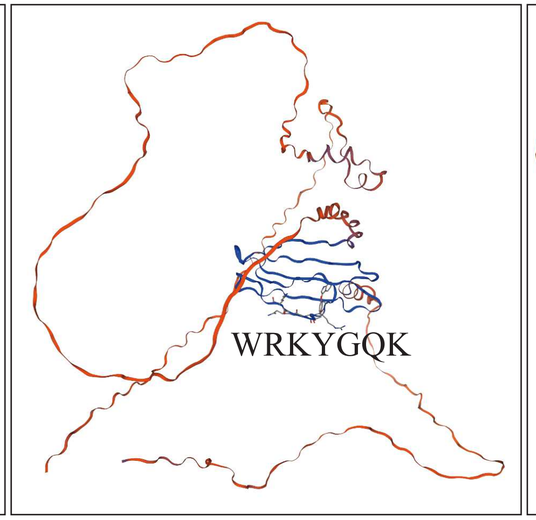
PaWRKY34



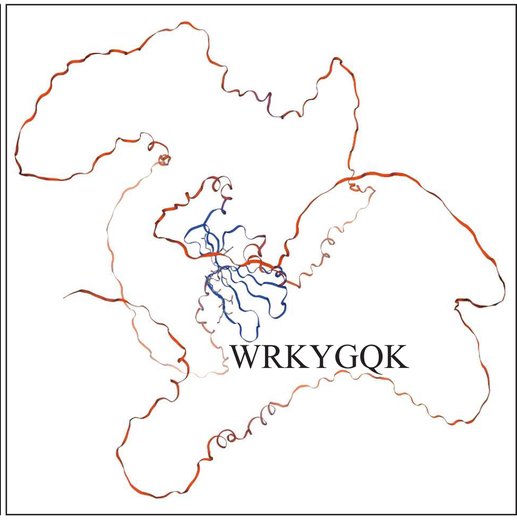
PaWRKY35



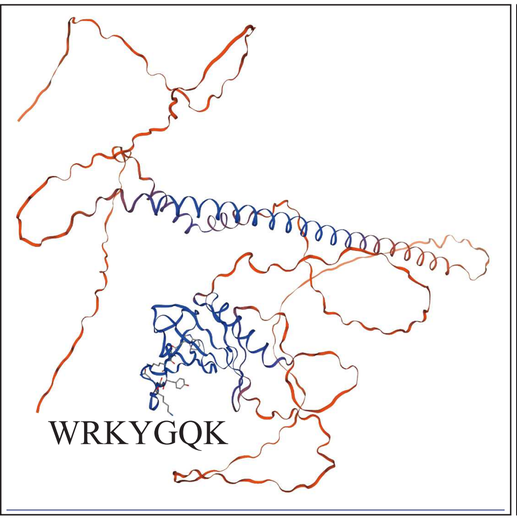
PaWRKY38



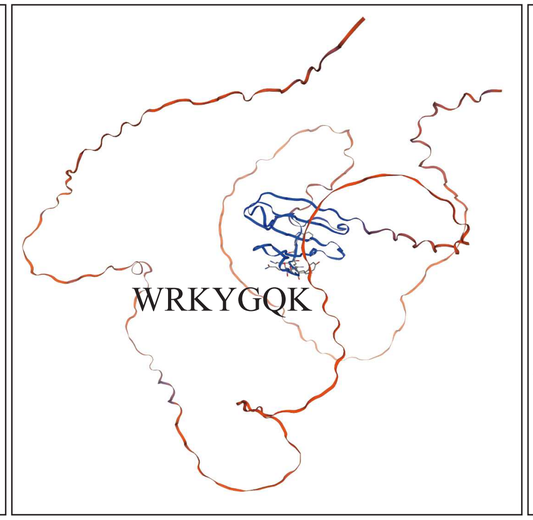
PaWRKY44



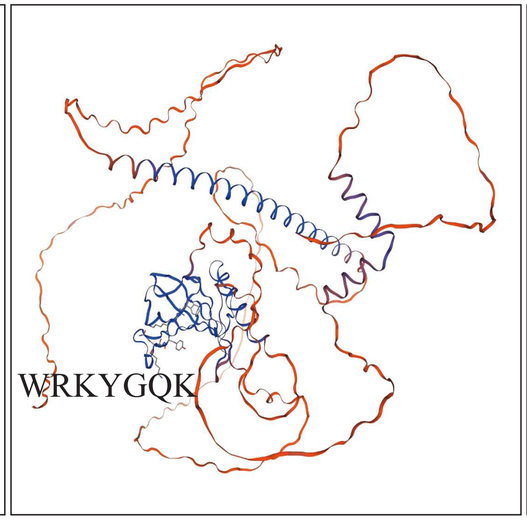
PaWRKY46



PaWRKY57



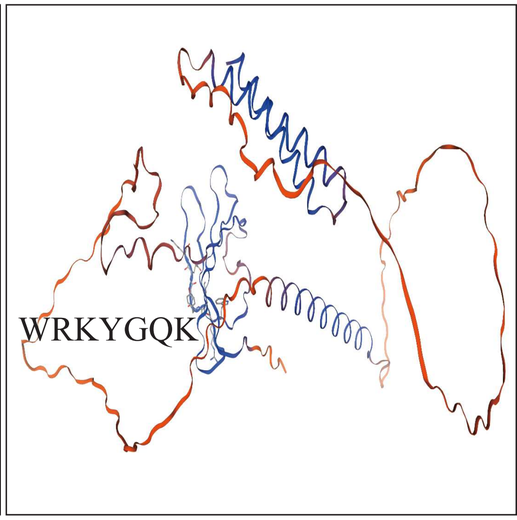
PaWRKY64



PaWRKY65



PaWRKY69



PaWRKY42

Supplementary Files

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

- SupplementaryMaterial1.docx
- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx