**Original Research**

**JrbHLH Gene Family: Genome-Wide Identification and Transcriptional Expression in Persian Walnut (Juglans regia L.)**

Irfan Ullah1#, Na Hou2#, Feng Yan1, Hanif Khan1, Xurong Bai1, Sana Ullah1, Ming Yue3, Peng Zhao1*

1Laboratory of Resource Biology and Biotechnology in Western China, Ministry of Education, College of Life Sciences, Northwest University, Xi’an, Shaanxi 710069, China
2Guizhou Academy of Forestry, Guiyang, China
3Xi’an Botanical Garden of Shaanxi Province, Xi’an, Shaanxi 710061, China

Received: 14 March 2020
Accepted: 29 May 2020

**Abstract**

In plants, the basic helix-loop-helix (bHLH) transcription factors play key roles in diverse biological processes. Genome-wide comprehensive and systematic analyses of bHLH proteins have been well conducted in Arabidopsis (AtbHLH), rice (OsbHLH) and tomato (SlbHLH) species. However, there are no reports of bHLH family genes have been functionally characterized in Persian walnut (Juglans regia). We identified a total of 189 JrbHLH genes in J. regia, which are clustering into 18 subfamilies. All JrbHLH contained the conserved domains HLH. A total of 69 JrbHLH genes were expanded as whole genome duplications (WGD). The syntenic analysis was conducted for Persian walnut (J. regia) with other two woody plants (P. trichocarpa and O. europaea) to infer the bHLH genes’ evolutionary relationship between these species. We identified a total of 42 pairs of orthologous bHLH genes between J. regia and P. trichocarpa, while only 29 orthologous gene pairs between J. regia and O. europaea, while 25 collinear gene pairs were found in three species J. regia, O. europaea and P. trichocarpa. Furthermore, JrbHLHs had different expressed patterns between reproduction and vegetative tissues based on the transcriptome expression profiles of female flowers, male flowers, leaves, and hulls, respectively. Specifically, both transcriptome data and quantitative polymerase chain reaction (qRT-PCR) assessment showed that two JrbHLH genes (JrbHLH3-5 and JrbHLH13-13) were highly represented in female and male flowers, while one JrbHLH gene (JrbHLH13-4) was highly expressed in leaves. Genome-wide identification, gene structure, phylogeny and expression analysis of the JrbHLH

# These authors contributed equally to this work.
*e-mail: pengzhao@nwu.edu.cn

DOI: 10.15244/pjoes/123113
ONLINE PUBLICATION DATE: 2020-11-26
Introduction

The basic helix-loop-helix (bHLH) gene, its transcriptional factors were first observed in mammals and subsequently detected in eukaryotic species [1-3]. In many species, members of the bHLH gene family play a role in eukaryotic organisms with transcriptional regulators [4, 5]. The bHLH superfamily is the second-largest transcription factor (TF) family, defined by a highly conserved bHLH domain with a length of approximately 50-60 amino acids and divided into two separate areas, the critical region and the HLH region [2, 5]. The base area consists about 10-15 amino acids for DNA binding located at N-terminus domain and C-terminal site involving of two amphipathic α-helices acting as a dimerization domain [6-8]. Concerning the highly conserved bHLH domain, there are also other preserved motifs in some bHLH superfamily [9, 10].

Genome-wide comprehensive and systematic analyses of bHLH proteins in plants have been carried out in many annual plant species, such as rice (Oryza sativa), maize (Zea mays), thaliana (Arabidopsis thaliana), and tomatoes (Solanum lycopersicum) [11, 12], including fruitful plants species such as apple (Malus domestica) and poplar (Populus trichocarpa) [2, 13]. However, the bHLH genes have not been functionally reported in Persian walnut (Juglans regia) previously. The function of JrbHLH proteins in plants is also involved in biological and biochemical processes, including the development of petals and the reaction to drought and salt stress and biosynthesis [13-15]. Besides, bHLH proteins are involved in the regulation of fruit growth and anthocyanin biosynthesis in flowers and fruit [16, 17].

Due to the importance of the bHLH genes in various physiological programs, it would be of interest to make a systematic investigation of the bHLH family in Persian walnut. The walnut genome sequencing provided an opportunity to reveal the organization, expression and evolutionary traits of walnut bHLH gene family at the genome-wide level [18]. In the present study, we identified 189 walnut bHLH genes and classified them into 18 main groups. The comprehensive analysis including the exon-intron organization, motif compositions, gene duplications, phylogenetic and synteny analysis were further investigated. Expression and quantitative polymerase chain response (qRT-PCR) analysis was performed to identify involvement of specific bHLH gene family members in flowering process in walnut. This study provided valuable clues for functional characterization of bHLH gene family members in walnut.

Material and Methods

Identification of bHLH Transcription Factors in J. regia

The whole protein sequence of walnuts downloaded from the NCBI Genbank (https://treegenesdb.org/FTP/Genomes/Jure/v1.4/annotation/). The members of Arabidopsis JrbHLH family were downloaded from the TAIR website (https://www.arabidopsis.org/index.jsp). We used Arabidopsis bHLH protein sequences as a query using a local alignment search tool (BLAST) to search against walnut protein sequences, considering those with an E-value of less than $1 \times 10^{-10}$. Therefore, we implemented the Markov hidden profile model (HMM) in HMMER v.3.2.1 for window, with default parameters for the search of bHLH proteins and domains in the protein family (Pfam) database [19].

Phylogenetic, Domain Analysis, Motif, Gene Structure Analysis of JrbHLH Transcription Factors

The neighbor-joining (NJ) tree of 189 bHLH transcription factors were constructed using MEGA v.7.0 software [16, 20], with the pairwise deletion of 1,000 bootstraps and a Poisson model, in the presence of potential domains using the Pfam webserver [21]. The same domains were also obtained from Pfam with an E-value cutoff of 1.0 to validate the result using the simple modular architecture research tool (SMART) program through BLASTN, the whole genome sequence of J. regia were examined JrbHLH transcription factors. Similarly, from the NCBI the conserved domain database (CDD) [22], and the whole CDS database was downloaded (https://treegenesdb.org/FTP/Genomes/Jure/v1.4/annotation/). The exon and intron structure were displayed using the online gene structure display server (http://gsds.cbi.pku.edu.cn/) [11]. The genome explorer (https://www.ncbi.nlm.nih.gov/genome/) was used to report the specific sequences of walnut genes and the motif identification was done by MEME software with default parameters, the maximum number of motifs was (20) and the optimum motifs were (30-50) [23].

Synteny Analysis

To identify the potential homologous gene pairs (E<1 × 10^{-5}, top 5 matches) across multiple genomes using BLASTP. The identification of syntenic chains was determined using homologous gene pairs through MCScanX and detected the duplicate pairs of two
woody plants (*P. trichocarpa* and *O. europaea*) [24]. These detected duplicate gene pairs further analyzed through MCScanX including whole-genome duplication (WGD)/segmental and tandem duplication of walnut genes pair with *P. trichocarpa* and *O. europaea* species gene pairs.

**GO Annotation**

The Blast2GO v2.5 with a cutoff E-value of 1×10^{-6} was used to conduct GO annotations [25]. First, we use the *JrbHLH* protein sequence to perform a blastp with an E-value of 1e-05. The GO mapping analysis was used to obtain the GO annotation from the *JrbHLH* members the GO enrichment analysis was carried out via the online GO enrichment program on the Omicshare website (https://www.omicsshare.com/tools/Home/Sof/gogsea).

Interaction Network of *JrbHLH* Proteins

The function of *JrbHLH* protein based on *bHLH* protein predicted in *Arabidopsis* due to lack of relevant information on walnut protein. The 189 walnut *bHLH* proteins aligned with *Arabidopsis* *bHLH* proteins using the Blastp program with an E-value of 1e-05. The *JrbHLH* protein interaction network was developed with *Arabidopsis* using homologous *bHLH* proteins. The network was developed with default parameters on the String database (https://string-db.org/) using the input proteins of *J. regia* and six predicted input proteins [26].

**Plant Materials, Treatments and Collections**

A total of 26 samples from walnut were collected in this study, including 15 female flowers in different stages, the first opening of female flowers took place on April 10, April 15 and 22, and the full opening of female flowers took place (specifically, on April 15 and 22, stigma was not fully developed) and May 1 was the last date, 3 male flowers, 3 leaves, and 3 hulls. The female flowers (consist of 3 replicates) were collected on March 23, April 1, 8, 16 and April 23 respectively, the male flowers were collected on April 10, 11 and May 2 respectively, the leaves and the hull transcriptome data were as a query to identify the gene family members in walnut, the results showed that a total of 189 genes belong to the walnut *bHLH* gene family. The *JrbHLHs* were used to construct a phylogenetic tree with NJ method, it is clear that the *JrbHLHs* were divided into 18 subfamilies (Fig. 1) classified into 18 groups, and the number in these groups were unevenly distributed (Fig. 1). Based on their phylogenetic relationship walnut *bHLH* genes were divided into 18 subfamilies (Fig. 1). Subfamily No. 13 reported the minimum number of *JrbHLH* genes (25), while family No. 11 reported the highest number of *JrbHLH* genes (4).

**Results and Discussion**

The conservated domain of *bHLH* genes from *Arabidopsis* were as a query to identify the gene family members in walnut, the results showed that a total of 189 genes belong to the walnut *bHLH* gene family. The *JrbHLHs* were used to construct a phylogenetic tree with NJ method, it is clear that the *JrbHLHs* were classified into 18 groups, and the number in these groups were unevenly distributed (Fig. 1). Based on their phylogenetic relationship walnut *bHLH* genes were divided into 18 subfamilies (Fig. 1). Subfamily No. 13 reported the highest number of *JrbHLH* genes (25), while family No. 11 reported the minimum number of *JrbHLH* genes (4). In this study, we identified a total of 189 *bHLH* genes in Persian walnut, compared to other plant members such as *Arabidopsis thaliana*, which contains 133 *bHLH* genes [4, 16] and significant plants such as peanuts, potatoes, wheat, beans, rice, carrots, and tomatoes have *JrbHLH* genes 132, 124, 188, 225, 155, 167, 146 and 159 [29, 30] (Fig. 1).

**Gene Structure and Protein Domains of the *bHLH* Gene Family of Persian Walnut**

Structures including exons and introns were mapped to compare the structural parts of the 189 *JrbHLH* genes (Fig. 2). The precise number and location of the domains of each walnut protein were determined.
using the data available in the Pfam database (Fig. 2b). Conserved domain and motif valuation indicated that Motif 1 was present in all prevalent walnut \textit{JrbHLH} genes. The subfamily of 16 and 17 consists of various motifs, such as 7, 8, 10, 15, 16, 17, and 19 also, as shown in the 18 family includes only motif 5, as well as a conserved motif (Fig. 2b). The analysis of the \textit{JrbHLH} gene structure revealed large gene variations. It has been noted that the number of exons of Persian walnut \textit{JrbHLH} transcript varies from 1 to 17 (Fig. 2c).

MEME analysis has shown that the walnut \textit{JrbHLH} gene family consists of twenty-four residues of amino acids has been conserved and has unique domains for 189 walnut \textit{JrbHLH} gene families. Among all these twenty-four conserved domains, one conserved domain (\textit{HLH}) is made up of walnut \textit{JrbHLH} genes, while the other 23 are unique domains. All of these conserved amino acid residues are compatible with previous \textit{JrbHLH} domain research in other crops, such as peaches twenty-nine conserved amino acid residues. Based on conserved domains, walnut \textit{JrbHLH} genes consist of 23 unique domains and one conserved domain (\textit{HLH}) found in all \textit{JrbHLH} genes [4, 22].

Synteny Analysis of \textit{JrbHLH} Genes

The syntenic analysis was conducted for Persian walnut (\textit{J. regia}) with other two woody plants (\textit{P. trichocarpa} and \textit{O. europaea}) to infer \textit{bHLH} genes evolutionary relationship between these species (Fig. 3). We identified 42 pairs of orthologous \textit{bHLH} genes between \textit{J. regia} and \textit{P. trichocarpa}, while only 29 orthologous gene pairs between \textit{J. regia} and \textit{O. europaea} (Fig. 3). A total of 24 collinear gene pairs identified between \textit{J. regia} and \textit{P. trichocarpa} were not found between \textit{J. regia} and \textit{O. europaea}, 3 collinear gene pairs identified between \textit{J. regia} and \textit{O. europaea} were not found between \textit{J. regia} and \textit{P. trichocarpa}, importantly, 25 collinear gene pairs were found in three species \textit{J. regia}, \textit{O. europaea} and \textit{P. trichocarpa} (Fig. 3). Gene organization plays a vital role in the evolution of multigene families [28]. Analysis of the gene structure indicated that exons present in the walnut \textit{JrbHLH} gene ranged from 1-17 considered to exhibit the lower expression levels in plants [31, 32]. In the case of the \textit{JrbHLH} gene in Persian walnuts, almost all gene structures have an intron 0 phase; therefore, the \textit{JrbHLH} gene is evolutionarily preserved. The position of intron at 1, 2, 3, 4 and 5 were present throughout the \textit{JrbHLH} family gene showed that the \textit{JrbHLH} family is evolutionarily protected [14]. Since the evolutionary studies of \textit{JrbHLH} transcription factors are limited in walnut need to be explored in further study (Fig. 3). There are collinear genes between the walnut and \textit{P. trichocarpa} and \textit{O. europaea} these results suggested that the \textit{bHLH} genes may have evolved from the Persian ancestor in different plants [29, 33, 34].
Fig. 2. The Phylogenetic relationship, motif compositions, and gene structure of JrbHLH transcription factors. a) The phylogenetic relationships of JrbHLHs based on NJ method. The various colors characterized the ten subfamilies. b) motif compositions of JrbHLHs. Gray lines indicated non-conserved sequences, and colored boxes represent conserved motifs. The motifs are displayed proportionally in each protein. c) gene structure of JrbHLHs. CDS represents coding sequence, green boxes indicate CDS, and gray lines represent Introns; 0, 1, and 2 represent different types of phase. Phase 0: located between two consecutive codons; Phase 1: splitting codon between the first and second nucleotides; Phase 2: between the second and third nucleotides of a codon.
The Expansion and Evolution of the \textit{JrbHLH} Gene Family of Persian Walnut

The analysis of GO enrichment based on 189 \textit{JrbHLH} proteins annotated in the GO database. In the category of biological processes, bio-regulation, metabolic processes, cellular processes, and stimulus-response are significantly enriched terms. In the cellular component category, cell, organelle and cell parts are significant. GO terms for the transcription factor activities of nucleic acid binding were highly represented in the molecular function category. Besides, the most GO term enriched by members of the \textit{JrbHLH} is GO: 003674 (molecular function). Tandem duplication, segmental duplication, and WGD were most likely chosen by gene families as forms of expansion [10]. However, a total of 69 \textit{JrbHLHs} were duplicated as WGD, and no gene pair experienced tandem duplication events. These results indicated that the evolutionary expansion patterns of \textit{bHLH} transcription factor family members were duplicated by WGD events (Fig. 1 and 2). These finding contrasts with several previous reports in which a similar phenomenon was analyzed [27, 35]. For example, the expansions of the Gossypium, \textit{bHLH} genes were tandem duplication events [29, 34].

Transcriptome Expression Profile Analysis, qRT-PCR, and Interaction Work of \textit{JrbHLH} Genes in Persian Walnut

Transcript expression analysis of different tissues of Persian walnut (female flower, male flower, young leaf, and hull) showed that the \textit{JrbHLHs} expression levels were different in female flowers, and male flowers, leaves, and hulls (Fig. 4). Expression profiling revealed that a total of 28 \textit{JrbHLH} genes have higher expression in multiple tissues of Persian walnut [the values of FPKM (fragments per kilobase per million) were more than 60] (Fig. 4). A total of 4 and 8 members of \textit{JrbHLH} genes show higher expression in female flowers and male flowers, respectively. For vegetative tissues transcriptome expression levels, we found that a total of 4 and 12 \textit{JrbHLH} genes were expressed highly in leaves, while 12 \textit{JrbHLH} genes were expressed highly in hulls. Among these high expressed genes, the \textit{JrbHLH13-10}, \textit{JrbHLH14-15}, and \textit{JrbHLH14-8} showed highly expression levels in both tissues (Fig. 4). Moreover, five \textit{JrbHLH} genes were extremely high expressed (the values of FPKM were more than 400) (Fig. 4). Furthermore, by compared the differences between the \textit{JrbHLH} gene expression levels in reproductive tissues and vegetative tissues, and we found that the two genes (\textit{JrbHLH13-13} and \textit{JrbHLH14-8}) both expressed high in reproductive tissues (female flowers and male flowers), while three genes (\textit{JrbHLH13-4}, \textit{JrbHLH14-3}, and \textit{JrbHLH14-4}) both expressed high in vegetative tissues (leaves and hulls) (Fig. 4). In woody plants, previous research showed that the \textit{JrbHLH} gene plays an important role in flowering, such as peach and other fruit varieties, and some \textit{JrbHLH} superfamily genes may be involved in fruit development [16]. This finding is consistent with previous findings of the \textit{JrbHLH} protein, which also plays a regulatory role in tomato fruit ripening [29]. Especially in the reproductive development members of rice and \textit{A. thaliana}, the functional subfamilies, OsbHLH142, AbhHLH091, OsbHLH141, AbhHLH010 and AbhHLH089 of these genes have similar expression patterns. While from the present we
reported \textit{JrbHLH13-4}, \textit{JrbHLH1-15} and \textit{JrbHLH14-8} in reproductive and vegetative tissues (Fig. 4). The same applies in the potato as the \textit{StbHLHs} (\textit{StbHLH26, 85 and 99}) have relatively high expression in flowers, in this study, 41 are highly expressed in flower tissues. These genes may be involved in flower development [19, 36, 37].

The different expression patterns involving gradual increases and decreases observed in this study illustrate that some \textit{bHLH} superfamily genes
may be associated with fruit development. This finding was consistent with previous findings that bHLH proteins play a regulatory role in fruit ripening in walnut and other fruit species [37]. Additionally, the findings of this study are in agreement with transcriptomic results [37]. During the developmental stage (1st-5th) of the female flower, the transcriptome expression levels of JrbHLH genes were increased i.e. JrbHLH3-5, JrbHLH13-4, JrbHLH13-13, and JrbHLH14-4 [37] (Fig. 5a). We further verified the difference gene expression level patterns in Persian walnut flowers and leaves used by qRT-PCR experiment (Fig. 5b). The results showed that JrbHLH3-5 and JrbHLH13-13 were highly expressed in female compared to male flowers, while JrbHLH13-25 expressed highly in leaves of J. regia (Fig. 5b). These genes differentially expressed in female flowers, male flowers and leaves (Fig. 5b), which could be subsequently prioritized in plant functional studies for further analysis. Based on transcriptome and relative (qRT-PCR) expression, the walnut JrbHLH genes JrbHLH3-5 and JrbHLH13-13 show high expression in the female flower followed by the male flower, while the JrbHLH13-4 shows greater expression in the vegetative part of the leaves followed by the male flower and also decreased expression in the female walnut family members of the JrbHLH gene. Therefore, further characterization of the 28 JrbHLHs is highly important and will provide a new insight to understand the molecular mechanism which may play some roles in the expression of the Persian walnut in multiple tissues.

Conclusions

In summary, we have recognized 189 JrbHLH genes in Persian walnut (Juglans regia) in this research. Phylogenetic analysis showed the division of the JrbHLH gene into 18 subfamilies. Also, expression profiling revealed that the JrbHLH gene showed distinct patterns of expression in distinctly prevalent walnut tissues. A total of 10 genes (JrbHLH14-8, JrbHLH13-2, JrbHLH13-13, JrbHLH13-10, JrbHLH13-1, JrbHLH16-8, JrbHLH16-9, JrbHLH16-4, JrbHLH13-5, and JrbHLH11-9) were highly reflected in multiple phases of female flowers and male flowers, which may play a significant role in Persian walnut flowers. During the flowering of female flowers, it was shown that some genes had increased expression rates as females grew. Overall, 28 members of the JrbHLH gene family are highly expressed in reproductive and vegetative tissues, while JrbHLH13-4, JrbHLH1-15, and JrbHLH14-8 have been highly expressed in all tissues. Specifically, both transcript information and (qRT-PCR) analysis showed that two JrbHLH genes (JrbHLH3-5, JrbHLH13-13) were extremely expressed in female and male flowers, while one JrbHLH gene (JrbHLH13-4) was strongly expressed in leaves. Our outcome was consistent with that in Persian walnuts, but without any combination with other tissues, the expression of the JrbHLH gene in male flowers was fully associated with female flowers.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 41471038; No. 31200500; No. 31860215), the Program for Excellent Young Academic Backbones funding by Northwest University, Shaanxi Academy of Science Research Funding Project (2019K-06), and Natural Science Foundation of Shaanxi Province of China (2019JM-008); Guizhou walnut Engineering Technology Research Center (Qiankehe Platform Talents [2019] 5202), the Propagation and high-yield cultivation demonstration of improved variety of Qianlin He from 3 to 6 walnut ([2018] TG03;[2018]4320).

Conflict of Interest

The authors declare no conflict of interest.

References


