Genome-wide identification of WRKY transcription factors and their target genes associated with drought and salt stress tolerance in *Robinia pseudoacacia* L.

Sergey Alexandrovich Vishnyakov * and Pavel Andreevich Krylov

Federal State Budget Scientific Institution "Federal Scientific Centre of Agroecology, Complex Melioration and Protective Afforestation of the Russian Academy of Sciences", Volgograd, Russia

Abstract

Robinia pseudoacacia L. (black locust) is a valuable tree species widely used in afforestation and land reclamation. This species plays a crucial role in establishing resilient forest ecosystems in arid regions, contributing to soil erosion prevention and environmental improvement. The high adaptability of *R. pseudoacacia* makes it one of the key species for restoring degraded lands. In this study, a comprehensive bioinformatics analysis was conducted, including multiple sequence alignment, phylogenetic analysis, gene structure examination, and conserved motif identification of WRKY transcription factors. Special attention was given to the promoter regions of *R. pseudoacacia* genes associated with drought and salt stress responses to identify WRKY-recognized motifs (W-box) within their sequences. The results highlight the diversity of biological processes regulated by the WRKY transcription factor family. The findings provide valuable insights for selecting candidate genes that can be utilized in breeding new R. *pseudoacacia* cultivars with enhanced drought and salt stress tolerance.

Keywords: afforestation, drought tolerance, *Robinia pseudoacacia* L., salt tolerance, W-box, WRKY

Vishnyakov S. A. and P. A. Krylov.2025. 'Genome-wide identification of WRKY transcription factors and their target genes associated with drought and salt stress tolerance in *Robinia pseudoacacia* L.'. *Iranian Journal of Plant Physiology 15* (3), 5597-5603.

Introduction

WRKY transcription factor family in higher plants plays a key role in regulating the expression of various target genes involved in plant development and responses to biotic and abiotic stresses, thereby influencing plant resilience to environmental challenges such as drought, salinity, and high temperatures (Li et al., 2020; Wang et al., 2023a; Wu et al., 2024). Additionally, certain WRKY proteins are involved in the regulation of secondary metabolite biosynthesis, including phenolic compounds (Huang et al., 2022).

A key structural feature of WRKY proteins is the presence of the WRKY domain. This domain is approximately 60 amino acids long and contains the conserved WRKYGQK amino acid sequence along with a zinc finger motif (C_2H_2 or C_2HC) at the C-terminal region (Chen et al., 2017a). Based on the number of WRKY domains and the structure of their zinc fingers, WRKY proteins are classified into three distinct groups (Goyal et al., 2023). The WRKYGQK structure and zinc finger motifs play a crucial role in ensuring the high binding affinity of WRKY transcription factors to the consensus cis-

* *Corresponding Author*
E-mail Address: vishnyskov-s@vfanc.ru
Received: March, 2025
Accepted: June, 2025

regulatory DNA element 5'-C/TTGACT/C-3', known as the W-box. The presence of this region in the promoter sequences of target genes enables WRKY transcription factors to regulate their expression (Banerjee and Roychoudhury, 2015; Chen et al., 2019; Wani et al., 2021).

The number of genes encoding this transcription factor varies significantly among plant species. In most cases, plant genomes contain between 50 and 130 WRKY genes. Even closely related species can exhibit substantial differences in WRKY gene numbers. For example, Glycine max possesses 296 WRKY genes, the highest number among all plant species according to Plant TFDB, whereas its close relative Glycine soja has 169 WRKY genes. This variation in WRKY gene numbers across different plant species reflects the complexity of the evolutionary history of this transcription factor family.

Robinia pseudoacacia L. is one of the most valuable and fast-growing tree species in arid regions. Owing to its high drought tolerance, resistance to saline soils, and rapid growth, this species is widely used in artificial forest plantations as a primary forest-forming species (Li et al., 2020). *R. pseudoacacia* plays a crucial role in establishing resilient forest ecosystems, contributing to soil erosion prevention and environmental improvement in areas with arid climates and soil salinity. Its ability to adapt to harsh environmental conditions makes it a key species for afforestation and the restoration of degraded lands (KUZMIN and KRYLOV; Schwärzel et al., 2020).

Recently, the genome sequence and assembly of R. *pseudoacacia* were published (Wang et al., 2023b), providing new opportunities for a detailed structural analysis of WRKY genes and proteins in this species. This study aimed to conduct a comprehensive bioinformatic analysis, including multiple sequence alignment, phylogenetic analysis, gene structure examination, and conserved motif identification. Additionally, WRKY proteins were classified into groups based on their structural features. Special attention was given to the promoter regions of *R. pseudoacacia* genes associated with drought and salt stress responses to identify W-box motifs.

Materials and Methods

The complete genome sequence of *R. pseudoacacia*, along with transcript and protein sequences, was obtained from Figshare (doi.org/10.6084/m9.figshare.23301668) (Wang et al., 2023a).

The hidden Markov model (HMM) profile for the WRKY domain (PF03106) was retrieved from the Pfam protein family database (http://pfam.xfam.org/) and was utilized to identify putative WRKY proteins using the HMMER Search program (http://hmmer.janelia.org/, version 3.4) with an E-value threshold of less than 1e−10. All non-redundant protein sequences containing complete WRKY domains were selected as putative WRKY proteins. The integrity of WRKY domains was verified using the InterPro Scan tool (Jones et al., 2014)and the Batch CD-Search Tool (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi) (Wang J. et al., 2023a).

Protein properties, including sequence length (aa length), molecular weight (MW), isoelectric point (IP), instability index (II), aliphatic index (AI), and grand average of hydropathy (GRAVY), were calculated using the ExPASy web service (http://web.expasy.org/protparam/).

The structural information of each gene was visualized using GSDS2.0 (Hu et al., 2015). Conserved protein motifs were predicted using MEME in the classic motif discovery mode (Bailey et al., 2015).

Multiple sequence alignments were performed using MAFFT with the BLOSUM62 scoring matrix. A phylogenetic tree was constructed using IQ-TREE with exhaustive tree search parameters for each model, ultimately employing the JTT+F+R5 model (Minh et al., 2020).

Gene ontology (GO) terms were obtained using PANNZER2 (Protein ANNotation with Z-scoRE) (Törönen and Holm, 2022).

The search for W-box motifs in the promoter region, spanning -1000 to +100 bp from the transcription start site, was conducted using FIMO (version 5.5.7) (Grant et al., 2011). The W-box motif sequences were retrieved from the JASPAR

database (Rauluseviciute et al., 2024). GO term visualization was performed using (Supek et al., 2011).

Results

A total of 84 protein sequences containing a complete conserved WRKY domain was identified in the *R. pseudoacacia* genome sequence. The length of RpWRKY proteins ranged from 156 to 761 amino acids, with an average of 386 amino acids. The isoelectric point varied between 4.88 and 9.81 while the relative molecular weight ranged from 18.06 to 83.15 kDa. Sequence analysis using the PSORT program indicated that all 84 RpWRKY proteins are localized in the cell nucleus.

A phylogenetic tree was constructed based on multiple sequence alignment of RpWRKY proteins (Fig. I). The WRKY proteins were classified into seven clusters, divided into three groups: Group I, Group II (further subdivided into five subgroups), and Group III. Group II contained the highest number of RpWRKY proteins (59 members), with subgroup IIa comprising 5 members, IIb – 17, IIc – 9, IId – 16, and IIe – 12. Subgroups IIa and IIb were found to cluster together, as did subgroups IIc and IIe while subgroup IId had the highest similarity to Group I, which consists of 14 proteins. Group III, containing 10 proteins, was positioned adjacent to the IIc and IIe branches. One protein (CHAGGene29207) could not be assigned to any specific group but showed the highest similarity to Group III.

All *R. pseudoacacia* WRKY sequences had high similarity and conservation of the WRKY domain. Group I RpWRKYs contained two conserved WRKYGQK motifs at both the N- and C-terminal regions, along with a C_2H_2-type zinc finger structure. Groups II and III contained a single WRKY domain at the N-terminal region, but their zinc finger structures differed: Group II had a C_2H_2 zinc finger domain, whereas the zinc finger structure of Group III was distinct from the other groups, adopting a C_2HC configuration.

The gene lengths of RpWRKYs ranged from 471 bp (CHAGGene02964) to 2286 bp (CHAGGene24860). The exon-intron structures of RpWRKY genes and

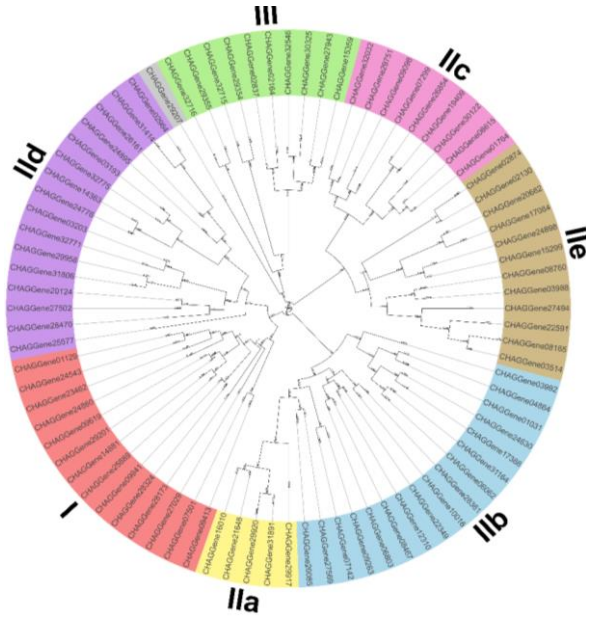

Fig. I. Phylogenetic tree of RpWRKY proteins. each color represents a distinct WRKY protein subgroup (I-III).

the conserved protein motifs are illustrated in Fig. (II). The number of exons varied from 2 to 7, with nearly half of the RpWRKY genes (39 genes) containing 3 exons.

A total of 20 conserved motifs were identified in the 84 RpWRKY protein sequences. The conserved WRKY domain was identified as motif 1 while motifs 2, 3, and 7 partially represented zinc finger structures. Zinc finger motifs 3 and 7 were found in both Group I and subgroup IId, whereas all other groups contained zinc finger motif 2. Other motifs were distributed around the WRKY domain in a unique manner, such as motifs 14 and 15, which were exclusive to Group III.

GO analysis of more than 30,000 *R. pseudoacacia* protein sequences identified 193 genes associated with water balance and drought tolerance, as well as 228 genes related to salt stress response, with 25 genes overlapping between these two groups. The promoter regions of the identified genes were analyzed for the presence of cis-regulatory elements recognized by WRKY transcription factor family proteins (W-box). Five characteristic motif variants from the JASPAR database were used for this analysis: MA1080.1 (AGTCAACG), MA0589.2 (TTGACCGAGC), MA1077.2 (GGTCAA), MA1295.2 (CGTTGACT), and MA1317.2 (TTGACTTTT).

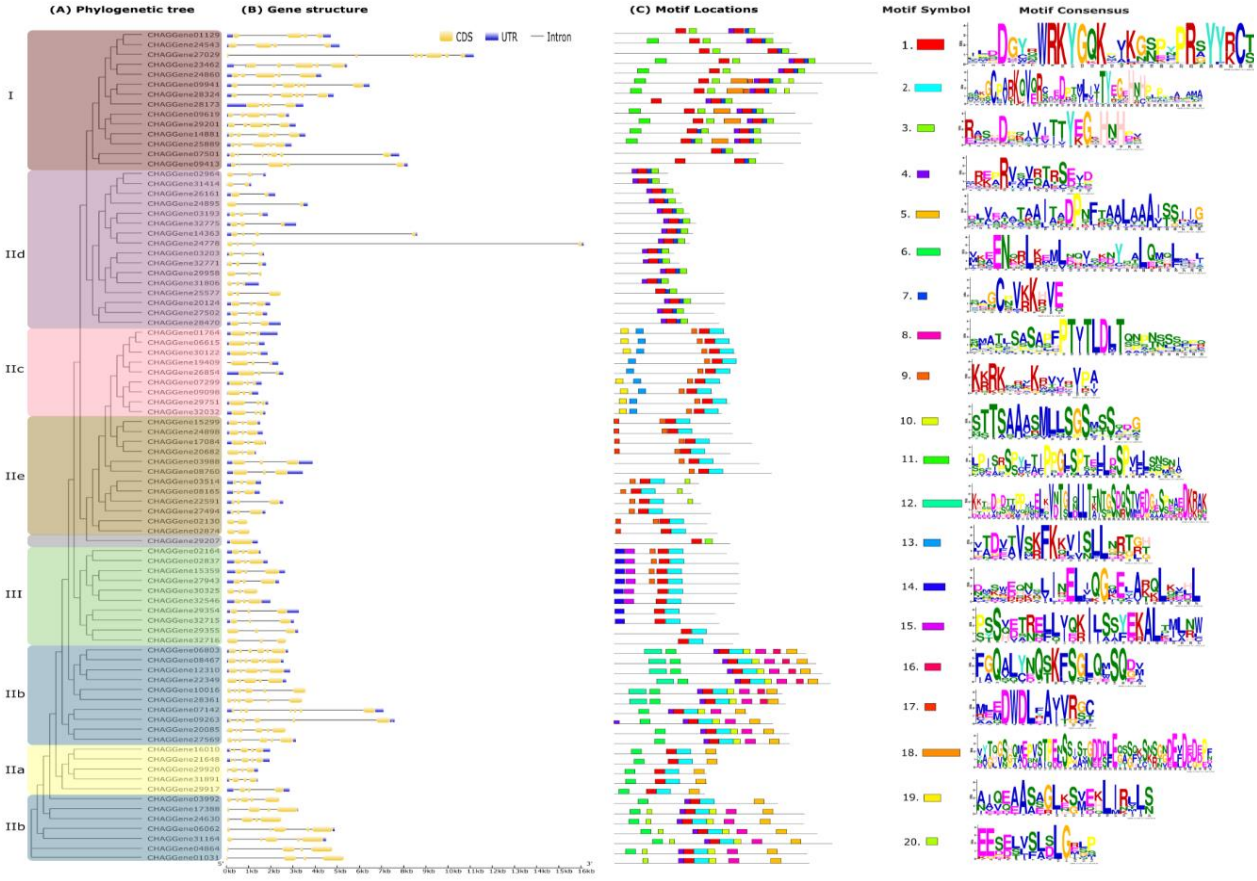

Fig. II. Evolutionary relationships (A), exon-intron structure (B), and conserved motifs (C) of the WRKY transcription factor family in *R. pseudoacacia*.

Analysis revealed the presence of 129 W-box motifs ($p<0.0001$) in the promoter regions of 68 genes associated with drought stress response. Among these, 26 genes contained more than one W-box motif within their promoters. Similarly, an analysis of the promoter regions of 93 genes involved in salt stress response identified 159 W-box occurrences, with 33 promoters containing multiple motifs.

For the 161 genes whose promoter regions contained W-box motifs, a repeated search for GO terms was performed, which resulted in 322 GO terms. Figure (III) illustrates the semantic distribution of these terms based on cluster analysis. Each cluster represents a set of GO terms that are biologically related.

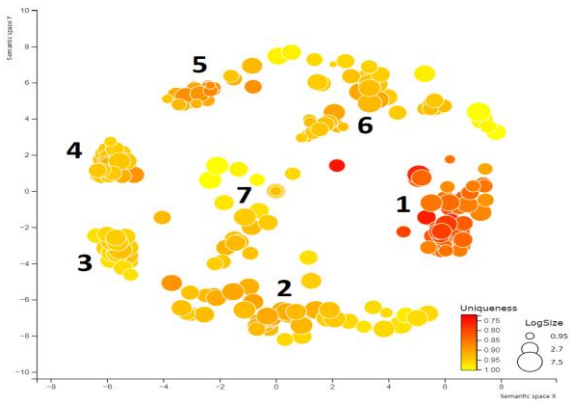

Fig. III. Semantic distribution of GO terms for 161 *R. pseudoacacia* genes with W-box motifs in their promoter regions; terms describing specific or narrowly defined functions/processes are shown in darker colors while broader or more general terms are represented in lighter colors. The first cluster includes terms related to responses to biotic and abiotic stresses; the second cluster relates to biosynthetic, metabolic, and genetic processes; the third involves molecular and ion transport; the fourth covers intracellular structure organization and assembly; the fifth is associated with organogenesis and cell differentiation; the sixth relates to gene expression regulation and cellular organization; and the seventh encompasses catabolism and cellular activity regulation.

Discussion

The physicochemical characterization of RpWRKY proteins revealed that their average grand average of hydropathy (GRAVY) values were negative, indicating predominantly hydrophilic

properties. Furthermore, intracellular localization analysis confirmed that all RpWRKY proteins are localized in the nucleus, emphasizing their role in transcriptional regulation.

Phylogenetic and structural analyses of RpWRKY proteins revealed that groups III, IIe, and IIc contained fewer introns than other groups, suggesting that these genes have undergone fewer duplication events and are relatively young in evolutionary terms. In contrast, Group I had the highest number of introns, which may indicate an earlier evolutionary origin. These phylogenetic findings are consistent with previous studies on WRKY family evolution, particularly supporting the hypothesis of the independent evolution of groups IIa and IIb, as well as the proposition that other WRKY groups originated from the C-terminal domain of Group I (Dai et al., 2023; Rinerson et al., 2015; Song et al., 2018). The presence of specific conserved motifs in different WRKY groups and subgroups allows for structural differentiation, yet the functional mechanisms underlying their divergence remain unclear. This highlights the need for further research to elucidate the molecular and functional characteristics defining the role of WRKY proteins in transcriptional regulation and plant stress adaptation.

GO analysis suggests that WRKY proteins function within intricate regulatory networks, simultaneously influencing multiple biological processes, including responses to biotic and abiotic stresses, as well as internal physiological processes. This observation aligns with findings by Phukan et al., (2016). A notable example of such multifaceted regulation is MdWRKY56, which enhances drought resistance in *Malus domestica* (Suckow) Borkh. by modulating malondialdehyde levels, reactive oxygen species accumulation, proline content, and antioxidant enzyme activity (Duan et al., 2023). MfWRKY40 promotes primary root elongation, enhances water uptake, reduces water loss under drought conditions, and significantly improves antioxidant defense in overexpression lines (Huang et al., 2022). Similarly, AtWRKY46 regulates developmental processes and hormonal responses, facilitating lateral root growth under osmotic and salt stress conditions via ABA signaling and auxin homeostasis maintenance (Ding et al., 2015).

WRKY transcription factors are key regulators that coordinate a wide range of cellular processes, including metabolism, gene expression, and responses to environmental stress. The multifunctionality of the WRKY family makes these proteins crucial for maintaining cellular homeostasis and adapting to changing environmental conditions. Further investigation of their functions could enhance our understanding of plant regulatory mechanisms in response to biotic and abiotic stresses.

The presence of W-box elements in gene promoter regions suggests their potential role in the regulation of target gene expression. While a single W-box is sufficient for WRKY-dependent regulation, multiple W-box motifs are frequently found in the promoter regions of downstream genes (Gu et al., 2019). Most genes associated with abiotic stress responses contain more than one W-box sequence, suggesting that adjacent W-box motifs within a target gene promoter may be recognized by interacting WRKY proteins, which can regulate gene expression either cooperatively or antagonistically (Grzechowiak et al., 2022). For instance, in *Arabidopsis thaliana* L., WRKY46, WRKY54, and WRKY70 proteins jointly bind to the BES1 gene promoter, participating in drought stress regulation (Chen et al., 2017b).

Conclusion

In this study, bioinformatics analysis of the *R. pseudoacacia* genome identified a total of 84 WRKY transcription factors (RpWRKY). Additionally, 161 genes associated with drought tolerance and salt stress response were predicted to be regulated by WRKY family proteins through binding to cis-regulatory W-box elements. The findings highlight the remarkable functional plasticity of the WRKY transcription factor family. Their ability to regulate multiple processes simultaneously, thereby facilitating adaptation to stressful conditions, presents significant opportunities for improving the tolerance of tree species to environmental challenges.

This study contributes to a deeper understanding of RpWRKY protein functions and provides a foundation for further research aimed at molecular breeding of *R. pseudoacacia* to develop more resilient, productive, and protective forest plantations.

Acknowledgment

This research was carried out as part of the state assignment of the Ministry of Science and Higher Education of the Russian Federation No. FNFE-2025-0013 'Regulation of gene expression involved in adaptation to adverse environmental factors to improve economically valuable traits of woody and cultivated plants used in agroforestry'.

Reference

Bailey, T. L., J. Johnson, C. E. Grant and W. S. Noble. 2015. The MEME suite. *Nucleic acids research,* 43, (W1) W39-W49.

Banerjee, A. and A. Roychoudhury. 2015. WRKY proteins: signaling and regulation of expression during abiotic stress responses. *The Scientific World Journal,* 2015, (1) 807560.

Chen, F., Y. Hu, A. Vannozzi, K. Wu, H. Cai, Y. Qin, A. Mullis, Z. Lin and L. Zhang. 2017a. The WRKY transcription factor family in model plants and crops. *Critical Reviews in Plant Sciences,* 36, (5-6) 311-335.

Chen, J., T. M. Nolan, H. Ye, M. Zhang, H. Tong, P. Xin, J. Chu, C. Chu, Z. Li and Y. Yin. 2017b. Arabidopsis WRKY46, WRKY54, and WRKY70 transcription factors are involved in brassinosteroid-regulated plant growth and drought responses. *The Plant Cell,* 29, (6) 1425-1439.

Chen, X., C. Li, H. Wang and Z. Guo. 2019. WRKY transcription factors: evolution, binding, and action. *Phytopathology Research,* 1, (1) 1-15.

Dai, W.-S., T. Peng, M. Wang and J.-H. Liu. 2023. Genome-wide identification and comparative expression profiling of the WRKY transcription factor family in two *Citrus* species with different Candidatus Liberibacter asiaticus susceptibility. *BMC Plant Biology,* 23, (1) 159.

Ding, Z. J., J. Y. Yan, C. X. Li, G. X. Li, Y. R. Wu and S. J. Zheng. 2015. Transcription factor WRKY 46 modulates the development of *Arabidopsis* lateral roots in osmotic/salt stress conditions via regulation of ABA signaling and auxin homeostasis. *The Plant Journal,* 84, (1) 56-69.

Duan, D., R. Yi, Y. Ma, Q. Dong, K. Mao and F. Ma. 2023. Apple WRKY transcription factor MdWRKY56 positively modulates drought stress tolerance. *Environmental and Experimental Botany,* 212, 105400.

Goyal, P., R. Devi, B. Verma, S. Hussain, P. Arora, R. Tabassum and S. Gupta. 2023. WRKY transcription factors: Evolution, regulation, and functional diversity in plants. *Protoplasma,* 260, (2) 331-348.

Grant, C. E., T. L. Bailey and W. S. Noble. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics,* 27, (7) 1017-1018.

Grzechowiak, M., A. Ruszkowska, J. Sliwiak, A. Urbanowicz, M. Jaskolski and M. Ruszkowski. 2022. New aspects of DNA recognition by group II WRKY transcription factor revealed by structural and functional study of AtWRKY18 DNA binding domain. *International journal of biological macromolecules,* 213, 589-601.

Gu, L., L. Dou, Y. Guo, H. Wang, L. Li, C. Wang, L. Ma, H. Wei and S. Yu. 2019. The WRKY transcription factor GhWRKY27 coordinates the senescence regulatory pathway in upland cotton (*Gossypium hirsutum* L.). *BMC plant biology,* 19, 1-14.

Hu, B., J. Jin, A.-Y. Guo, H. Zhang, J. Luo and G. Gao. 2015. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics,* 31, (8) 1296-1297.

Huang, Z., J. Wang, Y. Li, L. Song, D. E. Chen, L. Liu and C.-Z. Jiang. 2022. A WRKY protein, MfWRKY40, of resurrection plant *Myrothamnus flabellifolia* plays a positive role in regulating tolerance to drought and salinity stresses of Arabidopsis. *International Journal of Molecular Sciences,* 23, (15) 8145.

Jones, P., D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. Mcanulla, H. Mcwilliam, J. Maslen, A. Mitchell and G. Nuka. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics,* 30, (9) 1236-1240.

Kuzmin, P. and P. Krylov.2024. GENETIC DIVERSITY *ROBINIA PSEUDOACACIA* L.

GROWING IN DROUGHT CONDITIONS USING ISSR. *AGRICULTURAL SCIENCE DIGEST,* 1-6.

Li, Q., M. Zhao, N. Wang, S. Liu, J. Wang, W. Zhang, N. Yang, P. Fan, R. Wang and H. Wang. 2020. Water use strategies and drought intensity define the relative contributions of hydraulic failure and carbohydrate depletion during seedling mortality. *Plant Physiology and Biochemistry,* 153, 106-118.

Minh, B. Q., H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. Von Haeseler and R. Lanfear. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular biology and evolution,* 37, (5) 1530-1534.

Phukan, U. J., G. S. Jeena and R. K. Shukla. 2016. WRKY transcription factors: molecular regulation and stress responses in plants. *Frontiers in plant science,* 7, 760.

Rauluseviciute, I., R. Riudavets-Puig, R. Blanc-Mathieu, J. A. Castro-Mondragon, K. Ferenc, V. Kumar, R. B. Lemma, J. Lucas, J. Chèneby and D. Baranasic. 2024. JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic acids research,* 52, (D1) D174-D182.

Rinerson, C. I., R. C. Rabara, P. Tripathi, Q. J. Shen and P. J. Rushton. 2015. The evolution of WRKY transcription factors. *BMC plant biology,* 15, 1-18.

Schwärzel, K., L. Zhang, L. Montanarella, Y. Wang and G. Sun. 2020. How afforestation affects the water cycle in drylands: A process-based comparative analysis. *Global change biology,* 26, (2) 944-959.

Song, H., W. Sun, G. Yang and J. Sun. 2018. WRKY transcription factors in legumes. *BMC plant biology,* 18, 1-13.

Supek, F., M. Bošnjak, N. Škunca and T. Šmuc. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS one,* 6, (7) e21800.

Törönen, P. and L. Holm. 2022. PANNZER—a practical tool for protein function prediction. *Protein Science,* 31, (1) 118-128.

Wang, H., X. Cheng, D. Yin, D. Chen, C. Luo, H. Liu and C. Huang. 2023a. Advances in the research on plant WRKY transcription factors responsive to external stresses. *Current issues in molecular biology,* 45, (4).

Wang, Z., X. Zhang, W. Lei, H. Zhu, S. Wu, B. Liu and D. Ru. 2023b. Chromosome-level genome assembly and population genomics of *Robinia pseudoacacia* reveal the genetic basis for its wide cultivation. *Communications Biology,* 6, (1) 797.

Wani, S. H., S. Anand, B. Singh, A. Bohra and R. Joshi. 2021. WRKY transcription factors and plant defense responses: latest discoveries and future prospects. *Plant Cell Reports,* 40, 1071-1085.

Wu, W., J. Yang, N. Yu, R. Li, Z. Yuan, J. Shi and J. Chen. 2024. Evolution of the WRKY Family in Angiosperms and Functional Diversity under Environmental Stress. *International Journal of Molecular Sciences,* 25, (6) 3551.