Research Square

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

Genome-wide Analysis of the bHLH Transcription Factor Family and Functional Identification of bHLH Genes Related to Puerarin Biosynthesis in Pueraria lobata var. thomsonii (Benth.)

Liang Xiao
Cash Crops Research Institute, Guangxi Academy of Agricultural Sciences
Ding Huang
Guangxi University of Chinese Medicine
Zhengdan Wu
Cash Crops Research Institute, Guangxi Academy of Agricultural Sciences
Xiaohong Shang
Cash Crops Research Institute, Guangxi Academy of Agricultural Sciences
Sheng Cao
Cash Crops Research Institute, Guangxi Academy of Agricultural Sciences
Wendan Zeng
Cash Crops Research Institute, Guangxi Academy of Agricultural Sciences
Liuying Lu
Cash Crops Research Institute, Guangxi Academy of Agricultural Sciences
Pingli Shi
Cash Crops Research Institute, Guangxi Academy of Agricultural Sciences
Huabing Yan
hbyan1979@163.com

Cash Crops Research Institute, Guangxi Academy of Agricultural Sciences

Research Article

Keywords: Pueraria thomsonii, Genome-wide analysis, bHLH transcriptional factors, Expression analysis, Pueraria biosynthesis pathway

Posted Date: June 2nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1703982/v1

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

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Plant Gene on March 1st, 2023. See the published version at https://doi.org/10.1016/j.plgene.2022.100390.

Abstract

Plant basic helix-loop-helix (bHLH) transcription factors are involved in diverse biological process. Several bHLH transcription factors have been functionally characterized in various crop species. However, no systematic and comprehensively identification of *bHLH* members has been in *Pueraria lobata var. thomsonii* (Benth.) (*P. thomsonii*), a traditional Chinese herb that is an excellent source of puerarin. Phylogenetic analysis showed that 219 *bHLH* genes were identified in the *P. thomsonii* and classified into 13 subfamilies. All the *PlbHLH* genes were distributed on the 11 chromosomes unevenly and the number of *PlbHLH* genes on each chromosome varied from 12 to 44. Both of tandem and segmental duplications were key factors driving *bHLH* gene family expansion in *P. thomsonii*. The gene structure analysis indicated that 20 conserved motifs were found in *PlbHLH* gene family, 185 *PlbHLH* members contain more than two introns, and genes within the same subfamily appeared to share similar intron-exon gene structures. The Ka/Ks analysis indicated *PlbHLH* gene family undergo purifying selection during evolution. The hierarchical clustering analysis indicated that 41 *PlbHLHs* showed root-specific expression patterns, and seven of them exhibited up-regulated expression pattern after MeJA treatment within 24 h. Using qRT-PCR validation, five puarian synthesis genes had a similar expression pattern with the seven *PlbHLH* genes in response to MeJA within 24 h. Together, our study provided a good foundation for further research into the regulatory mechanism of the seven candidate PlbHLH proteins for pueraria biosynthesis in *P. thomsonii.*

Introduction

The basic/helix–loop–helix (bHLH) transcription factors have been widely found in all three eukaryotic kingdoms (Duek and Fankhauser 2005; Carretero-Paulet et al. 2010). bHLH transcription factors are identified by a conserved 50–60 amino acid sequence responsible for DNA binding, according to the typical bifunctional structure, namely the basic and HLH regions. The basic region is a 15–20 residues located at the N-terminal end, which determine the recognition of the cis-acting element E-box (5'-CANNTG-3') or G-box (CACGTG) and allows the binding of the HLH region (Li et al. 2006). The basic region contains about 15 amino acids of which six are typically basic amino acid residues. In contrast, the HLH domain, located at the C-terminal end, is mainly composed of two relatively conserved amphipathic helices with a linking loop that varies in length or sequence. bHLH transcription factors play crucial role in plant second metabolic regulation (Heim et al. 2003).

Since the bHLH domain was initially defined, a variety of bHLH subgroups have been identified based on the frequencies of 19 conserved amino acids in the bHLH consensus motif (Atchley and Terhalle 1999). The first bHLH, Lc (L-myc), is required for the flavonoid/anthocyanin biosynthetic pathway in maize (*Zea mays*) (Ludwig et al. 1989). Subsequently, many more functions for plant bHLH genes have been identified, including the regulation of light signaling (Roig-Villanova et al. 2007); hormone signal transduction (Lee et al. 2006); responses to abiotic stress (Jin et al. 2016). In addition to this, bHLHs participate in regulating flavonoid biosynthesis (Matus et al. 2010). The heterogeneous overexpression of liverwort (*Plagiochasma appendiculatum*) *PabHLH1* in *Arabidopsis* could activate the synthesis of

both flavonoids and anthocyanins, through the up-regulation of early and late structural genes in the synthesis pathway of flavonoids (Zhao et al. 2019). Two *bHLH* genes, *bHLH003* and *bHLH007*, were also found to be related to anthocyanin or flavonoids synthesis in grape (*Vitis vinifera*) (Wang et al. 2018a).

So far, extensive study on the *bHLH* gene family members in many crop species has increased our understanding of their transcriptional regulatory mechanism (Guo et al. 2021; Qian et al. 2021). However, characteristics of *bHLH* gene family in *Pueraria lobata var. thomsonii Pueraria lobata var. thomsonii* (Benth.) (hereinafter referred as *P. thomsonii*), belong to the leguminosae family, has long been used as traditional Chinese herb, have not yet been classified. The roots of *P. thomsonii* have long been used for treating fever, indigestion, indigestion, toxicosis, and liver damage from alcohol abuse in traditional Chinese medicine (Wang et al. 2020), which was recorded in The Divine Husbandman's Classic of Materia Medica (*Shen Nong Ben Cao Jing)* complied in the Eastern Han Dynasty (25–250 AD) (Zhao and Guo 2018). Puerarin, the 8-C-glucoside of daidzein, was synthesized through the phenylpropanoid pathway under the actions of phenylalanine ammonia-lyase (PAL), cinnamic acid 4-hydroxylase (C4H), 4-coumarate: coenzyme A ligase (4CL), chalcone synthase (CHS), chalcone isomerase (CHI), flavanone 3-hydroxylase (F3H), isoflavone synthesis (IFS), 2-hydroxyisoflavanone dehydratase (HID) and then modified by glucosyltransferase (GT) (Wang et al. 2017). Although some putative structural genes involved in pueraria biosynthesis has been identified through transcriptome sequencing and biochemical analysis (Wang et al. 2015). However, it remains largely unknown whether the *bHLH* gene family members are involved in the regulation of puerarin biosynthesis in *P. thomsonii.*

Currently, we finished the genome sequencing of *P. thomsonii*, which provides an excellent opportunity for further genome-wide analysis of *P. thomsonii* genes (Shang et al. 2022). In this study, we performed a genome-wide analysis of *bHLH* genes in *P. thomsonii* and 219 *PlbHLH* genes were identified. Further, a comprehensive analysis including phylogenetic relationship, chromosomal location, gene duplication, collinearity, gene structure, conserved motifs was performed. Based on different tissues expression pattern of *PlbHLH* genes as well as up-regulated expression in response to methyl jasmonate (MeJA) treatment, several *PlbHLH* genes were considered as candidate regulators involving in puerarin biosynthesis.

Materials And Methods

Plant Materials and Treatment

P. thomsonii were grown in MS medium in an artificial climate at 26±2℃, with a photoperiod of 16 h light and 8 h dark, at Guangxi Agricultural Academy Sciences (Nanning, China). For MeJA treatment, 4-week-old seedlings were cultured in nutrient solutions with 100 uM MeJA treatment. Roots, stems and leaves for RNA extraction were collected at 0 h, 12 h and 24 h after the treatment, respectively, immediately frozen in liquid nitrogen and stored at −80°C for RNA-sequencing. Three seedlings were used as biology replicate.

*P. thomsonii*bHLH Sequence Retrieval and Analysis

We used blastp to search for candidate bHLH genes in *P. thomsonii* genome with an e-value cutoff of 1e-5 using 167 *Arabidopsis thaliana* (*A. thaliana*) bHLH proteins as query sequences (Tian et al. 2019). In addition, the bHLH encoding genes in *P. thomsonii* with Pfam00010 were identified using Hidden Markov Model (HMM) (El-Gebali et al. 2019). The remaining sequences were checked for the conserved bHLH domain using Conserved Domain Database (CDD) (Lu et al. 2020) and SMART (Letunic and Bork 2007). The theoretical isoelectric point (PI) and molecular weight (MW) were predicted using ExPASy (Gasteiger 2003). ProtComp v 9.0 was used to predicted protein subcellular location.

Phylogenetic Analysis

Multiple sequence alignments using MUSCLE with default parameters were applied to confirm the conserved domains of predicted PlbHLH proteins. A phylogenetic tree was created using the NJ method of MEGA X version 10.1.7 with 1000 bootstrap replicates based on the sequence alignments (Kumar et al. 2018), and was visualized by the Interactive Tree of Life (iTOL) (Letunic and Bork 2007). The phylogenetic tree was visualized in MEGA5.0.

Chromosomal Location and Gene Synteny Analysis

PlbHLH genes mapping was performed based on physical location information from the database of *P. thomsonii* genome using TBtools (Chen et al. 2018). The gene duplication events were conducted using the Multiple Collinearity Scan toolkit (MCScanX), with the default parameters (Wang et al. 2012). The orthologous *bHLH* genes between *P. thomsonii* and *A. thaliana* as well as those between *P. thomsonii* and *Glycine max* were identified using OrthoVenn2 (Xu et al. 2019). Multiple Collinearity Scan toolkit (MCScanX) was used to analyze the gene duplication events with default parameters (Wang et al. 2012). Non-synonymous (ka) and synonymous (ks) substitutions of each duplicated *bHLH* gene were calculated using the Tbtools software (Chen et al. 2020).

Conserved Motifs and Genes Structure Characterization

MEME version 5.1.1 program was used to detect the conserved motifs of PlbHLH proteins using the parameters: Zero or One Occurrence per Sequence (zoops), maximum number of motifs, 15; and the optimum width of each motif, 6–50 residues (Bailey et al. 2009). Gene Structure Display Server (http://gsds.cbi.pku.edu.cn) was used to analyze the exon–intron structures within the coding sequences and the genomic sequences of each predicted *PlbHLH*.

Transcriptome Data Analysis

Expression pattern of each *PlbHLH* gene was analyzed from our previously reported RNA-seq data (Shang et al. 2022). The data included three tissues, including root, stem and leave. The expression level of each predicted transcript in each RNA-seq library was calculated as the fragments per kilobase of transcript per million fragments mapped (FPKM) using RSEM (Li and Dewey 2011). To assess the three biological replicates, the ratio of root/stem and root/leaf-transformed FPKM values were performed. The expression level of *PlbHLH* genes in roots with more than 2-fold change stem and leave were deemed as

the the genes which were expressed the highest in root. Each *PlbHLH* gene was analyzed and normalized and displayed in heat map (Chen et al. 2018).

Quantitative Real-time RT-PCR Validation

Changes in the expression of *PlbHLH* genes in response to MeJA were detected by quantitative real-time polymerase chain reaction (qRT-PCR) analysis using SYBR® Premix Ex TaqTM (TaKaRa) chemistry on a Stratagene Mx3000P (Stratagene, CA, USA) instrument. For each biological replicate, quantifications were conducted in triplicate. The primers were designed using Primer 3 online (http://bioinfo.ut.ee/primer3-0.4.0/primer3/). Primer information is presented in Supplementary Table S1. The *Plactin* gene was used as the internal. The relative expression levels of the target genes were assessed based on 2− *C*t method (Livak and Schmittgen 2001).

Statistical Analyses

All the expression data were analyzed using Excel and SPSS Statistical Software (version 17.0; Chicago, IL, United States), by ANOVA, followed by Tukey's signi cant difference test at $P<0.05$.

Results

Identification and Classification of PtbHLH Proteins in *P. thomsonii* genome

A total of 219 putative *PlbHLHs* were defined from *P. thomsonii* based on the methods in "Materials and methods". Sequence analysis indicated that the PlbHLH proteins varied from 91 amino acids to 799 amino acids, with molecular weight ranging from 10356.7 to 89787.63 kDa, the theoretical isoelectric point values ranging of 4.29-9.61. A total of 197 PlbHLH proteins localized in nucleus, while 14 PlbHLH proteins were located in chloroplast, one in endoplasmic reticulum (PlbHLH199), four in mitochondrion, two in plasma membrane (PlbHLH57/169), one in peroxisome (PlbHLH196) (Supplementary Table S2).

Phylogenetic Analysis

To explore the evolutionary relationship and the classification of *PlbHLHs*, a neighbor-joining phylogenetic tree was constructed with conserved sequences of 219 *PlbHLHs* and 167 *AtbHLHs*. Based on this phylogenetic analysis and the previous reported classification of the *AtbHLHs*, we classified the *PlbHLH* family with 219 members into 13 subfamilies, designated as I to XIII (**Fig. 1)**. This result corroborated the previously proposed classification of the bHLH family members (Pires and Dolan 2010). As is shown in **Fig. 1**, group I was the largest subfamily with 36 *PlbHLH* members, followed by IV with 34 members. In contrast, Group II, the smallest, has only two members (*PlbHLH76/93*), followed by group VI with four *PlbHLHs* (*PlbHLH22/23/49/50*).

Chromosomal Distribution and Synteny Analyses of *PlbHLH* gene Family

Based on their chromosomal locations, PlbHLH proteins were assigned from *PlbHLH1* to *PlbHLH219* (Supplementary Table S2). All the *PlbHLH* genes were distributed on the 11 chromosomes (chr) unevenly. The number of *PlbHLH* genes on each chr varied from 12 to 44. Chr11 had only 12 *PtbHLH* genes, both of chr6 and chr9 had 15 *PlbHLH* members. In contrast, chr1 contain the largest number, 44, of *PlbHLH* genes followed by 23 on chr4. Most members are closely tandem distributed on chromosomes. In addition, *PlbHLH219* was positioned near the telomeric regions on chr11 (**Fig. 2**). According to a previous study, if two or more genes are present within 200 kb, the elements are considered a tandem repeat event (Holub 2001). It is shown that 35 *PlbHLHs* were identified to be tandem duplicated genes which were distributed on each chr. The duplicated genes were *PlbHLH4/5, 12/13, 18/19, 25/26, 29/30, 33/34/35, 36/37/38, 39/40, 47/48, 54/55, 59/60, 62/63, 67/68, 70/71, 84/85/86, 87/88, 90/91/92, 102/103, 105/106, 114/115/116, 121/122, 125/126, 136/137, 138/139, 144/145, 161/162, 163/164/165/166, 173/174, 175/176, 180/181, 182/183, 197/198, 202/203, 213/214, 216/217* (**Fig. 2;** Supplementary Table S2). Basic local alignment search tool for proteins (BLASTP) and MCScanX were performed to construct the collinearity of the *PlbHLH* gene family and identify the possible relationship and potential duplication events between them. Intrachromosomal duplications of the *PlbHLH* gene family were observed in the *P. thomsonii* genome (Supplementary Fig. S1). In detail, 30 pairs of *PlbHLH* genes duplicated tandemly on the 11 chromosomes (Supplementary Table S3). Together, the results suggested that both of tandem and segmental duplications are key factors driving *bHLH* gene family expansion in *P. thomsonii* (Guo et al. 2013).

To further illustrate the potential evolutionary patterns of the *PlbHLH* gene family, a comparative orthologous analysis was performed between *P. thomsonii* and the other two plant species, *A. thaliana* and soybean (*Glycine max*), which belong to the Brassicaceae and leguminosae families, respectively (**Fig. 3**). The orthologous gene pairs between *P. thomsonii* and *A. thaliana* and *P. thomsonii* and *Glycine max* were 153 and 584, respectively (Supplementary Tables S4 and S5). These results revealed that the identified orthologous events of *PlbHLH-GmbHLH* were greatly more than those of *PlbHLH-AtbHLH* based on the close evolutionary relationship between *P. thomsonii* and *Glycine max.* An extensive level of synteny conservation and increased number of orthologous events of *PlbHLH-GmbHLH* indicated that *PlbHLH* genes in *P. thomsonii* shared a similar structure and function with *GmbHLH* genes in *Glycine max*.

To further investigate the driving force behind the duplication of *PlbHLH* gene pairs in *P. thomsonii*, Ka/Ks (non-synonymous/synonymous substitution ratio) calculation of the duplicated *bHLH* gene pairs was performed to determine whether a selective pressure acted on the *bHLH* genes. Interestingly, all the Ka/Ks values of orthologous *PlbHLH* gene pairs were less than 1, indicating that these genes were subjected to purifying selection with limited functional divergence during evolution after duplication events (Supplementary Tables S4 and S5).

Conserved Motifs, Domain and Gene structure Characterization

To support the phylogenetic analysis, motifs of *PlbHLH* family members was investigated. The results showed that 20 motifs were found in the *PlbHLH* gene family (**Fig. 4a**). In details, motif 1 was presented in all the *PlbHLH* members. All the *PlbHLH* members contained motif 2 except *PlbHLH192*. We found that motif 11 was only presented in *PlbHLH51/192/133*, and *PlbHLH161-166*. Motif 2 and Motif 9 were repeated twice in *PlbHLH136/137/190/191/215*, and *PlbHLH19/20/121/122*, respectively. Motif 19 was only presented in *PlbHLH54/55/156/185*, and repeated twice in *PlbHLH54/55*. The members that phylogenetically clustered in the same subfamily often showed a similar motif pattern. For example, the closely clustered pair *PlbHLH24/45/47/48/136/137/1901/191/167* in subfamily I had identical motifs 1-3, 5, 6 and 14. All the 12 members of group IV contained motifs 2, 13 and 15 (Supplementary Table S6). Noticeable, all the *PlbHLHs* contained the conserved HLH domain except *PlbHLH1/2/18/29/35/36/37/39/52/53* (**Fig.4b)**.

Exon-intron structural diversity, an important part in the evolution of gene families, provides additional evidence supporting phylogenetic groupings (Shiu and Bleecker 2003). In order to better understand the structural features of *PlbHLHs*, the intron structure was detected based on their evolutionary relationships. As shown in **Fig. 4c**, the number of introns in *PlbHLH* members varied from 0 (*PlbHLH8/9/16/17/46/61/104/113/119/120/122/135/177/205*) to 16 (*PlbHLH137*). Twenty *PlbHLH* genes contain one intron. *PlbHLHs* genes within the same subfamily appeared to share similar intron-exon gene structures. A total of 185 *PlbHLH* genes contain more than two introns. Nevertheless, the number of exons and introns varied greatly among other subgroups **(Fig. 4c)**.

Tissue Expression Pattern of *PlbHLH* Genes

In order to investigate the possible function of *PlbHLH* genes, hierarchical cluster analysis was performed to assess the tissue-specific transcript abundance of each *PlbHLH* gene, represented by FPKM, using the expression patterns of 219 *PlbHLH* genes corresponding to root, stem and leaf of *P. thomsonii* (Supplementary Fig. S2). The results indicated that five *PlbHLH* genes (*PlbHLH13/15/78/129/155*) showed no expression in the three tissues, in particular, 20 *PlbHLHs* (*PlbHLH3/16/17/18/40/46/48/59/77/86/93/125/131/138/139/162/163/174/177/195*) showed high expression level in the three tissues. High expression in root and stem, but low expression in leaf tissue was prevalent among in 13 *PlbHLHs* (*PlbHLH25/31/34/47/53/100/115/121/126/127/165/195/196*). Four *PlbHLHs* (*PlbHLH60/63/67/145*) showed high expression level in stem (Supplementary Table S7). In order to identify the highest expressed *PlbHLH* genes which were deemed as the root-specific expressed, the transcriptional abundance of *PlbHLH* genes in root with more than twice than stem and leaf were selected, consequently, 41 *PlbHLH* genes were screened (**Fig. 5;** Supplementary Table S8).

Identification of *PlbHLH* Genes Related to Pueraria Biosynthesis

In *P. thomsonii*, pueraria was produced in root, and pueraria content were much higher in the root than in stem and leaf (Chen and Zhang 2001; Han et al. 2015). MeJA is a signaling molecule associated with biosynthesis of flavonoids (Shelton et al. 2012; Zhou and Memelink 2016; Du et al. 2021), for instance, the pueraria yield could be increased to 25 times by different concentration MeJA treatment (Goyal and

Ramawat 2008). In view of this, MeJA treatment analysis was performed to narrow down the candidate structural genes. The RNA-seq result of Shang et al. (2022) and results of the qRT-PCR detection indicated *PlCHS1* (Pmon008G00376), *PlCHR3* (Pmon003G00389), *PlIFS1* (Pmon003G02838), *PlIFS2* (Pmon003G02830) and *PlUGT155* (Pmon006G01850) were up-regulated expressed after treatment with MeJA within 24 h (**Fig. 6**).

The regulation of isoflavone is at the transcriptional level (Ferreyra and Rius 2012; Li 2014). The positive regulators often exhibit a similar expression pattern to structural genes in biosynthetic pathway (Zhang et al. 2008; Xiang et al. 2019; Zhu et al. 2019). Using qRT-PCR detection, we found that the expression of 19 *PlbHLH* genes were not influenced by MeJA treatment within 24 h (Supplementary Fig. S3). Compared with untreated root, nine *PlbHLH* genes were down-regulated in response to MeJA treatment at 12 h and 24 h (Supplementary Fig. S3). In contrast, seven *PlbHLH* genes (*PlbHLH6/56/73/79/94/149/196*), were up-regulated in response to MeJA within 24 h of treatment compared with the untreated root (**Fig. 7**). Importantly, the expression of *PlCHS1*, *PlCHR3, PlIFS1*, *PlIFS2,* and *PlUGT155* was positively correlated with the expression of the seven *PlbHLH* genes. Together, the seven bHLH genes were the candidate genes which acts regulator of puerarin biosynthesis in *P. thomsonii*.

Discussion

Puerarin, an isoflavone compound, only comes from *P. thomsonii* which is a traditional Chinese medicinal herb. It often serves as a natural source of puerarin in medicinal products for human health. The genome sequencing of *P. thomsonii* was completed recently (Shang et al. 2022), providing the opportunity to perform a genome-wide analysis of *bHLH* gene family in *P. thomsonii*. Numerous studies have shown that bHLH-type transcription factors are involved in diverse biological processes in plant growth, development, and stress responses and regulating specialized secondary metabolite biosynthesis (Guo et al. 2021; Qian et al. 2021). However, a systematic characterization of *bHLH* genes in *P. thomsonii* is lacked.

The bHLH transcription factors comprised a large family in plants. Based on sequence homology and phylogenetic relationships, 128 *bHLH* genes in pummelo (*Citrus grandis*), 133 *bHLH* genes in alfalfa (*Medicago sativa*), 167 bHLH proteins in *A. thaliana*, 197 bHLH proteins in pear (*Pyrus bretschneideri*), 230 *bHLH* genes in Chinese cabbage (*Brassics rapa ssp. pekinensis*), potato (*Solanum tuberosum*) with 259 bHLH proteins, 319 bHLH proteins in soybean, and 602 bHLH proteins in *Brassica napus* (Toledo-Ortiz and Huq 2003; Li et al. 2006; Hudson and Hudson 2015; Mao et al. 2017; Wang et al. 2018b; Ke et al. 2020; Dong et al. 2021; Ndayambaza et al. 2021). In our study, there were 219 *PlbHLH* genes divided into 13 subfamilies. These results indicated that *bHLH* gene family has undergone substantial expansion, with potential subfunctionalization and neofunctionalization, as well as changes in gene expression patterns and protein–protein interactions.

The gain and loss of intron is highly related with rate of the gene molecular evolution (Pires and Dolan 2010). Intron loss might occur through recombination between the genomic copy of the gene and a reverse transcript of a spliced mRNA copy of the gene (Roy and Gibert 2006). It was demonstrated that genes with almost no intron often underwent positive selection during evolution, for instance, SAUR family (Chen and Hao 2014). In contrast, for *bHLH* gene family, a large number of *bHLH* genes with more than two introns were widely distributed in *bHLH* gene family of lotus (*Nelumbo nucifers*), cucumber (*Cucumis sativus*), pummelo and pear, and these *bHLH* genes underwent negative selection in the corresponding crop species (Mao et al. 2019; Li et al. 2020; Zhang et al. 2020; Dong et al. 2021). In the 219 *PlbHLH* genes identified in this study, 185 *PlbHLH* genes contain more than two introns, supporting the *PlbHLH* gene family undergo purifying selection during evolution process, which is consistent with the result of Ka/Ks analysis. Additionally, the distribution of introns in some gene families were not exactly the same because introns were constantly gained and lost (Chen and Cao 2014). Whether the *PlbHLH* gene family has a similar effect needs to study further in *P. thomsonii*.

Genes with similar expression patterns were consistent with their diverse functions in regulation of plant growth and development, and secondary metabolism. Transcriptional activators are often co-expressed with its downstream structural genes, therefore, co-expressed analysis is an effective way to identify regulators controlling secondary metabolism. For instance, the expression pattern analysis indicated that the expression of *GmMYB29* was closely associated with that of *IFS2*, the expression of the two genes was consistent with the isoflavone content in different tissues. The *GmMYB29*-overexpressing transgenic lines leaded to higher isoflavones content in soybean hair roots via activating the transcription level of IFS2 (Chu et al. 2017). *CsMYB31* was a transcriptional activator and positively regulates genus-specialized metabolite production in pepper, and the *CsMYB31* expression pattern was similar with that of the 15 capsaicinoids-biosynthetic genes (Zhu et al. 2019). *AabHLH112* could induce artemisnin biosynthesis via enhanced the promoter activity of artemisnin biosynthesis genes, *ADS* and *CYP71AV1*, and *AabHLH112* was tightly associated with artemisinin biosynthesis genes (Xiang et al. 2019). In this study, the seven *bHLH* genes (*PlbHLH6/56/73/79/94/149/196*) and five key structure genes (*PlCHS1*, *PlCHR3, PlIFS1*, *PlIFS2*, and *PlUGT155*) were exhibited high expression in root. Furthermore, the five puerarin biosynthesis *PlbHLH* genes were responsive to MeJA, which could increase the content of puerarin. All these results implied that *PlbHLH6/56/73/79/94/149/196* played an important role in the regulation of the puerarin biosynthetic pathway, and their function need further verification in *P. thomsonii*.

Conclusion

In summary, 219 *bHLH* genes were identified from *P. thomsonii*. Chromosomal distribution, conserved domains, motifs distribution, and exon-intron characteristics of *PlbHLH* gene family were investigated. The Ka/Ks analysis indicated *PlbHLH* gene family undergo purifying selection during evolution. The expression patterns of 219 *PlbHLH* genes in root, stem and leaf were analyzed. RNA-seq data, and qRT-PCR results revealed that seven *PlbHLH* genes had a similar expression patterns with five structural genes of the puerarin biosynthesis in response to MeJA treatment. The collected data from this study

provided a foundation for advanced studies to evaluate the mechanisms of regulating puerarin biosynthesis for *bHLH* genes in *P. thomsonii*.

Declarations

Author Contributions

HY, and LX designed the research. WZ prepared the materials. ZW and XS prepared all the Figures. SC, LL and PS prepared the supplementary Figures and Tables. LX performed the experiments and analyzed the data. LX and DH wrote the main text. HY revised the paper. All authors approved the final manuscript.

Funding

This study was supported by Guangxi Natural Science Foundation Project (2021GXNSFBA220026), Guangxi Natural Science Foundation Project (2020GXNSFAA297187), National Natural Science Foundation of China (31960420), and Special Project for Basic Scientific Research of Guangxi Academy of Agricultural Sciences (Guinongke 2021YT057).

Declaration of competing interest

The authors confirm no conflict of interest.

Data Availability

The transcriptome sequencing data can be found in DDBJ/ENA/GenBank and SRA under the umbrella of BioProject Accession PRJNA723378.

References

1. Atchley W R, Terhalle W, Dress A (1999) Positional dependence, cliques, and predictive motifs in the bHLH protein domain. J Mol Evol 48: 501–516
2. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L et al (2009) MEME Suite: tools for motif discovery and searching. Nucleic Acids Res 37: W202–W208
3. Carretero-Paulet L, Galstyan A, Roig-Villanova I, Martinez-Garcia JF, Bilbao-Castro JR, Robertson DL (2010) Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in Arabidopsis, poplar, rice, moss, and algae. Plant Physiol 153: 1398–1412
4. Chen C, Xia R, Chen H, He Y (2018) TBtools, a toolkit for biologists integrating various biological data handling tools with a user-friendly interface. BioRxiv: 289660
5. Chen CJ, Chen H, Zhang Y, Thomas HR, Frank MH, He YH et al (2020) TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant 13: 1194–1202
6. Chen G, Zhang JX, Ye JN (2001) Determination of puerarin, daidzein and rutin in *Pueraria lobata* (Wild.) Ohwi by capillary electrophoresis with electrochemical detection. J Chromatogr A 923: 255–

262

7. Chen YZ, Cao J (2014) Comparative genomic analysis of the Sm gene family in rice and maize. Gene 539: 238–249
8. Chen YZ, Hao X, Cao J (2014) Small auxin upregulated RNA (SAUR) gene family in maize: Identification, evolution, and its phylogenetic comparison with *Arabidopsis*, rice, and sorghum. J Integr Plant Biol 56: 133–150
9. Chu S, Wang J, Zhu Y, Liu SL, Zhou XQ, Zhang HR et al (2017) An R2R3-type MYB transcription factor, GmMYB29, regulates isoflavone biosynthesis in soybean. PLOS Genet 13: e1006770
10. Dong HZ, Chen QM, Dai YQ, Hu WJ, Zhang SL, Huang XS (2021) Genome-wide identification of PbrbHLH family genes, and expression analysis in response to drought and cold stresses in pear (*Pyrus bretschneideri*). BMC Biol 21: 86
11. Du TT, Fan YX, Cao HY, Song ZH, Dong BY, Liu TY et al (2021) Transcriptome analysis revealed key genes involved in flavonoid metabolism in response to jasmonic acid in pigeon pea (*Cajanus cajan* (L.) Millsp.). Plant Physiol Biotech 168: 40–422
12. Duek PD, Fankhauser C (2005) bHLH class transcription factors take centre stage in phytochrome signalling. Trends Plant Sci 10: 51–54
13. Egan AN (2020) Economic and ethnobotanical uses of tubers in the genus *Pueraria* DC. Legume 19: 24
14. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC et al (2019) The Pfam protein families database in 2019. Nucleic Acids Res 47(Database issue): D427–D432
15. Ferreyra MF, Rius SP, Casati P (2012) Flavonoids: biosynthesis, biological functions, and biotechnological applications. Front Plant Sci 3: 222
16. Gasteiger E (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 31: 3784–3788
17. Goyal S, Ramawat K (2008) Increased isoflavonoids accumulation in cell suspension cultures of Pueraria tuberosa by elicitors. Indian J Biotechnol 7: 378–382
18. Guo JR, Sun BX, He HR, Zhang YF, Tian HY, Wang BS (2021) Current understanding of bHLH transcription factors in plant abiotic stress tolerance. Int J Mol Sci 22: 4921
19. Guo RR, Xu XZ, Carole B, Li XQ, Gao M, Zheng Y et al (2013) Genome-wide identification, evolutionary and expression analysis of the aspartic protease gene superfamily in grape. BMC Genomics 14: 1–18
20. Han RC, Takahsshi H, Nakamura M, Yoshimoto N, Suzuki H, Shibata D et al (2015) Transcriptomic landscape of *Pueraria lobata* demonstrates potential for phytochemical study. Front Plant Sci 6: 426
21. Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC (2003) The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. Mol Biol Evol 20: 735–747

22. Hien TT, Kim HG, Han EH, Wang KW, Jeong HG (2010) Molecular mechanism of suppression of MDR1 by puerarin from *Pueraria lobata* via NF-kappaB pathway and cAMP-responsive element transcriptional activity-dependent up-regulation of AMP-activated protein kinase in breast cancer MCF-7/adr cells. Mol Nutr Food Res 54: 918–928
23. Holub EB (2001) The arms race is ancient history in *Arabidopsis*, the wildflower. Nat Rev Genet 2: 516–527
24. Hudson KA, Hudson ME (2015) A classification of basic helix-loop-helix transcription factors of soybean. Int J Genomics 2015: 603182
25. Jin C, Huang XS, Li KQ, Yin H, Li LT, Yao ZH et al (2016) Overexpression of a *bHLH1* transcription factor of *Pyrus ussuriensis* confers enhanced cold tolerance and increases expression of stress-responsive genes. Front Plant Sci 7: 441
26. Ke YZ, Wu YW, Zhou HJ, Chen P, Wang MM, Liu MM et al (2020) Genome-wide survey of the bHLH super gene family in *Brassica napus*. BMC Plant Biol 20: 115
27. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35: 1547–1549
28. Lee S, Lee S, Yang KY, Kim YM, Park SY, Kim SY et al (2006) Overexpression of *PRE1* and its homologous genes activates Gibberellin-dependent responses in *Arabidopsis thaliana*. Plant Cell Physiol 47: 591–600
29. Letunic I, Bork P (2018) 20 years of the SMART protein domain annotation resource. Nucleic Acids Res 46: D493–D496
30. Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf 12: 323
31. Li JL, Wang T, Han J, Ren ZH (2020) Genome-wide identification and characterization of cucumber bHLH family genes and the functional characterization of CsbHLH041 in NaCl and ABA tolerance in *Arabidopsis* and *cucumber*. BMC Plant Biol 20: 272
32. Li S (2014) Transcriptional control of flavonoid biosynthesis. Plant Signal Behav 9: e27522
33. Li X, Duan X, Jiang H, Sun Y, Tang Y, Yuan Z et al (2006) Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and *Arabidopsis*. Plant Physiol 141: 1167–1184
34. Liu B, Wu ZY, Li YP, Ou CW, Huang ZJ, Zhang JW et al (2015) Puerarin prevents cardiac hypertrophy induced by pressure overload through activation of autophagy. Biochem Bioph Res Co 464: 908–915
35. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2 - \Delta\Delta Ct$ method. Methods 25 402–408
36. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR et al (2020) CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res 48: D265–D268
37. Ludwig SR, Habera LF, Dellaporta SL, Wessler SR (1989) Lc, a member of the maize R gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the myc-homology region. Proc Natl Acad Sci USA 86: 7092–7096

38. Mao TY, Liu YY, Zhu HH, Zhang J, Yang JX, Fu Q et al (2019) Genome-wide analyses of the bHLH gene family reveals structural and functional characteristics in the aquatic plant *Nelumbo nucifera*. Peer J 7: e7153
39. Matus JT, Poupin MJ, Cañón P, Bordeu, Alcalde JA, Arce-Johnson P (2010) Isolation of WDR and bHLH genes related to flavonoid synthesis in grapevine (*Vitisvinifera* L.). Plant Mol Biol 72: 607–620
40. Mao K, Dong Q, Li C, Liu C, Ma F (2017) Genome wide identification and characterization of apple bHLH transcription factors and expression analysis in response to drought and salt stress. Front Plant Sci 8: 480
41. Meezan E, Meezan EM, Jones K, Moore R, Barnes S, Prasain JK (2005) Contrasting effects of puerarin and daidzin on glucose homeostasis in mice. J Agric Food Chem 53: 8760–8767
42. Ndayambaza B, Jin XY, Min XY, Lin XS, Yin XF, Liu WX (2021) Genome-wide identification and expression analysis of the barrel medic (*Medicago truncatula*) and alfalfa (*Medicago sativa* L.) basic helix-loop-helix transcription factor family under salt and drought stresses. J Plant Growth Regul 40: 2058–2078
43. Pires N, Dolan L (2010) Origin and diversification of basic-helix-loop-helix proteins in plants. Mol Biol Evol 27: 862–874
44. Qian YC, Zhang TY, Yu Y, Gou LP, Yang JT, Xu J et al (2021) Regulatory mechanisms of bHLH transcription factors in plant adaptive responses to various abiotic stresses. Front Plant Sci 12: 677611
45. Roig-Villanova I, Bou-Torrent J, Galstyan A, Carretero-Paulet L, Portolés S, Rodríguez-Concepción M et al (2007) Interaction of shade avoidance and auxin responses: a role for two novel atypical bHLH proteins. EMBO J 26: 4756–4767
46. Roy SW, Gilbert W (2006) The evolution of spliceosomal introns: patterns, puzzles and progress. Nat Rev Genetic 7: 211–221
47. Shang XH, Yi XX, Xiao L, Zhang YS, Huang D, Xia ZB et al (2022) Chromosomal-level genome and multi-omics dataset of *Pueraria lobata var. thomsonii* provide new insights into legume family and the isoflavone and puerarin biosynthesis pathways. Hortic Res 9: uhab035
48. Shelton D, Stranne M, Mikkelsen L, Pakseresht N, Welham T, Hiraka H et al (2012) Transcription factors of Lotus: regulation of isoflavonoid biosynthesis requires coordinated changes in transcription factor activity. Plant Physiol 159: 531–547
49. Shiu S H, Bleecker AB (2003) Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in Arabidopsis. Plant Physiol 132: 530–543
50. Song XM, Huang ZN, Duan WK, Ren J, Liu TK, Li Y et al (2014) Genome-wide analysis of the bHLH transcription factor family in Chinese cabbage (*Brassica rapa ssp. pekinensis*). Mol Genet Genom 289: 77–91
51. Tian F, Yang DC, Meng YQ, Jin J, Gao G (2019) PlantRegMap: charting functional regulatory maps in plants. Nucleic Acids Res 48: D1104–D1113

52. Toledo-Ortiz G, Huq E, Quail PH (2003) The Arabidopsis basic/helix-loop-helix transcription factor family. Plant C 15: 1749–1770
53. Wang PF, Su L, Gao HH, Jiang XL, Wu XY, Li Y (2018a) Genome-wide characterization of bHLH genes in grape and analysis of their potential relevance to abiotic stress tolerance and secondary metabolite biosynthesis. Front Plant Sci 9: 64
54. Wang RQ, Zhao P, Kong NN, Lu RZ, Pei Y, Huang CX et al (2018b) Genome-wide identification and characterization of the potato bHLH transcription factor family. Genes 9: 54
55. Wang SG, Zhang SM, Wang SP, Gao P (2020) A comprehensive review on *Pueraria*: insights on its chemistry and medicinal value. Biomed Pharmacother 131: 110734
56. Wang X, Li CF, Zhou C, Li J, Zhang YS (2017) Molecular characterization of the C-glucosylation for puerarin biosynthesis in *Pueraria lobata*. Plant J 90: 535–546
57. Wang X, Li S, Li J, Li CF, Zhang YS (2015) De novo transcriptome sequencing in *Pueraria lobato* to identify putative genes involved in isoflavones biosynthesis. Plant Cell Rep 34: 733–743
58. Wang YP, Tang HB, DeBarry JD, Tan X, Li JP, Wang XY et al (2012) MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res 40: e49
59. Xiang L, Jian DQ, Zhang FY, Yang CX, Bai G, Lan XZ et al (2019) The cold-induced bHLH transcription factor AabHLH112 promotes artemisinin biosynthesis indirectly via ERF1 in *Artemisia annua*. J Exp Bot 70: 4835–4848
60. Xu L, Dong Z, Fang L, Luo Y, Wei Z, Guo H et al (2019) OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res 47: W52–W58
61. Zhang GY, Chen M, Chen XP, Xu ZS, Guan S, Li LC et al (2008) Phylogeny, gene structures, and expression patterns of the ERF gene family in soybean (*Glycine max* L.). J Exp Bot 59: 4095–4107
62. Zhang XY, Qiu JY, Hui QL, Xu YY, He ZY, Peng LZ et al (2020) Systematic analysis of the basic/helix-loop- helix (bHLH) transcription factor family in pummelo (*Citrus grandis*) and identification of the key members involved in the response to iron deficiency. BMC Genomics 21: 233
63. Zhao Y, Zhang YY, Liu H, Zhang XS, Ni R, Wang PY (2019) Functional characterization of a liverworts bHLH transcription factor involved in the regulation of bisbibenzyls and flavonoids biosynthesis. BMC Plant Biol 19: 497
64. Zhao Z, Guo P, Brand E (2018) A concise classification of bencao (*materia medica*). Chin Med 13: 18
65. Zhou M, Memelink J (2016) Jasmonate-responsive transcription factors regulating plant secondary metabolism. Biotechnol Adv 34: 441–449
66. Zhou YX, Zhang H, Peng C (2014) Puerarin: a review of pharmacological effects. Phytother Res 28: 961–75
67. Zhu ZS, Sun BS, Cai W, Zhou X, Mao YH, Chen CJ et al (2019) Natural variations in the MYB transcription factor MYB31 determine the evolution of extremely pungent peppers. New Phytol 223: 922–938

Figures

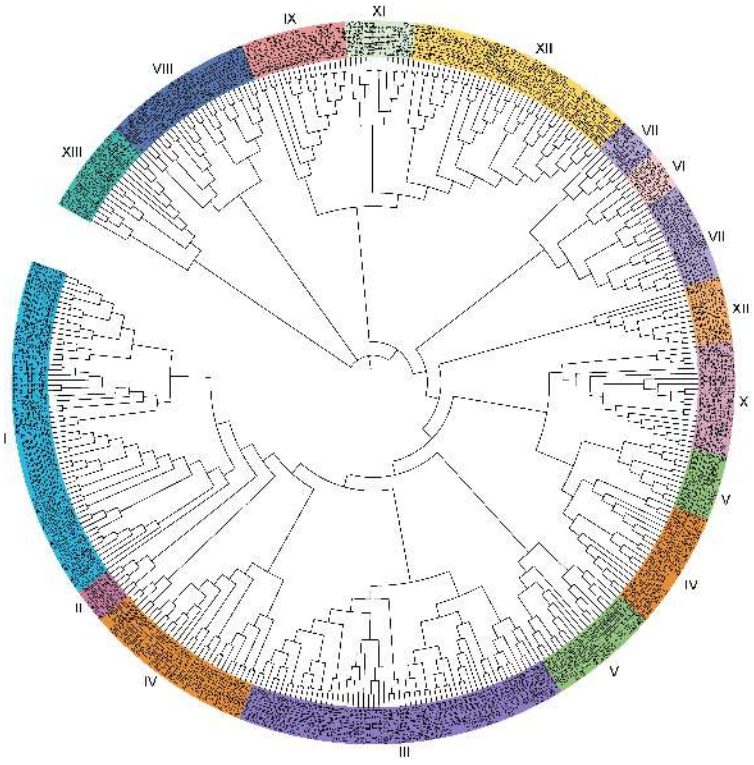


Figure 1

Phylogenetic analysis of bHLH proteins from *P. thomsonii* and *A. thaliana*. A total of 219 bHLH proteins from *P. thomsonii* and 167 bHLH proteins from *A. thaliana* were used to construct the neighbor-joining (NJ) tree.

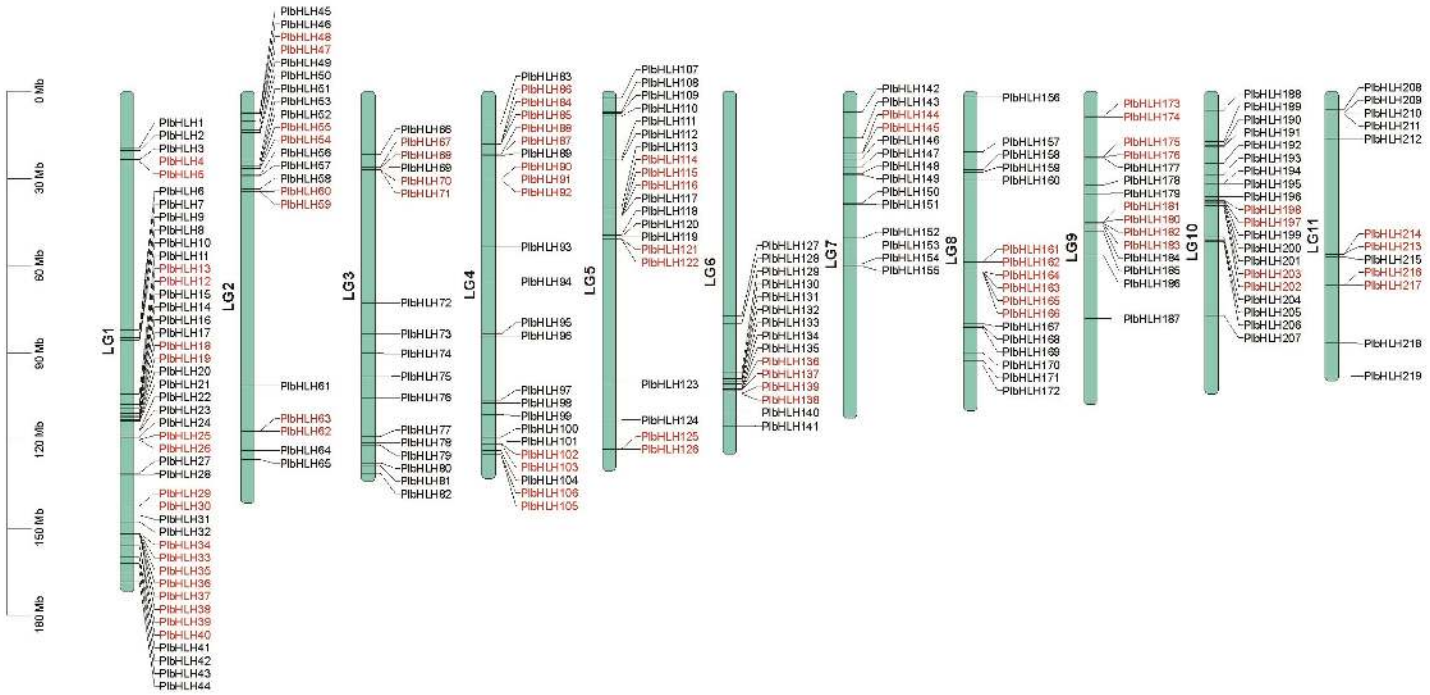


Figure 2

The distributions of 219 *PlbHLH* genes on *P. thomsonii* chromosomes. Centromeres are indicated by black ovals. The scale is in megabases (Mb). The *PlbHLH* genes highlighted in red on chromosomes indicate tandem duplications.

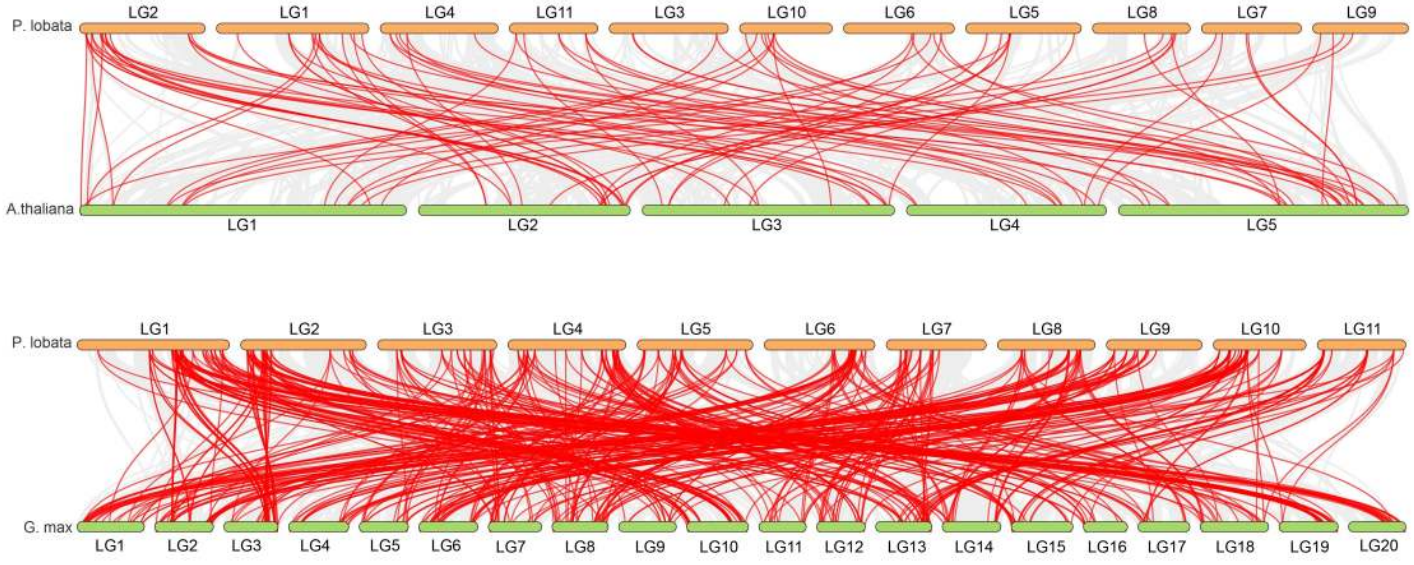


Figure 3

Collinearity analysis of *PlbHLHs* between *P. thomsonii* and *Glycine max* and *A. thaliana*. Chr represents chromosome-scale scaffolds. All gene pairs and syntenic *PlbHLH* gene pairs were presented by gray lines and red lines, respectively.

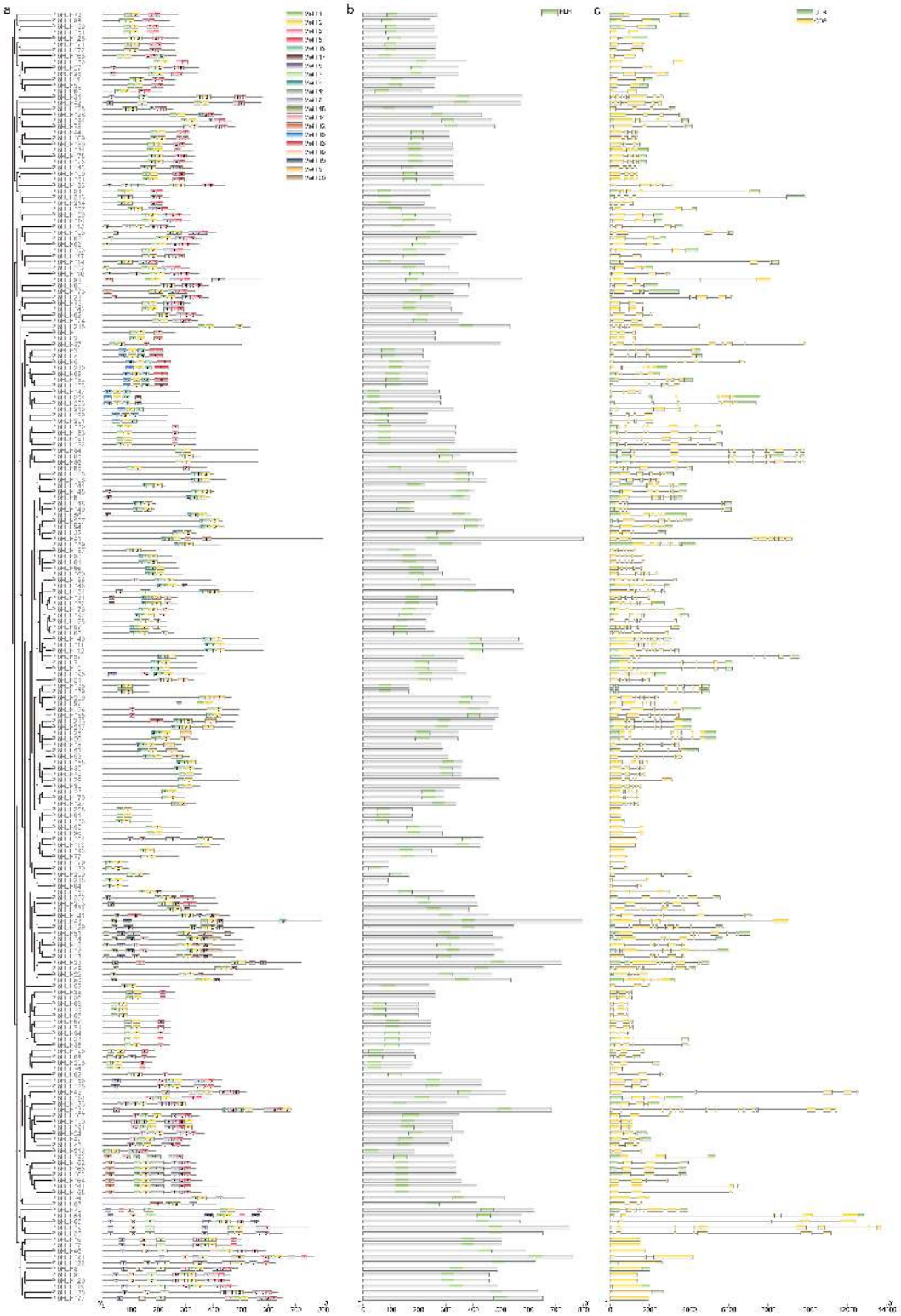


Figure 4

Schematic of the conserved motifs, domain and gene structure depending on the phylogenetic relationships in *PlbHLH* genes. **a** The unrooted phylogenetic tree was built using neighbor-joining (NJ)

method. Schematic representation of conserved motifs elucidated by MEME. The conserved motifs are represented by the different colored boxes. The scale bar of each *PlbHLH* gene is shown at the bottom. **b** Exon/intron structure analysis of *PlbHLH* genes. Gray lines, yellow boxes, and green boxes indicate introns, exons, and untranslated regions (UTRs), respectively. **c** The conserved bHLH DNA-binding domains were examined using Pfam and SMART. Green boxes indicate conserved bHLH DNA-binding domains.

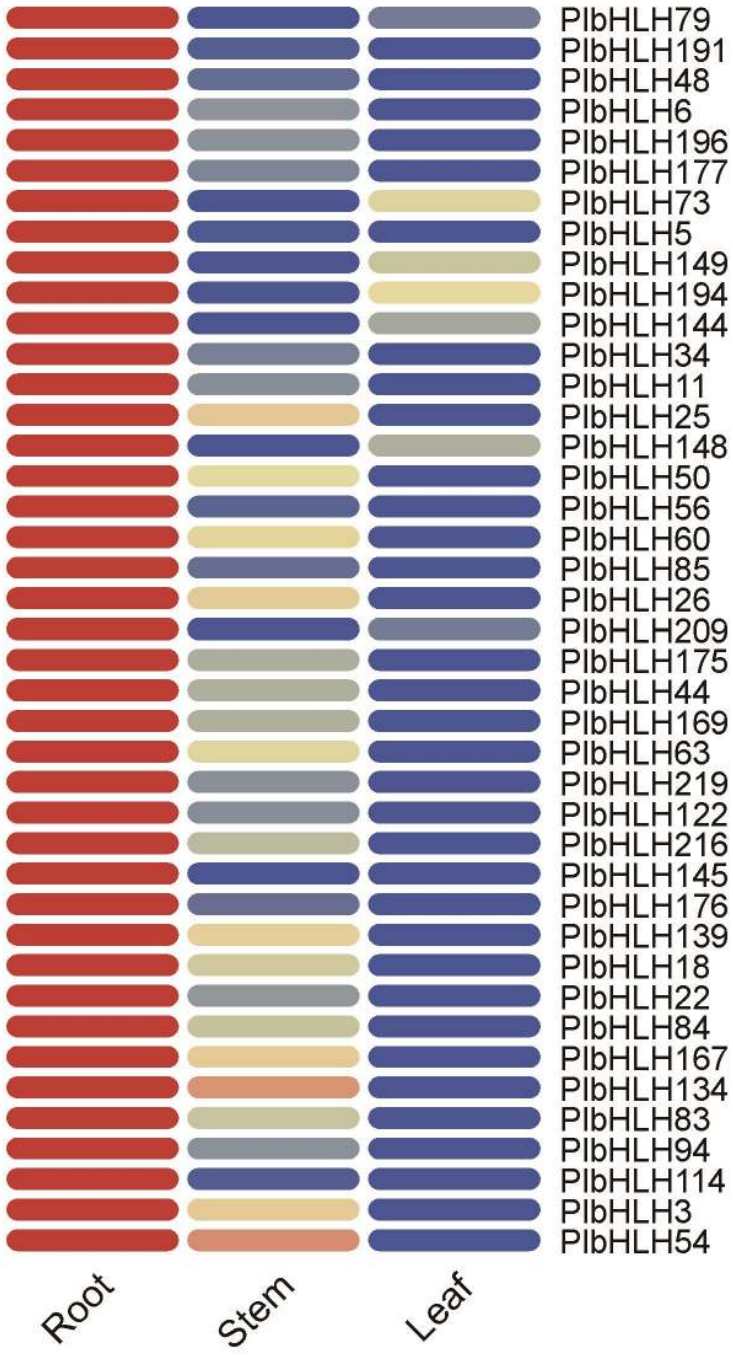


Figure 5

Heat map representation and hierarchical clustering of root-specific *PlbHLHs*. The cluster was generated using the Pearson clustering algorithm according to gene expression profiles from the transcriptome data. For each row, red and blue correspond to high and low expression values, respectively, after z-score normalized transformation.

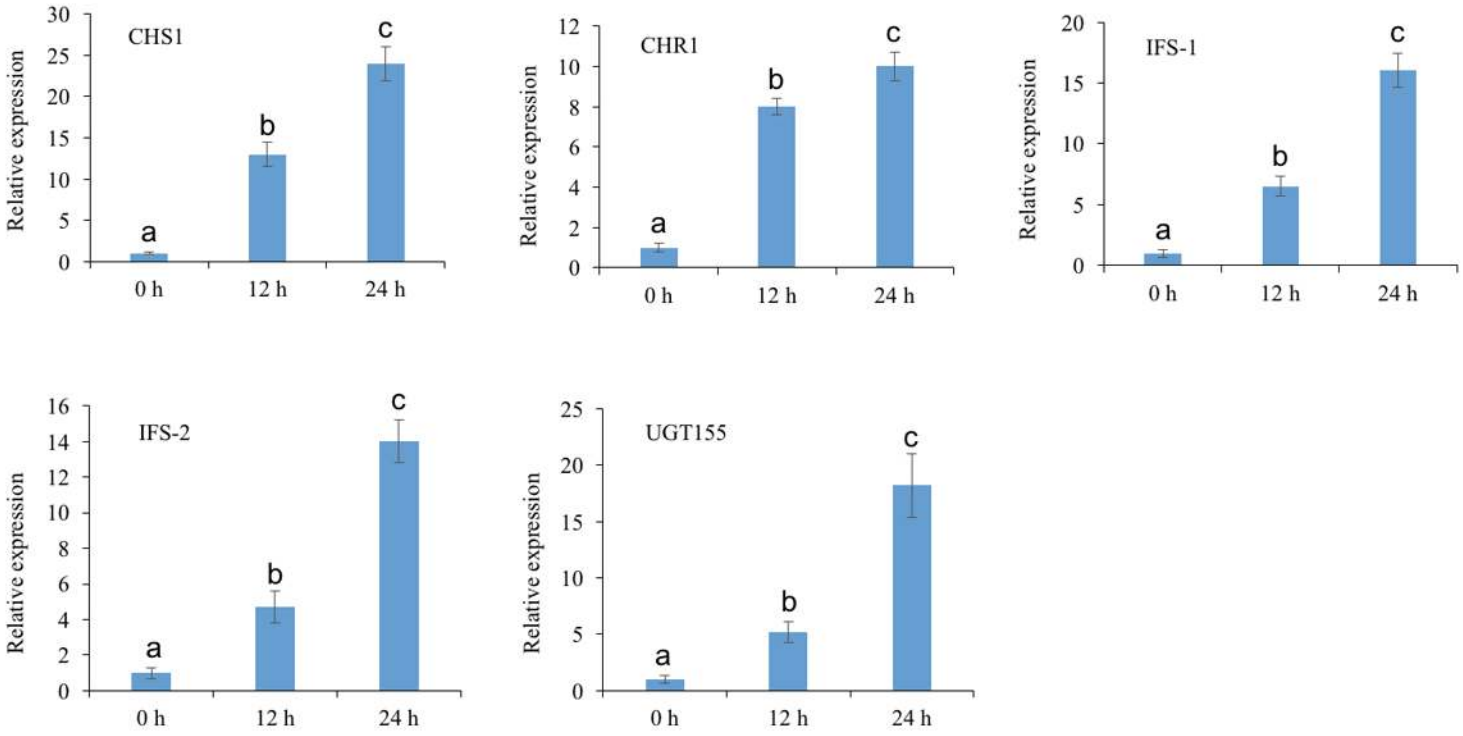


Figure 6

The relative expression pattern of five puarian biosynthesis genes under MeJA at 0 h, 12 h and 24 h. Mean values and SDs were obtained from three biology and three technical replicates. Different lower-case letters represent significant different at 0.05 level.

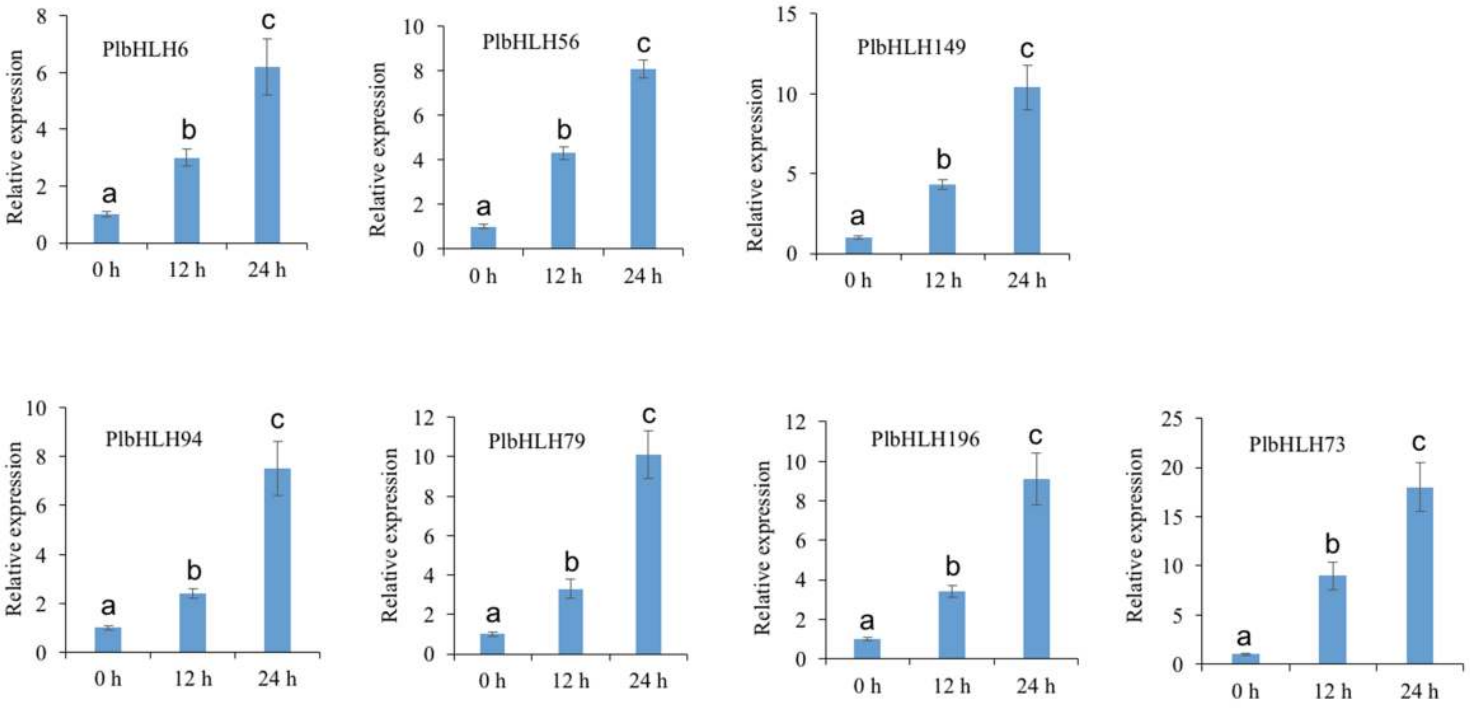


Figure 7

The relative expression pattern of seven root-specific *PlbHLH* genes in response to MeJA at 0 h, 12 h and 24 h. Mean values and SDs were obtained from three biology and three technical replicates. Different lower-case letters represent significant different at 0.05 level.

Supplementary Files

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

- FigS1.jpg
- FigS2.jpg
- FigS3.jpeg
- TableS1.xls
- TableS2.xls
- TableS3.xls
- TableS4.xls
- TableS5.xls
- TableS6.xls
- TableS7.xls
- TableS8.xls