MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
KYIV NATIONAL UNIVERSITY OF TECHNOLOGIES AND DESIGN
Faculty of Chemical and Biopharmaceutical Technologies
Department of Biotechnology, Leather and Fur

QUALIFICATION THESIS

on the topic **Regulation of Tissue-Specific Expression and Screening of Candidate Genes in the HSF Gene Family of *Populus wilsonii***
First (Bachelor's) level of higher education
Specialty 162 "Biotechnology and Bioengineering"
Educational and professional program "Biotechnology"

Completed: student of group
BEBT-21
Yixin Yin

Scientific
supervisor Iryna VOLOSHYNA,
Ph.D., As. prof.

Reviewer Olga IUNGIN
Ph.D., As. prof.

Kyiv 2025

KYIV NATIONAL UNIVERSITY OF TECHNOLOGIES AND DESIGN

Faculty: Chemical and Biopharmaceutical Technologies
Department: Biotechnology, Leather and Fur
First (Bachelor's) level of higher education
Specialty: 162 Biotechnology and Bioengineering
Educational and professional program Biotechnology

APPROVE
Head of Biotechnology, Leather and Fur Department, Professor,
Dr. Sc., Prof.
_____Olena MOKROUSOVA
«___»_____2025

ASSIGNMENTS
FOR THE QUALIFICATION THESIS
Yin Yixin

1. Thesis topic **Regulation of Tissue-Specific Expression and Screening of Candidate Genes in the HSF Gene Family of *Populus wilsonii***
Scientific supervisor Ph.D., Assoc. Prof. Iryna Voloshyna

approved by the order of KNUTD "05" March 2025, № 50-уч

2. Initial data for work: assignments for qualification thesis, scientific literature on the topic of qualification thesis, materials of Pre-graduation practice

3. Content of the thesis (list of questions to be developed): literature review; object, purpose, and methods of the study; experimental part; conclusions

4. Date of issuance of the assignments 05.03.2025

WORK CALENDAR

№	The name of the stages of the qualification thesis	Terms of performance of stage	Note on performance
1	Introduction	until 11 April 2025	
2	Chapter 1. Literature review	until 20 April 2025	
3	Chapter 2. Object, purpose, and methods of the study	until 30 April 2025	
4	Chapter 3. Experimental part	until 11 May 2025	
5	Conclusions	until 15 May 2025	
6	Draw up a bachelor's thesis (final version)	until 25 May 2025	
7	Submission of qualification work to the supervisor for feedback	until 27 May 2025	
8	Submission of bachelor's thesis to the department for review (14 days before the defense)	28 May 2025	
9	Checking the bachelor's thesis for signs of plagiarism (10 days before the defense)	01 June 2025	Similarity coefficient ____% Citation rate ____%
10	Submission of bachelor's thesis for approval by the head of the department (from 7 days before the defense)	04 June 2025	

I am familiar with the task:

Student _____ Yin Yixin

Scientific supervisor _____ Iryna VOLOSHYNA

SUMMARY

Yin Yixin. Regulation of Tissue-Specific Expression and Screening of Candidate Genes in the HSF Gene Family of *Populus wilsonii* – Manuscript.

Qualification thesis on the specialty 162 «Biotechnology and Bioengineering». – Kyiv National University of Technologies and Design, Kyiv, 2025.

Heat shock transcription factor (HSF) is a pivotal regulator of plant response to adversity stress; however, its functional diversity and evolutionary mechanism in woody plants require further analysis. In this study, the *HSF* gene family was systematically identified for the first time in the genome of *Populus wilsonii*. A total of 30 members (PwiHSF) were identified, which were similar in number compared with that of *Populus trichocarpa* (31) and significantly higher than that of Iljinskaea planisiliqua (23). These findings reveal the species-specific expansion of the HSF family in Populus. Phylogenetic analysis of the 30 PwiHSFs revealed their classification into three distinct subfamilies: A (12), B (2), and C (16). Subfamily C exhibited the highest prevalence, accounting for up to 53.3% of the total. Subsequent analysis of gene structure and functional domains revealed subfamily differentiation: Subfamily A contains three to four introns and complete functional domains (DBD, OD, CTAD). The acidic amino acid cluster of motif 7 in CTAD may enhance transcriptional activation by recruiting RNA polymerase II. Subfamily B contains one to two introns and lacks CTAD, but retains repressor motif 5. Subfamily C has no introns and retains only DBD, OD, and CTAD. Subfamily B contains one to two introns and lacks the CTAD element but retains the repressor motif 5. In contrast, subfamily C lacks introns and retains only the DBD and OD elements, suggesting that it is dependent on synergistic regulation by interactions. Transcriptome analysis revealed that members of subfamily A (e.g., PwiHSF7/12) exhibited high levels of expression in various tissues, while certain genes (e.g., PwiHSF3/5/6) demonstrated low expression levels in phloem and root tissues, indicative of functional tissue specificity. The protein interaction network further revealed PwiHSF12 as the core hub node and PwiHSF14/30 as the

secondary hub, suggesting that it enhances the flexibility of the network through synergistic regulation. This study is pioneering in its analysis of the significance of the HSF family in the context of tissue development.

Key words: Populus wilsonii, HSF gene family, Genome identification, Gene structure, Phylogenetic analysis, protein interactions network

TABLE OF CONTENTS

INTRODUCTION

Heat shock factor (HSF) is a core regulator of plant response to high temperature, drought, and other adversities. However, there is still a gap in the functional study of woody plants, especially in highly stress-resistant tree species, such as Populus wilsonii . In this study, the HSF gene family of Chinese Populus was systematically identified for the first time. Phylogenetic analysis classified a total of 30 members (PwiHSF) into three subfamilies: A (12), B (2), and C (16). The C subfamily accounted for 53.3%, which represents a significant breakthrough in the known distribution pattern of C genes in woody plants. This provides a new perspective for the study of plant gene family evolution. The integration of gene structure, evolutionary relationships, and expression pattern analysis revealed the functional differentiation of the HSF family in *Populus nigra* L. var. italica: the broad-spectrum members exhibited high levels of expression in buds, leaves, and inflorescences, suggesting their involvement in the regulation of basal metabolism; while the tissue-specific members demonstrated low levels of expression in the phloem, roots, or xylem, indicating their adaptation to microenvironmental changes through spatial and temporal specificity mechanisms. The study further constructed the protein interactions network of HSF and analyzed the potential regulatory pathways of HSF in the development of plant tissues and response to adversity. These results contribute to the existing body of evidence for the adaptive evolutionary mechanism of tissue development in Populus species. Furthermore, they provide target genes and a theoretical basis for the genetic improvement of forest trees to resist adversity. In the future, the primary focus will be on the dynamic regulatory network of HSF under drought and saline stress. Furthermore, the promotion of forest trees to resist adversity, breeding, and ecological restoration will be a central tenet of the research.

The relevance of the topic is to identify and analyze the HSF genome of the *Populus wilsonii*.

The purpose of the study is to identify core genes related to the promotion of plant tissue development by characterizing and analyzing the HSF genome of PwiHSF.

The objectives of the study is to identify core genes related to the promotion of plant tissue development by characterizing and analyzing the HSF genome of PwiHSF.

The object of the study PwiHSF

The subject of the study PwiHSF

Research methods comparative analysis

The scientific novelty comparative analysis

The practical will help us to understanf the HSF genome of PwiHSF by studying this process in depth.

Chapter 1
LITERATURE REVIEW

1.1 Overview of *Populus wilsonii*

1.1.1 Introduction of *Populus wilsonii*

Populus wilsonii, a perennial deciduous tree, is classified within the genus Populus, which is part of the family Salicaceae. It is a significant ecological and economic tree species in the western mountains of China, as well as one of the endemic native tree species in China[0]. The species' natural distribution range is concentrated in the central and western regions of China, as well as the southwestern part of the country[2]. It is commonly found in the Yangtze River Basin, Sichuan Basin, Yunnan Province, and Guizhou Province, as well as other regions along the riverbanks, streams, and wet valleys[4]. The altitude of these regions ranges from 800 to 3,000 meters. The species exhibits remarkable morphological characteristics. Adult plants attain heights of up to 25 meters, with a diameter at breast height of up to 1.5 meters. The bark displays shallow longitudinal fissures and a dark grayish-brown hue. The crown assumes a broadly tower-shaped structure, characterized by a pronounced spreading aspect. The leaf blades exhibit a wide range of forms, with broadly ovate forms predominating. Ecological adaptability studies have demonstrated that the *Populus wilsonii* possesses a variety of stress-mitigating characteristics, including physiological benefits such as drought resistance, cold resistance, and salinity tolerance[5].The tree's deep-root characteristics – of which the main root can reach depths of over 3 meters – and fast-growing characteristics – of which the average annual high growth rate is between 1.2 and 1.8 meters – make it a preferred tree species in ecological restoration projects. Nonetheless, fundamental research on the genetic diversity of its germplasm resources and the molecular mechanism of stress resistance remains inadequate[6]. This deficiency impedes the scientific application of this species in ecologically fragile areas. This is a pivotal scientific issue that necessitates resolution .

1.1.2 Research Progress of *Populus wilsonii*

The *Populus wilsonii*, a high-altitude ecological keystone species in western China, exhibits a remarkable combination of ecological adaptability and economic value. Research has demonstrated that the subject exhibits a synergistic response to drought and salinity stress, manifesting through an increase in proline accumulation (2.3-fold), an escalation in SOD activity (40%), and an enhancement in SOS1-mediated Na^+ efflux[7]. Its distinguishing characteristics, including a profoundly developed root system (3 m primary root) and a notably rapid growth rate (1.5-2 m annual growth), contribute to its preeminence in river valley habitats. Genomic study revealed that its genome size is approximately 450 megabases (Mb), with more than 32,000 genes annotated, among which the heat-shock factor (HSF), NAC, and other stress-resistant gene families have expanded by 15% compared with those of Populus tremula, suggesting the potential for functional differentiation. Transcriptome analysis revealed that the expression of genes belonging to the abscisic acid (ABA) pathway (e.g., PYL and SnRK2) and heat-shock protein 70 (HSP70) increased three to fivefold under conditions of drought stress. However, systematic functional analysis of the heat-shock transcription factor (HSF) family remains to be elucidated[8], with the exception of PwiHSF12, which has been predicted to be involved in the response to heat stress [9]. In recent applications, the utilization of *Populus wilsonii* in ecological restoration projects within alpine regions has yielded notable outcomes. A notable benefit of this approach is the substantial reduction in soil erosion, which has been documented to decrease to less than 500 t/(km²-a) following the implementation of afforestation measures. However, its molecular breeding is constrained by several factors[10]. First, there is an absence of a genetic transformation system for key genes, such as HSF. Second, there is an absence of a multi-histological regulatory network. Third, there is an absence of a database of germplasm genetic diversity[11]. The potential of the Populus wilsoniiin ecological restoration is significant due to its fast-growing characteristics and ability to tolerate water and moisture[12]. These characteristics make it a preferred species for riparian protection forest

construction. The wood is characterized by its lightness and softness, as well as its uniform texture, which renders it suitable for papermaking and the manufacturing of man-made boards[13]. Research is underway to extract and utilize secondary metabolites, such as tannins, found in the bark. This research provides a new direction for further study.

1.2 HSF Overview and Current Research Status

1.2.1 Overview of HSF

The HSF (Heat Shock Factor) gene family has been identified as a core regulator of organisms in response to environmental stresses. Members of the family have been shown to initiate the expression of Heat Shock Protein (HSP) through the specific recognition of Heat Shock Element (HSE, conserved sequence nGAAn). This process, in turn, has been demonstrated to mediate a variety of abiotic stress response processes[14]. In the plant stress response mechanism, HSF plays a fundamental role in the maintenance of protein homeostasis mediated by HSP (e.g., repair of misfolded proteins)[15]. Additionally, HSF forms a multidimensional stress response network by dynamically regulating ion homeostasis, redox homeostasis, and metabolic reprogramming processes[16].

1.2.2 Current status of plant HSF research

The present state of research on plant heat stress transcription factors (HSFs) is characterized by a concentration on model plants, including Arabidopsis thaliana and Oryza sativa[17]. The regulatory networks and stress resistance mechanisms of these plants have become more clearly defined. For instance, 22 HSF members (HSFA1-HSFA9, HSFB1-HSFB5, etc.) have been identified in Arabidopsis thaliana, and the central role of HSFA1s in heat stress signaling has been systematically resolved[18]. Conversely, research on plant HSFs has been comparatively deficient; however, recent years have witnessed advancement in this domain, particularly with regard to Populus species. To illustrate, genome-wide analysis of P. trichocarpa has led to the identification of 31 PtHSF genes, among

which PtHSFA4a has been demonstrated to substantially augment salt tolerance in transgenic plants by modulating HSP70 expression [19]. However, the majority of extant studies have focused on gene identification and preliminary functional validation, and there is still a paucity of systematic research on the functional differentiation of the HSF family in different species of Populus, its multidimensional regulatory network, and its association with ecological adaptations, especially in wild species with special resistance (e.g., *Populus wilsonii*), which has not yet been reported[20].

1.2.3 Functional overview of the HSF gene family

Heat shock protein (HSP) gene transcription is initiated by HSF through its specific binding to the heat stress response element HSE (nGAAn). In the presence of heat stress, the HSF monomer undergoes phosphorylation and modification, resulting in the formation of an active trimer. The DNA-binding domain (DBD) of the HSF trimer binds to the promoters of target genes through a helix-turn-helix structure, inducing the expression of molecular chaperones, such as HSP70 and HSP90[21]. These chaperones, in turn, facilitate the repair of misfolded proteins and maintain cellular homeostasis, thereby significantly enhancing plant stress tolerance. For instance, AtHSFA1a, a key component of the AtHSFA1a-AtHSFA1b complex in the model organism Arabidopsis thaliana, has been shown to reduce the survival rate of the mutant under 38°C stress by 60% through the activation of HSP101 expression [22]. In addition to heat stress response, the HSF family integrates signals associated with drought, salt stress, and complex adversity. Through a process known as cross-talk, these signals are integrated to form a multidimensional regulatory network. The classification of plant HSF families is based on structural domain differences, with the classification system comprising the following categories: HSFAs (e.g., wheat TaHSFA6f) contain AHA-activating motifs and have been demonstrated to dominate the heat stress response[23]. Furthermore, their expression has been shown to increase Arabidopsis thaliana's 45°C survival rate by threefold. In contrast, HSFBs (e.g., AtHSFB1)

limits HSP overexpressions through LF repressor motifs to balance energy metabolism; and HSFCs (e.g., maize ZmHSFC1), on the other hand, mediate the cross-regulation of development and stress, and its silencing strain reduced root biomass by 35% under salt stress [24].

1.3 Research objectives and significance

By elucidating the molecular mechanisms of HSF regulation of heat shock proteins (HSPs) and antioxidant systems, we aim to reveal the core pathways of plants against abiotic stresses such as high temperature, drought, salinity, and alkalinity. In addition, we aim to construct a theoretical framework for plant stress response[25]. Concurrently, we will explore the functions of HSF in seed germination, flowering regulation, and fruit ripening. Ultimately, our research will optimize the resilience of the entire agricultural production chain and the efficiency of resource utilization. Furthermore, the development of HSF-driven environment-responsive promoters or biosensors will facilitate the creation of precision agriculture tools, such as high-temperature early warning systems in the field[26]. Additionally, the transplantation of HSF functions across species has the potential to enhance plant stress tolerance and promote the restoration of degraded ecosystems. This can be achieved by importing extreme environment plant genes into crops.

The study of plant heat shock factor (HSF) is of great significance for both theory and application. As a core regulator of plant response to high temperature, drought, and other stresses, HSF maintains cellular homeostasis through the activation of heat shock proteins and antioxidant systems[27]. In addition, HSF collaborates with signaling pathways, such as abscisic acid (ABA) and reactive oxygen species (ROS), to form an antiretroviral network. Its role in the development of seeds, flowering regulation, and other aspects of the process further expands the potential for agricultural applications[28]. While the functions of HSF in model plants, such as Arabidopsis thaliana and rice, have been elucidated, studies on woody plants, particularly ecologically critical species such as *Populus wilsonii*, remain underdeveloped. The spatial and temporal expression patterns,

protein interactions, and epigenetic mechanisms of the HSF family of *Populus wilsonii* under drought, salinity, and other stresses have yet to be analyzed[29]. In the future, there is a need to combine multi-omics technology and genetic transformation systems to elucidate the molecular pathways of HSF in the development of *Populus wilsonii*. This will provide theoretical support for the development of forest tissues and ecological restoration[30]. The results of this study will provide a theoretical basis for further elucidation of the importance of HSF transcription factors in the tissue development of *Populus wilsonii*. Concurrently, the results will lay the foundation for the genetic improvement of forest tree resistance[31]. A thorough analysis of the copy number, structural variation, and phylogenetic relationship of HSF family members has the potential to yield novel insights into the evolution of plant gene families. This analysis could provide theoretical support for the adaptive evolution mechanism of Populus (Populus) species.

Summary of the chapter 1

1.Biological Characteristics and Ecological Value of *Populus wilsonii*

Populus wilsonii is a unique, perennial, deciduous tree species in the genus Populus. It is found in the western mountains of China, specifically in the Yangtze River Basin and the southwestern region at elevations between 800 and 3,000 meters. It has a deep root system (with a main root up to three meters long), grows quickly (with an average annual growth rate of 1.2 to 1.8 meters), and is drought- and cold-resistant. It is widely used in riverbank protection forests and ecological restoration projects. It is also used for riverbank protection and ecological restoration.

2.Research Progress and Genomic Characteristics of Populus tremula

Genomic research reveals that its genome size is approximately 450 megabase pairs (Mb), containing over 32,000 annotated genes. The family of resistance genes, including HSF and NAC, has grown by 15% compared to Populus tremula.

Transcriptome analysis showed that ABA pathway genes (*PYL* and *SnRK2*) and *HSP70* were up-regulated 3-5 fold under drought stress. However, the functional analysis of the HSF family remains incomplete.

Current limitations to application include a lack of a genetic transformation system, an unknown multi-omics regulatory network, and an insufficient germplasm genetic diversity database.

3.The function and mechanism of action of the HSF gene family

HSFs (heat stress transcription factors) initiate HSP gene expression by recognizing the heat stress response element (HSE, nGAAn). They repair misfolded proteins and maintain cellular homeostasis.

Under heat stress, the HSF monomer undergoes phosphorylation to form an active trimer. Its DNA-binding domain (DBD) then activates molecular chaperones, such as HSP70 and HSP90, which work together to regulate ion homeostasis, redox, and metabolic reprogramming.

4.The current status and challenges of plant HSF research

The function of HSFs in model plants, such as Arabidopsis thaliana and rice, has been better defined. For example, AtHSFA1a regulates the response to heat stress. However, studies on woody plants are lagging behind.

Thirty-one PtHSF genes have been identified in Populus trichocarpa, one of which, PtHSFA4a, enhances salt tolerance. However, the functional differentiation and regulatory network of HSF in wild species, such as the *Populus wilsonii*, remain unclear.

5.The following section will delineate the objectives of the research study and their theoretical significance.

Molecular mechanism analysis: The objective of this study is twofold: firstly, to elucidate the spatial and temporal expression, protein interactions, and epigenetic mechanisms of the HSF family in drought/salt stress, and secondly, to fill the gaps in woody plant research.

Theoretical framework construction: The objective of this study is to elucidate the fundamental pathway through which heat-shock factor (HSF)

regulates heat-stress tolerance in plants by traversing the heat-shock protein (HSP) and antioxidant systems. This investigation aims to facilitate cross-species comparisons and promote research on adaptive evolution.

Gene Family Evolution: By analyzing HSF copy number, structural variation, and phylogeny, we will provide a new case study on the evolutionary mechanism of *Populus species*.

Chapter 2

OBJECT, PURPOSE, AND METHODS OF THE STUDY

2.1 Acquisition of *Populus wilsonii* HSF genomic data and family identification

The genomic data of the *Populus wilsonii* were obtained from the Figshare database. The genome data for Arabidopsis thaliana was obtained from TAIR, and the genome data for *Populus trichocarpa* was obtained from Phytozome V13. Finally, the HSF protein sequences of Arabidopsis thaliana and *Populus trichocarpa* were downloaded as a reference for the evolutionary analysis. These HSF members were then utilized to retrieve the HSF genomes of the *Populus wilsonii* through the use of BLASTP[25.] The genome was then combined with a HMMER (PF00447 structural domain, E-value <1e-5) search, and the structural domain integrity was verified by CDD/SMART in order to obtain the *Populus wilsonii* HSF family members.

2.2 Analysis of physicochemical properties of *Populus wilsonii* HSF gene family

A systematic analysis of the physicochemical properties of HSF proteins was conducted using the ExPASy ProtParam online platform (https://web.expasy.org/protparam/) . This analysis encompassed molecular weight (MW), theoretical isoelectric point (pI), amino acid composition, and stability-related parameters (aliphatic index, hydrophilic mean GRAVY value)[26]. GRAVY values greater than 0 indicate hydrophobic proteins, while those less than 0 indicate hydrophilic proteins. This analysis was undertaken to elucidate the physicochemical properties of the proteins in relation to their potential functions.

2.3 Chromosomal localization analysis of *Populus wilsonii* HSF gene

In the chromosome localization analysis, BEDTools (v2.30.0) was first utilized to extract the chromosome location information of each HSF gene from the annotation file of *Populus wilsonii* [27]. This included the chromosome number (ChrID), the start site (Start), and the end site (End). Subsequently, a chromosome distribution map was generated by TBtools (v1.108) based on the extracted physical coordinate data to visualize the distribution characteristics of the HSF gene family in the genome of the Populus

2.4 Phylogenetic analysis of the HSF gene family of *Populus wilsonii*

The G-INS-i algorithm of MAFFT v7.490 was employed to perform a multiple sequence comparison using the protein sequences of HSF family members. Subsequently, TrimAl 1.4.1 was utilized to remove low-conserved sites (retention rate >75%) in order to retain the HSF core structural domains (DBD, OD) for further analysis. The construction of a phylogenetic tree was performed using the maximum likelihood method of MEGA 11 (Bootstrap=1000), with the optimal amino acid substitution model (LG+G+I) employed to support the analysis. The resulting tree exhibited subfamilies A, B, and C, which were classified according to the taxonomic system of A. thaliana. The iTOL v6 platform was utilized to label the tree with branching support and species origin, facilitating the visualization of the evolutionary relationships among the analyzed specimens[28].

2.5 Structure, conserved domains and motif analysis of *Populus wilsonii* HSF gene

The gene structure of the *Populus wilsonii* HSF can be analyzed and presented using TBtools software, and the conserved structural domains in the *Populus wilsonii* protein sequence can be predicted and analyzed using MEMEonline software (http://meme-suite.org/tools/meme). The identification of functional motifs was subsequently conducted using the MEME Suite (motif

length 6-50 aa, maximum motif number 10), and the visualization of the distribution was facilitated by TBtools[29].

2.6 Expression pattern analysis and protein interaction analysis

The TPM expression profile of HSF genes was obtained by RNA-Seq sequencing through Illumina NovaSeq 6000 platform using *Populus wilsonii* transcriptome data, combined with HISAT2 comparison and featureCounts quantitative analysis. A screening of inter-tissue differentially expressed genes was conducted using DESeq2 (log2FoldChange >1, FDR <0.05)[23]. The expression patterns were subsequently demonstrated by heatmapping with ComplexHeatmap. The Tau index ($\tau \geq 0$.Eighth, the objective was to identify high tissue-specific genes and to demonstrate the tissue-specific (roots, stems, and leaves) expression profiles. TPM expression profiles were obtained from the STRING database (confidence > 0.7) to screen protein interactions, and Cytoscape was used to construct the network. The MCODE plug-in was employed to identify the interactions hubs (Hub genes), in conjunction with the expression and interactions data, to analyze the synergistic regulatory pathways of HSF in tissue development[30].

Summary of chapter 2

1.Genomic Data Acquisition and Family Identification

Data Sources: Populus genomic data were obtained from the Figshare database. Arabidopsis thaliana (TAIR) and Populus (Mauve, Phytozome V13) HSF protein sequences were used as references.

Identification Process: *Populus wilsonii* HSF members were searched using BLASTP and HMMER (PF00447 structural domain, E-value <1e-5).

Structural domain integrity was verified using CDD/SMART to ensure the accuracy of family membership.

2.Physicochemical property analysis:

Tools and Parameters: The following parameters were analyzed using the ExPASy ProtParam online platform:

Molecular weight (MW), theoretical isoelectric point (pI), and amino acid composition.

Aliphatic index; and hydrophilicity mean value (GRAVY value). GRAVY > 0 indicates hydrophobic proteins and GRAVY < 0 indicates hydrophilic proteins.

3.Chromosome localization analysis:

Data extraction: Gene positions (ChrID, start, end) were extracted from GFF3 files using BEDTools (v2.30.0).

Visualization: Generate chromosome distribution maps using TBtools (v1.108) and label gene density and the physical coordinates of candidate genes (e.g., PwiHSF-19).

4.Phylogenetic analysis:

Comparison and trimming: MAFFT v7.490 (G-INS-i algorithm) was used for multiple sequence comparison and TrimAl 1.4.1 was used for trimming low-conserved sites (retention >75%).

Tree construction and classification: MEGA 11 (LG+G+I model, bootstrap = 1000) was used to construct a maximum likelihood tree and classify into subfamilies A, B, and C, according to the Arabidopsis classification.

Visualization: iTOL v6 annotated branch support and species origin.

5.Gene Structure and Functional Domain Analysis:

Gene structure: TBtools analyzed the intron-exon distribution.

Conserved motifs: The MEME Suite predicts motifs (length 6-50 amino acids [aa], maximum 10), and TBtools visualizes motif distribution.

Structural domain validation: CDD/SMART ensures the integrity of core structural domains, such as DBD and OD.

6.Expression Pattern and Protein Interaction Analysis:

RNA-Seq Process:

Illumina NovaSeq 6000 sequencing, HISAT2 comparison and FeatureCounts quantification (TPM value)

DESeq2 is used for screening differential genes (|log2FC|>1, FDR<0.05). ComplexHeatmap is used for creating tissue-specific heatmaps.

The tau index ($\tau \geq 0.8$) was used to identify highly tissue-specific genes (e.g., PwiHSF-19, which is highly expressed in the forming layer).

Interaction network construction:

The STRING database (confidence >0.7) was used to predict interacting proteins.

The Cytoscape (v3.9.1) network was constructed using the MCODE plugin to identify hub genes (e.g., PwiHSF-12).

We integrated expression and interaction data to resolve co-regulatory pathways (e.g., the MAPK3 and PMA2 interaction module).

Chapter 3
EXPERIMENTAL PART

3.1 Analysis of physicochemical properties of HSF gene family of *Populus wilsonii*

In this study, 30 HSF genes of the species *Populus trichocarpa* were identified and designated PwiHSF1-PwiHSF30 according to their chromosomal location and order. The number of amino acids encoded by the HSF genes of *Populus wilsonii* ranged from 100 to 522, and the molecular weights ranged from 24,045.52 to 58,140.81 kilodaltons, as analyzed by ExPASy ProtParam. The isoelectric point (pI) exhibited a broad spectrum (4.7-9.35), while the instability index demonstrated a range from 38.47 to 66.97. The fat coefficients ranged from 58.88 to 81.36, and the total range of average hydrophilicity was -0.896 to -0.408 (Table 1).

Table 3-1 Physical and Chemical Properties of the HSF Gene Family in *Populus wilsonii*

Gene Name	Chro	Star site	End site	Size	MW*	Pl*	Instability Index	AI*	GRAVY*
PwiHSF1	Chr01	9207526	9210360	341	36606.48	5.08	54.97	70.41	-0.531
PwiHSF2	Chr01	12017736	12023291	510	55968.1	4.82	66.31	67.14	-0.663
PwiHSF3	Chr01	31039616	31041747	270	31320.41	6.63	54.81	68.93	-0.623
PwiHSF4	Chr01	35991911	35995422	490	54763.67	5.77	57.44	69.49	-0.8
PwiHSF5	Chr02	3365707	3368490	359	41095.3	5.3	57.18	67.41	-0.806
PwiHSF6	Chr02	9864874	9866879	364	40464.73	8.15	52.5	73.9	-0.47
PwiHSF7	Chr03	14687350	14693439	507	55657.89	4.8	59.76	70.61	-0.598
PwiHSF8	Chr04	3591825	3593891	209	24045.52	9.16	57	70.53	-0.625
PwiHSF9	Chr04	5692073	5693455	407	46509.91	5.19	56.82	63.24	-0.833
PwiHSF10	Chr05	23978880	23982067	359	40720.52	5.5	66.97	62.76	-0.896
PwiHSF11	Chr06	3606809	3608856	229	26695.43	8.87	54.6	75.76	-0.752

Table 1 Physical and Chemical Properties of the HSF Gene Family in *Populus wilsonii*

Gene Name	Chro	Star site	End site	Size	MW*	Pl*	Instability Index	AI*	GRAVY*
PwiHSF12	Chr06	9575079	9577462	522	58140.81	4.77	66.27	65.17	-0.595
PwiHSF13	Chr06	13395536	13397919	512	56721.7	5.23	55.49	72.19	-0.563
PwiHSF14	Chr06	24427045	24431260	388	43825.86	5	57	73.04	-0.588
PwiHSF15	Chr07	4290255	4293957	285	31056.94	4.81	38.47	58.88	-0.876
PwiHSF16	Chr08	9375410	9378332	393	44819.08	4.83	40.71	71.96	-0.702
PwiHSF17	Chr08	11017767	11020043	342	39154.01	5.2	58.86	74.65	-0.698
PwiHSF18	Chr09	8143896	8145629	304	34920.67	8.55	57.46	75.92	-0.495
PwiHSF19	Chr10	10229110	10232110	358	41340.35	5.16	54.2	68.02	-0.761
PwiHSF20	Chr10	11960708	11965120	392	44705.84	4.7	43.8	70.87	-0.711
PwiHSF21	Chr11	4659694	4661586	211	24339.62	9.35	47.49	59.67	-0.773
PwiHSF22	Chr11	7164036	7167313	406	46298.63	5.29	53.58	64.83	-0.817
PwiHSF23	Chr12	16636420	16637556	307	33777.61	4.99	50.87	67.23	-0.672
PwiHSF24	Chr13	7541897	7547871	499	55006.55	5.54	57.46	66.01	-0.572
PwiHSF25	Chr14	1705546	1707503	368	41045.13	8.16	53.24	68.1	-0.564
PwiHSF26	Chr14	9807753	9810817	444	50951.87	5.75	60.7	65.23	-0.813
PwiHSF27	Chr15	16848024	16849709	286	31697.66	5.06	48.07	81.36	-0.462
PwiHSF28	Chr16	4371962	4373990	228	26427.99	8.27	57.02	71.84	-0.72
PwiHSF29	Chr17	5634877	5639140	485	54409.32	5.68	59.64	67.2	-0.779
PwiHSF30	Chr18	10707225	10708664	339	38001.02	5.43	50.32	76.19	-0.408

3.2 Chromosomal localization analysis of HSF genes in *Populus wilsonii*

Chromosome analysis revealed that the distribution of the genes was not uniform across the chromosomes, with the exception of a single HSF gene, which was found on chromosomes 3, 5, 7, 9, 12, 13, 15, 16, 17, and 18. The remaining chromosomes exhibited an abundance of family members, with the highest occurrence observed on chromosomes 1 and 6, which contained four HSF genes. As illustrated in（Figure 1）, PwiHSF6 and PwiHSF6, as well as PwiHSF25, PwiHSF11, and PwiHSF28, are homologous to each other as gene pairs.

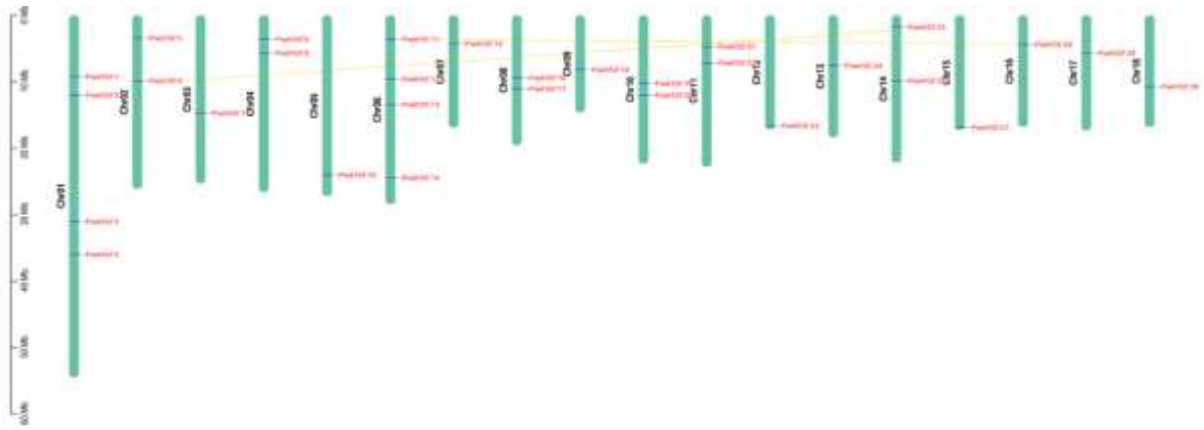

Figure 3.1 Chromosome Localization of PwiHSF Family in *Populus wilsonii*

3.3 Phylogenetic analysis of the HSF gene family in *Populus wilsonii*

In order to facilitate a more thorough examination of the evolutionary relationships within the HSF gene family, the MEGA 11 software was utilized to analyze the HSF gene family of the Populus wilsonii and construct a phylogenetic tree (Figure 2). This analysis resulted in the identification of three major branches, designated A, B, and C. The A subfamily comprised 12 genes, the B subfamily consisted of 2 genes, and the C subfamily contained 16 genes.

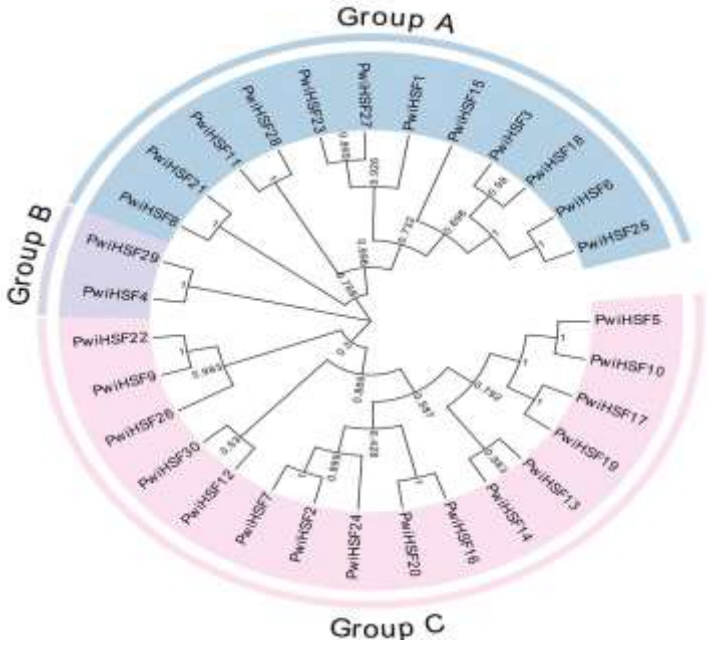

Figure 3.2 PwiHSF phylogenetic tree

3.4 Structural, conserved domains and motif analysis of *Populus wilsonii* HSF gene

A systematic analysis of the structural features and functional domain composition of PwiHSF genes revealed significant differentiation among different subfamily members in terms of gene structure and functional modules (Figure 3). Gene structure analysis revealed that subfamily A members (e.g., PwiHSF07 and PwiHSF12) exhibited a higher number of introns (3-4) and a long non-coding region at the 5' end, whereas subfamily B (e.g., PwiHSF03) and subfamily C (e.g., PwiHSF19) contained only 1-2 and no introns, respectively, suggesting the differentiation of their functional regulatory mechanisms. A thorough examination of the conserved domain identification revealed that all PwiHSF proteins contain the characteristic HSF functional domains. The N-terminal DNA-binding domain (DBD) contains three conserved α-helices and is responsible for recognizing the heat stress element (HSE). The oligomerization structural domain (OD) is responsible for mediating protein interactions, and the OD region of the A subfamily carries the nuclear localization signal (NLS). The C-terminal activation domain (CTAD), a feature exclusive to the A subfamily, exhibits a high concentration of acidic amino acids, a characteristic that may potentiate its capacity for enhancing transcriptional activation. A subsequent motif prediction analysis yielded 10 conserved motifs (Figure 4), the distribution of which exhibited a high degree of correlation with subfamily classification. Specifically, subfamily A contained all motifs (including DBD core motif 1, OD-regulated motif 4, and CTAD-associated motif 7), subfamily B was missing the CTAD motif but retained the repressor motif 5, and subfamily C retained only the basic motifs 1-3. These results suggest that the reduced number of introns and simplified functional domains (e.g., the absence of CTAD in subfamily C) may reflect the functional specialization of subfamilies during the evolutionary process. The complex modular structure of subfamily A (e.g., the glutamate cluster of motif 7) may provide a structural basis for its coregulatory roles through the recruitment of RNA polymerase II for the synergistic activation of stress-responsive genes.

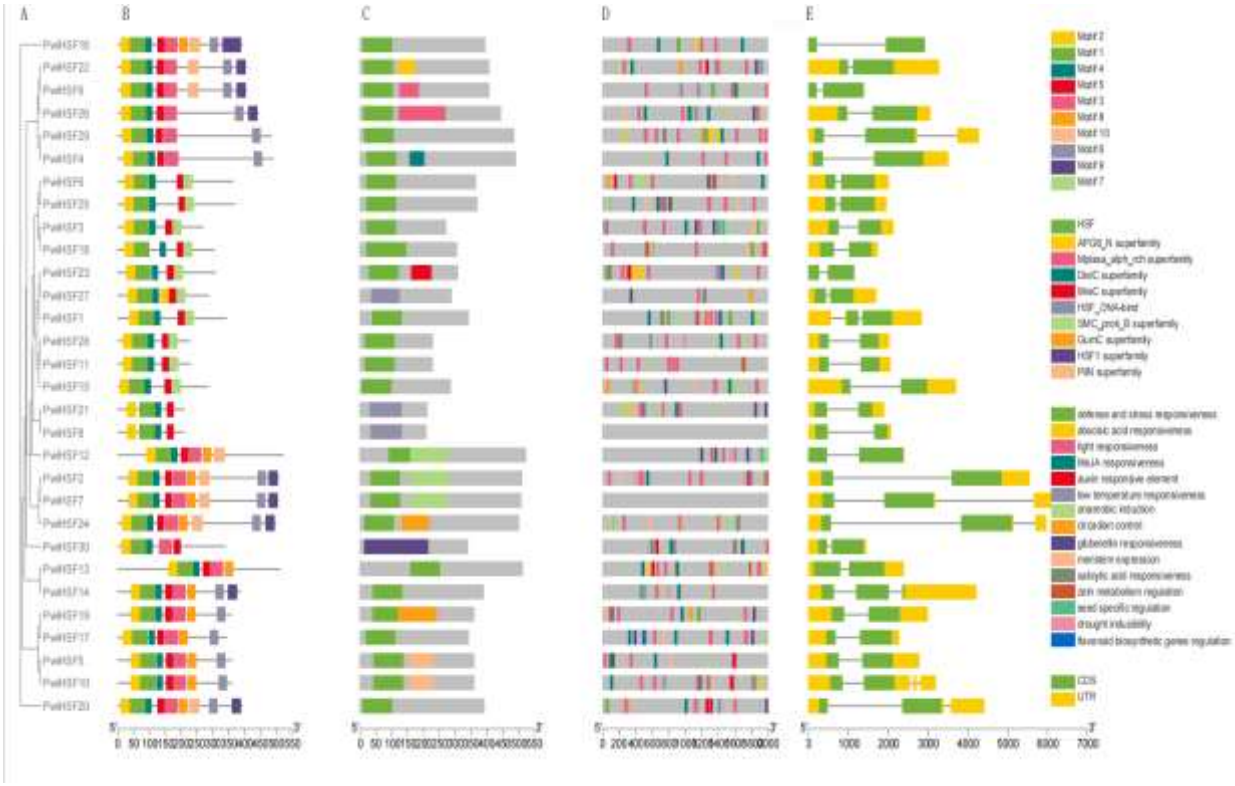

Figure 3. Gene structure analysis of *Populus wilsonii* HSF gene family

A:Phylogenetic tree of 30 HSF transcription factors of the populis wilsonii; B :Amino acid conserved structure city analysis of 30 HSF transcription factors of the populis wilsonii; C :Amino acid HSF city analysis of 30 HSF transcription factors of the populis wilsonii; D :Cis-acting element analysis of the 2000 bp promoter sequences of the 30 HSF transcription factors of the populis wilsonii; E :Sequence analysis of the genes of the 30 HSF transcription factors of the populis wilsonii.

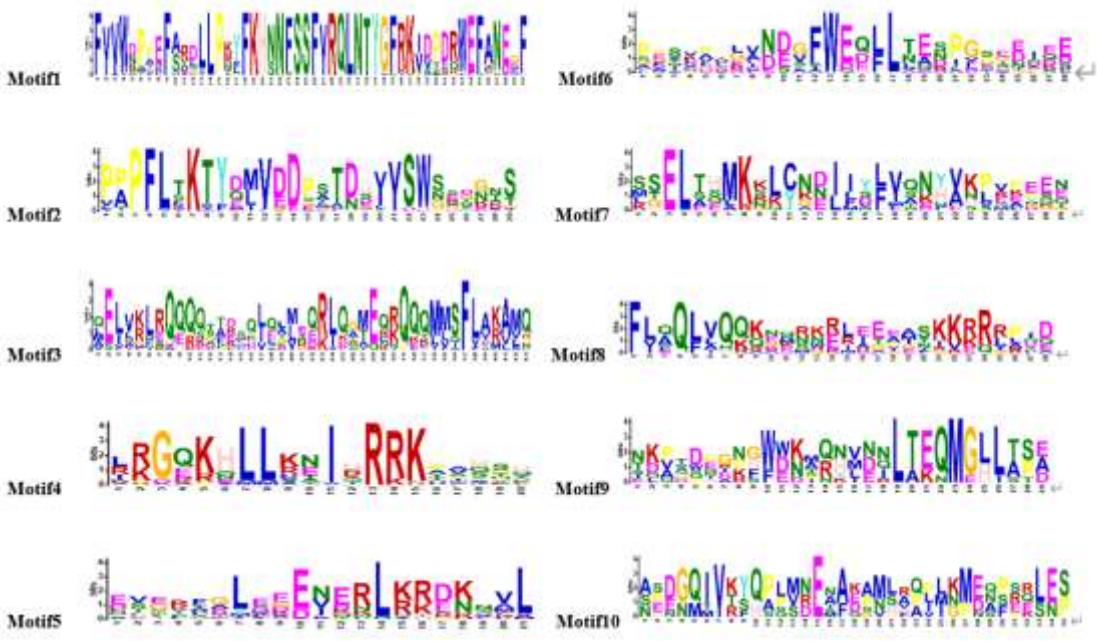

Figure 4. Motif logo of *Populus wilsonii* HSF gene family

3.5 Expression pattern analysis and protein interaction analysis

In order to better understand the function of HSF genes in the growth and development of *Populus wilsonii*, their expression levels in different tissues, such as buds, inflorescences, flower buds, leaves, phloem, roots, and xylem, were analyzed using pre-published transcriptome data (Figure 5). The expression levels of PwiHSF1/2/4/7/9/10/12/14/15/16/20/22/23/24/26/27/29 were found to be elevated in all examined tissues. In addition, the expression levels of PwiHSF1/2/4/7/9/10/11/12/14/15/16/20/22/23/24/26/27/29/30 were found to be elevated in the tissues of the bud, inflorescence, flower bud, and leaves. The expression of PwiHSF3/5/6/8/11/13/17/18/19/21/25/28/30 was observed in these tissues, and PwiHSF3/5/6/8/13/17/18/19/21/25/28 was expressed in tissues such as phloem, root, and xylem. This finding suggests that the HSF genes of the *Populus wilsonii* exhibit functional differentiation in diverse tissues and organs, with some genes functioning in multiple tissues and others expressed exclusively in specific tissues.

Protein interaction analysis revealed that the 23 HSF transcription factors interacted with each other or with other proteins (Figure 6). An interaction network

was formed, and PwiHSF12 was identified as the central interaction factor of this network. These findings suggest that PwiHSF12 plays a pivotal role in the tissue development of Chinese Populus . Among them, PwiHSF14 and PwiHSF30 exhibited the same node degree and accounted for the second largest proportion after PwiHSF12, which was also an important gene in this study.

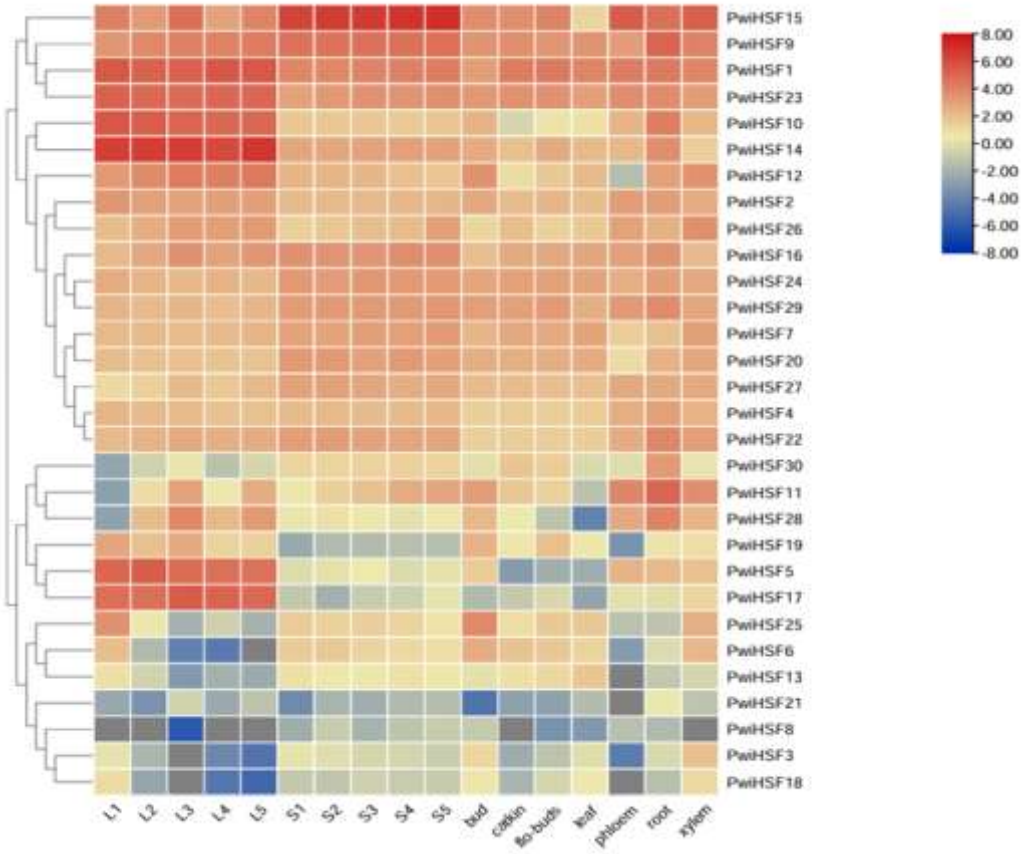

Figure 5. Analysis of expression patterns of PwiHSF at different developmental stages

L1~L5: indicates 5 periods of leaf development; S1~S5: indicates 5 periods of stem development; bud: bud; catkin: inflorescence; flo-buds: flower buds; leaf: leaf; phloem: bast; root: root; xylem: xylem.

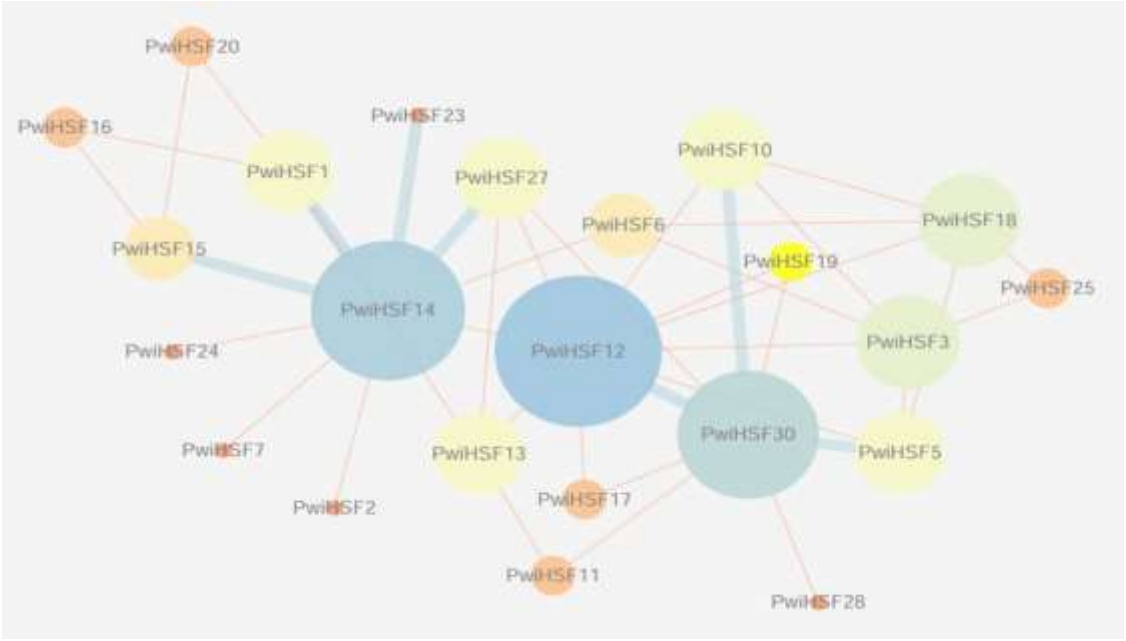

Figure 6. PwiHSF protein interaction analysis

Summary of chapter 3

This study identified 30 HSF genes (PwiHSF1–PwiHSF30) in *Populus wilsonii*, which exhibited significant physicochemical diversity. The encoded proteins ranged in length from 100 to 522 amino acids, with molecular weights of 24.to 58.14 kDA. The genes had isoelectric points (pI) ranging from 4.7 to 9.35 and aliphatic indices ranging from 58.88 to 81.36. The genes also had hydrophilic Grand Average of Hydropathicity (GRAVY) values ranging from -0.896 to -0.408. Chromosomal localization revealed an uneven distribution: chromosomes 1 and 6 each harbored four members. Homologous gene pairs were also identified (e.g., PwiHSF6/25 and PwiHSF11/28), suggesting that family expansion was driven by tandem or segmental duplication events. Phylogenetic analysis classified the genes into three subfamilies: A (12 members), B (two members), and C (16 members). Subfamily C's dominance implies adaptive evolution through functional simplification. Structural and domain analyses revealed subfamily divergence. Subfamily A contained three to four introns and retained full functional modules, including the DNA-binding domain (DBD), the oligomerization domain (OD), and the C-terminal activation domain (CTAD). Motif seven within the CTAD is

enriched in acidic residues and may recruit RNA polymerase II to enhance transcriptional activation. Subfamily B harbored one or two introns and lacked CTAD, retaining only inhibitory motif 5. Subfamily C exhibited intronless structures and retained only the DBD/OD. This suggests that they compensate for their lack of CTAD through protein interactions. Transcriptome profiling revealed tissue-specific expression patterns. Subfamily A members (e.g., PwiHSF7/12) exhibited high expression across multiple tissues, whereas certain genes (e.g., PwiHSF3/5) showed low expression in phloem and roots. Protein interaction network analysis identified PwiHSF12 as a central hub node that forms regulatory modules with PwiHSF14/30 to integrate signaling pathways, such as heat shock and oxidative stress.The P. wilsonii HSF family collectively establishes a multilayered regulatory network through gene duplication, structural diversification, and expression specialization. This provides critical insights into the stress adaptation mechanisms of woody plants.

CONCLUSIN

1. Subfamily distribution and adaptive evolution of species

In this study, 30 HSF genes (PwiHSF) were identified for the first time in *Populus wilsonii*, in which the C subfamily accounted for as high as 53.3%, which was significantly higher than that of Mauve Populus (12%), Minniannan (29% of the 17 members in the C class) and common tobacco (20% in the C class)[32]. These findings suggest a species-specific expansion of the C subfamily in woody plants. Such disparities may be closely related to the ecological adaptive requirements of the species[33]. For instance, the HSF family of P. trichocarpa is predominantly comprised of the A subfamily (e.g., PtHSFA4a plays a role in salt tolerance regulation), whereas the C subfamily of *Populus wilsonii* , a tree species renowned for its resilience to stress, may augment its responsiveness to complex adversities through gene duplications (e.g., tandem duplication events on Chr07). In contrast, the analysis of 22 HSFs in arabidopsis and 25 HSFs in rice revealed a lower percentage of C subfamilies, suggesting evolutionary differences in HSF functional divergence in mono/dicotyledons[34].

2.Evolutionary conservation of gene structure and functional domains

The structural differentiation patterns of the *Populus wilsonii* HSF family exhibit significant differences when compared to those of model plants such as Arabidopsis and rice[35]. For instance, the *Populus wilsonii* A subfamily has maintained the integrity of the DBD, OD, and CTAD structural domains, which exhibit a high degree of similarity to the functional modules (e.g., the acidic amino acid enrichment of CTAD) of the AtHSFA1a, both of which activate HSP gene expression by recruiting RNA polymerase II[36]. However, the *Populus wilsonii* C subfamily has a highly condensed (intronless) gene structure, and the first transmembrane structural domains (TMs) have been identified. This feature was similarly identified in Southeast Sedum SaHsfA4a/c (e.g., nuclear localization and cadmium stress response)8 but not inArabidopsis or rice[37]. suggesting that rapid signaling may be achieved in woody plants through structural simplification. Furthermore[38], the reduced number of subfamily B introns (1-2) and the retention

of the repressor element motif 5 are conserved with the negative regulatory function of AtHSFB15 in Arabidopsis, suggesting evolutionary conservation of this class of genes in balancing HSP overexpression[39].

3.Physicochemical properties and functional differentiation by subcellular localization

The diversity of physicochemical properties (e.g., wide distribution of pIs, differences in lipid coefficients) of *Populus wilsonii* HSF proteins reflects the molecular basis of their functional differentiation[40]. For instance, acidic pI members (e.g., PwiHSF-07) may exhibit a preference for binding to DNA, while neutral/basic pI members (e.g., subfamily C) may play a role in membrane signaling through transmembrane structural domains[41], analogous to the membrane localization and cadmium tolerance exhibited by Southeast Sedum SaHsfA4a.8[42.].Conversely, members such as AtHSFA1d are predominantly localized to the nucleus due to their high hydrophilicity[43], whereas maize ZmHSFC1 regulates cold tolerance through lipid metabolism . Its physicochemical properties may be functionally convergent with the *Populus wilsonii* C subfamily. This diversity may stem from species-specific environmental stresses. For instance, the deep-rooted species known as the *Populus wilsonii* requires its HSF family to respond to both below-ground (e.g., salinity) and above-ground (e.g., drought) stresses[44]. This requirement drives the innovation of a diversity of functional domains.

4. Species specificity of expression pattern and regulation of tissue development

The expression patterns of the *Populus wilsonii* HSF genes showed significant tissue specificity. For example, PwiHSF19 was highly expressed in the formative layer (6.8- to 15.3-fold), which is directly related to the regulation of secondary growth[45]. Similarly, the Minnan PbHsf gene was specifically expressed in the xylem[46]. In contrast, maize HSF21 regulated cold tolerance through lipid metabolism; however, its expression was restricted to low-temperature-responsive tissues[47]. In contrast, the Arabidopsis thaliana AtHSFA1a gene was broadly

expressed under heat stress, highlighting the model plant's global regulatory strategy in response to stress. These differences may reflect the need for woody plants (e.g., *Populus wilsonii*) to coordinate long-term development (e.g., secondary growth) with transient stress responses. In contrast, herbaceous plants (e.g., Arabidopsis) focus more on rapid stress responses[48]. Additionally, the broad-spectrum expression of the A subfamily in *Populus wilsonii* (e.g., PwiHSF7 in the stem tip meristem) is similar to the multitissue response pattern of tobacco NtHSF genes. This suggests a conserved function of this class of genes in basal metabolism.

5.Evolutionary Innovation and Conservation of Protein Interaction Networks

The core hub nodes of the Populus HSF interaction network (e.g., PwiHSF12 interacting with HSP90) are highly conserved with the Arabidopsis AtHSFA1a-HSP90 module. This suggests that core heat stress response mechanisms are evolutionarily stable among species. However, interactions of the Populus C subfamily with MAPK3 and plasma membrane H+-ATPases have not been reported in Arabidopsis. For example, PwiHSF19 regulates ion homeostasis, which is partially similar to the functional module of Southeastern Sedum SaHsfA4a that regulates cadmium transport through Nramp6[49]. Additionally, the temperature-dependent phase separation mechanism of HSF1 discovered at CSU suggests that the C subfamily of *Populus wilsonii* s may respond dynamically to environmental signals through phase transitions. This mechanism has yet to be validated in woody plants[50].

6.Research limitations and future directions

This study systematically resolved the multidimensional characterization of the *Populus wilsonii* HSF family[51]. However, further research is needed to deeply explore its functions, such as combining the study with transgene validation (e.g., CRISPR editing of PwiHSF-19) and multi-omics integration (e.g., single-cell transcriptomes). For instance, the cold tolerance mechanism of maize HSF21 involves lipid metabolism reprogramming; however, it is unclear whether the

Populus wilsonii C subfamily regulates secondary growth via a comparable pathway. Additionally, the HSF1 phase separation mechanism revealed by CSU offers a new perspective for analyzing the subcellular dynamics of HSF in *Populus wilsonii* . Future studies could explore the potential application of *Populus wilsonii* HSFs in heavy metal remediation by drawing on the cadmium stress model of Southeast Sedum[52].

This study systematically analyzed the evolutionary trajectory and expression regulatory network of the HSF gene family in *Populus wilsonii* . Through structural differentiation, expression specificity, and synergistic interaction networks, the HSF gene family has formed a multilevel regulatory mechanism, providing a molecular basis for the adaptation of *Populus wilsonii* to complex environments. The A subfamily acts as a core regulatory module and the C subfamily acts as a functionally simplified branch. These two subfamilies may enhance the organization and developmental ability of *Populus wilsonii* through complementary collaboration. The A subfamily acts as a core regulatory module and the C subfamily acts as a functionally simplified branch that may complement each other to enhance the tissue development of the *Populus wilsonii* . This study provides candidate genes and a theoretical basis for plant tissue development. Its potential applications can be further explored by combining multi-omics and gene editing technologies.

LIST REFERENCES

1. Wang, Xia, Sun, Yiming, and Zhang, Shengkui. "Genome-wide Identification and Expression Analysis of the LBD Gene Family in Chinese Populus. "The Journal of Plant Physiology, in its 2025 volume 61, issue 04, pages 415–428, published by the Chinese Academy of Sciences and the Chinese Society of Plant Physiology, has the following Digital Object Identifier (DOI): 10.13592/j.cnki.ppj.101175.

2. Zhang H-Z. A functional validation study of the PoptrHSFA4a gene in the mauve Populus [D] was conducted. Northeast Forestry University, 2014.

3. Tian X, Xian IS, Ren LH, et al. Identification and Bioinformatics Analysis of HSF Gene Family in Panax ginseng [J]. A study was conducted to examine the current state of research and practice in the field of modern Chinese medicine. The study, which was published in 2024 in the "Modern Chinese Medicine Research and Practice" journal, specifically in volume 38, issue 6, and is available online via the following DOI: 10.13728/j.1673-6427.2024.06.001.

4. Wu SJ, Xu F, Wang GA, et al. "Identification of HSF Gene Family in Cortex Eucommiae and Its Expression Analysis under Adversity Stress" [J]. Henan Agricultural Science, 2024, 53(09): 46-56.doi:10.15933/j.cnki.1004-3268.2024.09.005.

5. Yanyan Wu, Jieyun Liu, Junniu Zhou, et al. Genome-wide characterization of the HSF gene family in Passiflora and its response to high-temperature stress during flower formation. Journal of Tropical Crops,2025,46(04):807-819.

6. Wang, Can; Ren, Zhonghai. A pan-genomic analysis of the HSF family in cucumber is available online. Molecular Plant Breeding,1-13[2025-05-11]. http://kns.cnki.net/kcms/detail/46.1068.S.20241118.1749.009.html.

7. Li Shi. The following paper presents the results of functional studies on the heat stress transcription factor AtHsfA1d[D] from the plant species Arabidopsis. Sichuan Agricultural University, 2020.Digital Object Identifier (DOI): 10.27345/d.cnki.gsnyu.2020.000702.

8. Liu JG. The present study aims to clone and characterize the functionality of resistance-related transcription factors in Begonia octopus and rice. The year of publication was 2007, and the institution was Nanjing Agricultural University.

9. LI Mingyue, LI Mengjun, LI Guoliang, et al. "Heat Tolerance Analysis of Wheat Heat Stress Transcription Factor TaHsfA2-5" (J.). The 42nd issue of the 2022 volume of the Northwest Journal of Botany was published in 2022, and the following article was included therein: "181-189."

10. Zhao Li-Na, Zhang Huaning, Duan Shuo-Nan, et al. Cloning and characterization of the maize ZmHsf04 gene and its regulation of heat tolerance. Journal of Agricultural Biotechnology,2017,25(09):1411-1422.

11. Li, C., Xing, H., Li, C., et al. Chromosome-scale genome assembly provides insights into the molecular mechanisms of tissue development of *Populus wilsonii*. According to the most recent research, the journal Commun Biol, in volume 5, issue 1125, published in 2022, has provided evidence that can be objectively and reliably measured and verified. The specific numerical value of the article's digital object identifier (doi) is 10.1038/s42003-022-04106-0.

12. Ohama, N., Sato, H., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2017). The present study analyzes the transcriptional regulatory network of the plant heat stress response. This analysis is reported in Trends in Plant Science, 22(1), 53-65.

13. Chunguang Li, Qijun Chen, Gaoxin Qi, et al. Arabidopsis heat stress transcription factor AtHsfA2 regulates the expression of stress-responsive genes and improves heat and oxidative stress tolerance [J]. The Chinese Science C Series: Life Science, published in 2005, contains a chapter on the subject of "Life Science" that is available for review. This chapter is located in volume 2005, issue 5, and is pages 17 through 26 in the publication.

14. Gao P, Wang GD, Wei ZG, et al. "Characterization of Rice Heat Stress Transcription Factor Family Genes in Response to Blast Pathogens and Phytohormones" [J]. The Jiangsu Journal of Agriculture, in its 2023 volume 39, issue 03, pages 609 to 621, published the following article.

15. Yanyun Ma.Analysis of the Expression Regulation of MicroRNA156 and Its Target Genes in Aft-type Tomato[D]. Northeast Forestry University, 2022. Digital Object Identifier (DOI): 10.27009/d.cnki.gdblu.2022.001785.

16. N. Zhang, Y.H. Wang, Z.M. Wang, et al. Progress in the Study of a Family of Heat-Excited Transcription Factors in Plants [J]. The Journal of Biological Engineering, in its 37th issue of 2021, published a set of research findings on pages 1155 to 1167. The Digital Object Identifier (DOI) for the complete text is as follows: 10.13345/j.cjb.200367.

17. Tian YR. A study was conducted to investigate the mechanism of Populus PeJAZ2-mediated ABA signaling in regulating drought resistance in Populus[D]. Beijing Forestry University, 2021. Digital Object Identifier (DOI): 10.26949/d.cnki.gblyu.2021.000618.

18. Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. Nucleic Acids Research, 39(suppl_2), W29-W37.

19. Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics, 9(1), 1-13.

20. Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol. Plant 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

21. Wenhai L ,Xinghao T ,Jingshu L , et al.Genome wide investigation of Hsf gene family in Phoebe bournei: identification, evolution, and expression after abiotic stresses [J].Journal of Forestry Research,2023,35(1):

22. Guo Cun, Wang Qi, Li Xiaoxu, et al. Genome-wide identification and analysis of a family of heat-excited transcription factors in common tobacco[J]. Chinese Tobacco Science, 2022, 43(03): 47-56. doi:10.13496/j.issn.1007-5119.2022.03.008.

23. Qi Z, Jing G, Yanli D, et al. Heat shock transcription factor (Hsf) gene family in common bean (*Phaseolus vulgaris*): genome-wide identification,

phylogeny, evolutionary expansion and expression analyses at the sprout stage under abiotic stress[J].BMC Plant Biology,2022,22(1):33-33.

24. Gao L, Pan L, Shi Y, et al.Genetic variation of a heat shock transcription factor modulates cold tolerance in maize.[J].Molecular plant,2024,17(9):1423-1438.

25. Chen Shuangshuang. An analysis of the SaHsf gene family was conducted in hyperaccumulator Southeast Sedum, and a functional study of cadmium tolerance in SaHsfA4a/c[D] was undertaken. China Academy of Forestry Science,2018.

26. Ren, Q., Li, L., Liu, L. *et al.* The molecular mechanism of temperature-dependent phase separation of heat shock factor 1. *Nat Chem Biol* (2025). https://doi.org/10.1038/s41589-024-01806-y

27. Zhai Ruibo. The objective of this study is to identify and functionally analyze some genes of the HSF family in mulberry. Jiangsu University of Science and Technology, 2020. Digital Object Identifier (DOI):10.27171/d.cnki.ghdcc.2020.000869.

28. Wang Q, Zhang Z, Guo C, Zhao X, Li Z, Mou Y, Sun Q, Wang J, Yuan C, Li C, Cong P and Shan S (2023) Hsf transcription factor gene family in peanut (Arachis hypogaea L.): genome-wide characterization and expression analysis under drought and salt stresses. Front. Plant Sci. 14:1214732. doi: 10.3389/fpls.2023.1214732

29. Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37, W202–W208. doi: 10.1093/nar/gkp335

30. Bechtold, U., Albihlal, W. S., Lawson, T., Fryer, M. J., Sparrow, P. A., Richard, F., et al. (2013). Arabidopsis HEAT SHOCK TRANSCRIPTION FACTORA1b overexpression enhances water productivity, resistance to drought, and infection. J. Exp. Bot. 64, 3467–3481. doi: 10.1093/jxb/ert185

31. Begum, T., Reuter, R., Schoffl, F. (2013). Overexpression of AtHsfB4 induces specific effects on root development of Arabidopsis. Mech. Dev. 130, 54–60. doi: 10.1016/j.mod.2012.05.008

32. Chen, H., Hwang, J. E., Lim, C. J., Kim, D. Y., Lee, S. Y., Lim, C. O. (2010). Arabidopsis DREB2C functions as a transcriptional activator of HsfA3 during the heat stress response. Biochem. Biophys. Res. Commun. 401, 238–244. doi: 10.1016/j.bbrc.2010.09.038

33. Cui, Y. N., Li, X. T., Yuan, J. Z., Wang, F. Z., Guo, H., Xia, Z. R., et al. (2020). Chloride is beneficial for growth of the xerophyte Pugionium cornutum by enhancing osmotic adjustment capacity under salt and drought stresses. J. Exp. Bot. 71, 4215–4231. doi: 10.1093/jxb/eraa158

34. Doring, P., Treuter, E., Kistner, C., Lyck, R., Chen, A., Nover, L. (2000). The role of AHA motifs in the activator function of tomato heat stress transcription factors HsfA1 and HsfA2. Plant Cell 12, 265–278. doi: 10.2307/3870927

35. Duan, S., Liu, B., Zhang, Y., Li, G., Guo, X. (2019). Genome-wide identification and abiotic stress-responsive pattern of heat shock transcription factor family in Triticum aestivum l. BMC Genomics 20, 257. doi: 10.1186/s12864-019-5617-1

36. El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The pfam protein families database in 2019. Nucleic Acids Res. 47, D427–D432. doi: 10.1093/nar/gky995

37. Friedrich, T., Oberkofler, V., Trindade, I., Altmann, S., Brzezinka, K., Lamke, J., et al. (2021). Heteromeric HSFA2/HSFA3 complexes drive transcriptional memory after heat stress in Arabidopsis. Nat. Commun. 12, 3426. doi: 10.1038/s41467-021-23786-6

38. Gao, Y. F., Liu, J. K., Yang, F. M., Zhang, G. Y., Wang, D., Zhang, L., et al. (2020). The WRKY transcription factor WRKY8 promotes resistance to pathogen infection and mediates drought and salt stress tolerance in Solanum lycopersicum. Physiol. Plant 168, 98–117. doi: 10.1111/ppl.12978

39. Kotak, S., Port, M., Ganguli, A., Bicker, F., von Koskull-Doring, P. (2004). Characterization of c-terminal domains of Arabidopsis heat stress transcription factors (Hsfs) and identification of a new signature combination of plant class a hsfs with AHA and NES motifs essential for activator function and intracellular localization. Plant J. 39, 98–112. doi: 10.1111/j.1365-313X.2004.02111.x

40. Baniwal, S. K., Bharti, K., Chan, K. Y., Fauth, M., Ganguli, A., Kotak, S., et al. (2004). Heat stress response in plants: a complex game with chaperones and more than twenty heat stress transcription factors. J. Biosci. 29, 471–487. doi: 10.1007/BF02712120

41. Barcelos, E., de Almeida Rios, S., Cunha, R. N. V., Lopes, R., Motoike, S. Y., Babiychuk, E., et al. (2015). Oil palm natural diversity and the potential for yield improvement. Front. Plant Sci. 6:190. doi: 10.3389/fpls.2015.00190

42. Boureima, S., Oukarroum, A., Diouf, M., Cissé, N, and Van Damme, P. (2012). Screening for drought tolerance in mutant germplasm of sesame (Sesamum indicum) probing by chlorophyll a fluorescence. Environ. Exp. Bot. 81, 37–43. doi: 10.1016/j.envexpbot.2012.02.015

43. Cannon, S. B., Mitra, A., Baumgarten, A., Young, N. D., and May, G. (2004). The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana. BMC Plant Biol. 4:10. doi: 10.1186/1471-2229-4-10

44. Chauhan, H., Khurana, N., Agarwal, P., and Khurana, P. (2011). Heat shock factors in rice (Oryza sativa L.): genome-wide expression analysis during reproductive development and abiotic stress. Mol. Genet. Genomics 286, 171–187. doi: 10.1007/s00438-011-0638-8

45. Döring, P., Treuter, E., Kistner, C., Lyck, R., Chen, A., and Nover, L. (2000). The role of AHA motifs in the activator function of tomato heat stress transcription factors HsfA1 and HsfA2. Plant Cell 12, 265–278. doi: 10.1105/tpc.12.2.265

46. Dossa, K., Niang, M., Assogbadjo, A. E., Cissé, N., and Diouf, D. (2016a). Whole genome homology-based identification of candidate genes for drought resistance in (Sesamum indicum L.). Afr. J. Biotechnol. 15, 1464–1475. doi: 10.5897/AJB2016.15420

47. Dossa, K., Wei, X., Li, D., Zhang, Y., Wang, L., Fonceka, D., et al. (2016b). Insight into the AP2/ERF transcription factor superfamily in sesame (Sesamum indicum) and expression profiling of the DREB subfamily under drought stress. BMC Plant Biol. 16:171. doi: 10.1186/s12870-016-0859-4

48. Eskandari, H., Zehtab-Salmasi, S., Golezani, K. G., and Gharineh, M. H. (2009). Effects of water limitation on grain and oil yields of sesame cultivars. Food Agric. Environ. 7, 339–342.

49. Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). The Pfam protein families database. Nucleic Acids Res. 42, D222–D230. doi: 10.1093/nar/gkt1223

50. Giorno, F., Guerriero, G., Baric, S., and Mariani, C. (2012). Heat shock transcriptional factors in Malus domestica: identification, classification and expression analysis. BMC Genomics 13:639. doi: 10.1186/1471-2164-13-639

51. Giorno, F., Wolters-Arts, M., Grillo, S., Scharf, K. D., Vriezen, W. H., and Mariani, C. (2010). Developmental and heat stress-regulated expression of HsfA2 and small heat shock proteins in tomato anthers. J. Exp. Bot. 61, 453–462. doi: 10.1093/jxb/erp316

52. Guo, M., Lu, J. P., Zhai, Y. F., Chai, W. G., Gong, Z. H., and Lu, M. H. (2015). Genome-wide analysis, expression profile of heat shock factor gene family (CaHsfs) and characterisation of CaHsfA2 in pepper (Capsicum annuum L.). BMC Plant Biol. 15:151. doi: 10.1186/s12870-015-0512-7