

RESEARCH

Open Access



# Genome-wide identification of five fern *bHLH* families and functional analysis of *bHLHs* in lignin biosynthesis in *Alsophila spinulosa*

Xiong Huang<sup>1,2†</sup>, Jiangtao Fan<sup>1,2†</sup>, Cai Liu<sup>3</sup>, Peiyun Wang<sup>4</sup>, Hongfei Li<sup>1</sup>, Gang Wang<sup>1,2\*</sup> and Xiaohong Chen<sup>1\*</sup>

## Abstract

**Background** The basic helix-loop-helix (*bHLH*) transcription factors are involved in the biosynthesis of various secondary metabolites. However, genome-wide studies on the *bHLH* gene family in ferns and their role in lignin biosynthesis remain limited. As the second largest group of vascular plants, ferns are of significant interest for understanding plant evolution and secondary metabolism. Among ferns, *Alsophila spinulosa* stands out as one of the few tree ferns with a distinctive trunk structure. Investigating the genes potentially regulating lignin biosynthesis in *A. spinulosa* offers valuable insights into the growth and development mechanisms of its trunk, which is pivotal for the overall architecture and function of the plant.

**Results** In this study, we conducted a systematic study of *bHLH* gene families in five ferns, including 186 in *A. spinulosa*, 130 in *A. capillus*, 107 in *A. filiculoides*, 71 in *S. cucullata*, and 67 in *C. richardii*. Based on phylogenetic analysis, all *bHLH* genes were classified into 28 subgroups. The number of *bHLH* members in different ferns was closely related to their growth patterns and life habits, with the number in tree ferns being much larger than in other ferns. In addition, we identified tandem duplication in *C. richardii* and *A. capillus* as a key driver of their *bHLH* gene diversity, whereas in *A. spinulosa*, segmental duplication contributed more to gene expansion and evolution. Most of the *bHLH* genes in ferns are in a state of purifying selection. Additionally, tissue-specific expression patterns of *AspbHLH* genes suggest diverse functional roles in plant growth, development, and metabolite synthesis. We further focused on three genes, *AspbHLH80*, *AspbHLH120*, and *AspbHLH185*, which are specifically highly expressed in xylem. Results from weighted gene co-expression network analysis (WGCNA) and downstream target gene prediction indicate their potential regulatory roles in lignin biosynthesis.

**Conclusion** This study presents a comprehensive genomic analysis of the *bHLH* gene family in five fern species. We found a strong correlation between *bHLH* gene number and fern growth morphology, with tree ferns exhibiting a significantly higher number of *bHLH* genes. Tandem duplications were key to *bHLH* gene diversity in *C. richardii*, *A. capillus*, and *A. spinulosa*, while segmental duplications contributed more to *bHLH* gene expansion in *A. spinulosa*.

<sup>†</sup>Xiong Huang and Jiangtao Fan contributed equally to this work.

\*Correspondence:

Gang Wang  
wanggang@sicau.edu.cn  
Xiaohong Chen  
Xiaohong\_Chen@sicau.edu.cn

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



Evolutionary analysis indicated most fern *bHLH* genes are under purifying selection. Tissue-specific expression patterns of *AspbHLH* genes suggest roles in growth, development, and secondary metabolism. Furthermore, WGCNA and target gene predictions highlight three genes (*AspbHLH80*, *AspbHLH120*, and *AspbHLH185*) potentially involved in lignin biosynthesis. Overall, this work provides key insights into the mechanisms of wood formation in ferns and advances our understanding of plant secondary metabolism.

**Keywords** Fern, *bHLH* family genes, *AspbHLH* genes, Lignin biosynthesis, Expression pattern

## Background

The *bHLH* transcription factor (TF) is a member of a superfamily widely distributed in plants, animals, and fungi [1]. *bHLH* proteins contain a highly conserved *bHLH* domain, comprising approximately 60 amino acids, which is divided into two functional regions: the basic domain and helix-loop-helix (HLH) domain. The N-terminal basic region consists of approximately 10–17 amino acids, typically including five and six basic residues, which participate in DNA binding to regulate gene expression. At the C-terminus, the HLH domain, composed of roughly 40 amino acids, forms two amphipathic  $\alpha$ -helices connected by a loop, enabling the formation of homodimers or heterodimers to facilitate protein interactions [2, 3]. Plant *bHLH* TFs are typically categorized into 15–26 subfamilies based on conserved structural domains, phylogenetic relationships, and sequence homology [4]. When atypical *bHLH* proteins are included, the classification can extend to 35 subfamilies [5]. The identification of the *bHLH* gene family in plants dates back to 1989 [6]. Since then, advancements in genome sequencing technologies have enabled extensive identification and investigation of the *bHLH* superfamily across a diverse range of plant species, including *Prunus mume* [3], *Litsea cubeba* [7], *Cyclocarya paliurus* [8], *Malus domestica* [9], and others. Conducting a genome-wide characterization of *bHLH* TF is essential for advancing our understanding of the transcriptional mechanisms and functional roles attributed to the *bHLH* gene family. Numerous studies have highlighted the functional diversity of *bHLH* proteins, demonstrating their involvement in various biological processes. These include anthocyanin biosynthesis [10], the regulation of plant secondary metabolism [11], plant growth and development [12], and responses to stress [13, 14].

Pteridophytes are the second largest vascular plants on earth next to angiosperms. Vascular tissues in plants are crucial for providing physical support and transporting water, sugars, hormones, and other small signaling molecules throughout the plant [15]. The vascular system consists of two main tissue types, xylem and phloem, organized into vascular bundles [16]. Ferns possess tracheid-based xylem, which contains large amounts of lignin [17, 18]. The *bHLH* family plays a crucial role in lignin biosynthesis. In cotton (*Gossypium hirsutum*), *GhbHLH18* negatively regulates fiber strength and length

by enhancing lignin biosynthesis in fiber [19]. In *Populus trichocarpa*, overexpression of *PtrbHLH186* caused abnormal lignification and enhanced vessel cell development [20]. In *Arabidopsis thaliana*, the heterologous expression of the sorghum transcription factor *SbbHLH1* downregulates lignin biosynthesis [21]. Although *bHLH* TF families have been identified in many species, they have not yet been reported in ferns. To date, five fern genomes have been published, including *A. spinulosa* [22], *A. filiculoides* [23], *S. cucullata* [23], *C. richardii* [24], and *A. capillus* [25], paving the way for *bHLH* family investigations in ferns. These species represent different types of ferns, including aquatic ferns, terrestrial ferns, and tree ferns, spanning different evolutionary levels [22–25]. This selection provides a unique opportunity to study the evolution of gene families across distinct fern lineages, revealing how gene functions may diverge under varying environmental pressures and ecological niches. Among the five ferns, *A. spinulosa* is a tree fern with a large, erect rhizome compared to the other species, making it an ideal model for studying xylem development. In a previous study [22], we identified many G-type lignins in *A. spinulosa* xylem, but the transcriptional regulatory mechanisms of lignin biosynthesis remain unclear. Therefore, identifying *bHLH* genes in five ferns and determining the functional roles of *A. spinulosa* *bHLHs* in lignin biosynthesis are of great significance.

Here, we report a genome-wide analysis of the *bHLH* family in five ferns, identifying a non-redundant set of 558 *bHLH* genes in total. These *bHLH* proteins were comprehensively and systematically analyzed, including categorization, physicochemical properties, phylogenetic relationships, chromosome distribution, conserved motifs, and protein interactions. Expression patterns of *AspbHLH* gene in stem (xylem, pith, phloem, sclerenchymatic belt) and leaves were analyzed using transcriptome data and Quantitative Real-time PCR (qRT-PCR). Three candidate *AspbHLHs* (*AspbHLH80*, *AspbHLH120*, *AspbHLH185*) potentially involved in lignin biosynthesis were identified. This study provides not only the evolutionary relationship of the *bHLH* family in ferns but also a theoretical basis for understanding the regulatory mechanisms of *AspbHLH* TFs in lignin biosynthesis.

## Materials and methods

### Plant materials

The *A. spinulosa* materials used in this study were sampled from MeiShan City, Sichuan Province, China (29.90°N, 103.14°E). Tissues, including xylem (Xy), phloem (Ph), pith (Pi), sclerenchymatic belt (Sb), and leaf (Le), were collected from three individual *A. spinulosa* trees as biological replicates. The samples were immediately frozen in liquid nitrogen and stored at -80°C for RNA extraction.

### Identification and classification of *bHLH* genes in five ferns

The genome sequence, protein sequences, and annotation file of *A. spinulosa* and *A. capillus* were obtained from the FigShare database (<https://doi.org/10.6084/m9.figshare.19075346>; BioProject number: PRJCA006485) (<https://figshare.com/s/47be9fe90124b22d3c0e>; BioProject number: PRJNA593372) [22, 25]. The genome data of *A. filiculoides* and *S. cucullata* were obtained from the FernBase database (<https://www.fernbase.org>; BioProject number: PRJNA430527 and PRJNA430459.) [23]. The genome data for *C. richardii* were obtained from the Phytosome database ([https://phytosome-next.jgi.doe.gov/info/Crichardii\\_v2\\_1](https://phytosome-next.jgi.doe.gov/info/Crichardii_v2_1); BioProject number: PRJNA729743) [24]. A comprehensive hidden Markov model (HMM) of the bHLH domain (PF00010) was obtained from the Pfam database (<https://pfam.xfam.org>). The *bHLH* genes were identified using HMMER 3.0 with default settings and an E-value threshold of  $1e^{-10}$  [26]. The structural details of each bHLH gene identified in the five ferns were further analyzed using SMART and NCBI-CDD online tools to confirm the presence of the bHLH domain. The Expasy server ([http://web.expasy.org/prot\\_param/](http://web.expasy.org/prot_param/)) was used to calculate hydropathicity (GRAVY), isoelectric point (PI), and molecular weight (MW) [27]. Subcellular localization (SL) was predicted using WoLF PSORT online software (<https://www.genscript.com/wolf-psort.html>).

### Phylogenetic analysis and classification of *bHLH* genes

To understand evolutionary relationships and classify *bHLH* genes, a rooted neighbor-joining (NJ) phylogenetic tree of five ferns and *Arabidopsis* bHLH proteins was constructed using the MEGA 7 program [28]. The TAIR11 database (<https://www.arabidopsis.org>) provides sequences of *Arabidopsis* bHLH proteins. All protein sequences were aligned using the default settings of ClustalW. The phylogenetic tree was visualized and enhanced using Evolview2 [29].

### Analysis of gene architecture and conserved motifs

The bHLH protein sequences from the five ferns were analyzed using the MEME online program (<http://meme-suite.org/>) to predict conserved motifs [30]. The

discovery number of the pattern was set to 10, with the minimum and maximum motif widths and repetition parameters specified. The bHLH motif domains were visualized using TBtools [31]. The online Gene Structure Display Server (GSDS) (<http://gsds.cbi.pku.edu.cn/index.php>) was used to show the gene structure and extract the locations of every *bHLH* gene sequence from the genome annotation file.

### Analysis of chromosomal locations, gene duplication and collinearity

The *A. spinulosa*, *A. capillus*, and *C. richardii* bHLH genes were mapped to their respective chromosomes. Chromosomal localization analysis was not performed for *A. filiculoides* and *S. cucullata* due to the lack of chromosome-level genome assemblies. Map Chart software was used to visualize their distribution [32]. Software MCScanX was used to analyze gene duplication [33]. All protein sequences of *A. spinulosa* were compared against each other using BLASTP [34]. By using the KaKs-Calculator, the values of non-synonymous (Ka) and synonymous (Ks) words were calculated [35]. MCScanX was used to identify the synteny relationship of homologous *bHLH* genes obtained from *A. spinulosa* and other ferns (*A. capillus*, *A. filiculoides*, *S. cucullata*, and *C. richardii*).

### Promoter cis-regulatory element analysis of *bHLH* Genes

The 2000 bp upstream sequence of the *bHLH* gene was extracted from the genome file and uploaded to the Plant-Care website (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html>) for the prediction of cis-acting elements. The results of the Plant-Care study were condensed and TBtools was utilized to visualize them [31].

### Expression pattern analysis of *AspbHLH* genes

To further explore the expression patterns of the *AspbHLH* gene family in different tissues and organs, transcriptomic data for five different tissues were obtained from our previous studies [22], which are stored in the public database (BioProject: PRJCA006485). For xylem (Xy), phloem (Ph), sclerenchymatic belt (Sb) in the vascular bundle, pith (Pi) and leaf (Le), mature stems were harvested, and transverse sections were cut. The individual tissues were carefully separated using a scalpel, with xylem, phloem, and sclerenchymatic belt tissues dissected from the vascular bundle, and the pith tissue obtained from the central portion of the stem. All tissues were collected from three individual trees as biological replicates. Data processing and visualization were performed using TBtools software [31].

### qRT-PCR analysis

Total RNA was extracted from xylem (Xy), phloem (Ph), and sclerenchymatic belt (Sb) in the vascular bundle, pith

(Pi), and leaf (Le) sample using the CTAB method [36]. The qRT-PCR was performed with reference to previous studies, with *AspiActin* selected as an internal control, and the primers are shown in the Supplementary Table 1 [22]. In this study, the relative expression levels of genes in various samples were computed using the  $2^{-\Delta\Delta C_t}$  method. Excel 2019 software was used to align and compute all the data. Using the SPSS program, an analysis of variance (ANOVA) was carried out with  $p < 0.05$  as the significance threshold. GraphPad Prism 5 was used to create the graphs [37].

### WGCNA analysis and functional annotation

To further investigate the potential roles of *AspbHLH* genes in different tissues, we obtained transcriptome data from 10 tissues (leaves and stems at three different developmental periods, and xylem, phloem, pith, and sclerenchymatic belt) of *A. spinulosa*, totaling 30 samples from a previous study [22], which were subsequently used for WGCNA. The analysis was performed using the WGCNA (v1.69) package in R (v3.6.1) [38]. Using a topology-overlap-based dissimilarity metric, all genes were subjected to hierarchical clustering, and a gene dendrogram was constructed based on the Topological Overlap Matrix (TOM). Gene expression profiles for each module identified through the gene dendrogram were computed to assess associations with various plant tissues. Within the co-expression network, the edge weights between any two connected genes were determined based on their topological overlap measure, ranging from 0 to 1. Visualization of the network was performed using *Cytoscape* (v.3.8.2) software. Functional annotations of genes within modules, including Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses, were conducted following established methodologies from prior studies [22].

### Subcellular location analyses of *AspbHLH* genes

To further investigate the subcellular localization of *AspbHLH80*, *AspbHLH120*, and *AspbHLH185* genes, we amplified the stop codon-free coding sequences (CDS) and inserted them into the pCAMBIA1300 vector under the control of the CaMV 35 S promoter, with green

fluorescent protein (GFP) tagging. The amplification primers for plasmid construction were designed using SnapGene 6.0.2 software (Supplementary Table 2). Subsequently, the constructed plasmid was introduced into *Agrobacterium* GV3101 using the conventional freeze-thaw method. The experimental setup involved the nuclear localization vector NLS-RFP and the empty vector pCAMBIA1300-GFP as the control group. The experimental group consisted of the recombinant plasmid pCAMBIA1300-*AspbHLH*-GFP and the nuclear localization vector NLS-RFP. These constructs, both experimental and control, were infiltrated into 5-week-old tobacco leaves using an expression buffer (10 mM MES pH 5.6, 10 mM MgCl<sub>2</sub>, 200 μM acetosyringone) [39]. The infiltrated tobacco was then kept in the dark for 24 h, followed by a 48-hour incubation under low light conditions. Subsequently, the GFP signal was detected using Leica confocal microscopy.

### Transcription factor binding site analysis of *AspbHLH80*, *AspbHLH120*, and *AspbHLH185*

To further elucidate the transcription factor (TF) binding sites of *AspbHLH80*, *AspbHLH120*, and *AspbHLH185* within the *AspbHLHs* family, target genes were scanned using the Binding Site Prediction Tool available on the Plant Transcriptional Regulatory Maps website (<http://plantregmap.gao-lab.org/>).

## Results

### Identification and classification of *bHLH* genes in five ferns

After removing redundant and structurally incomplete candidate genes, a total of 561 *bHLH* genes were identified. Specifically, for the species *A. spinulosa*, *A. capillus*, *A. filiculoides*, *S. cucullata*, and *C. richardii*, 186, 130, 107, 71, and 67 *bHLH* genes were respectively identified, respectively (Table 1). Compared to the other four species, *A. spinulosa* exhibits the highest number of *bHLH* genes (186), and the proportion is the highest. The number of *bHLH* genes in *C. richardii* and *A. capillus* are similar, while *A. filiculoides* and *S. cucullata* have comparable numbers of *bHLH* genes. Based on the chromosomal locations of the genes, *bHLH* genes from the five species were named. The average molecular weight and

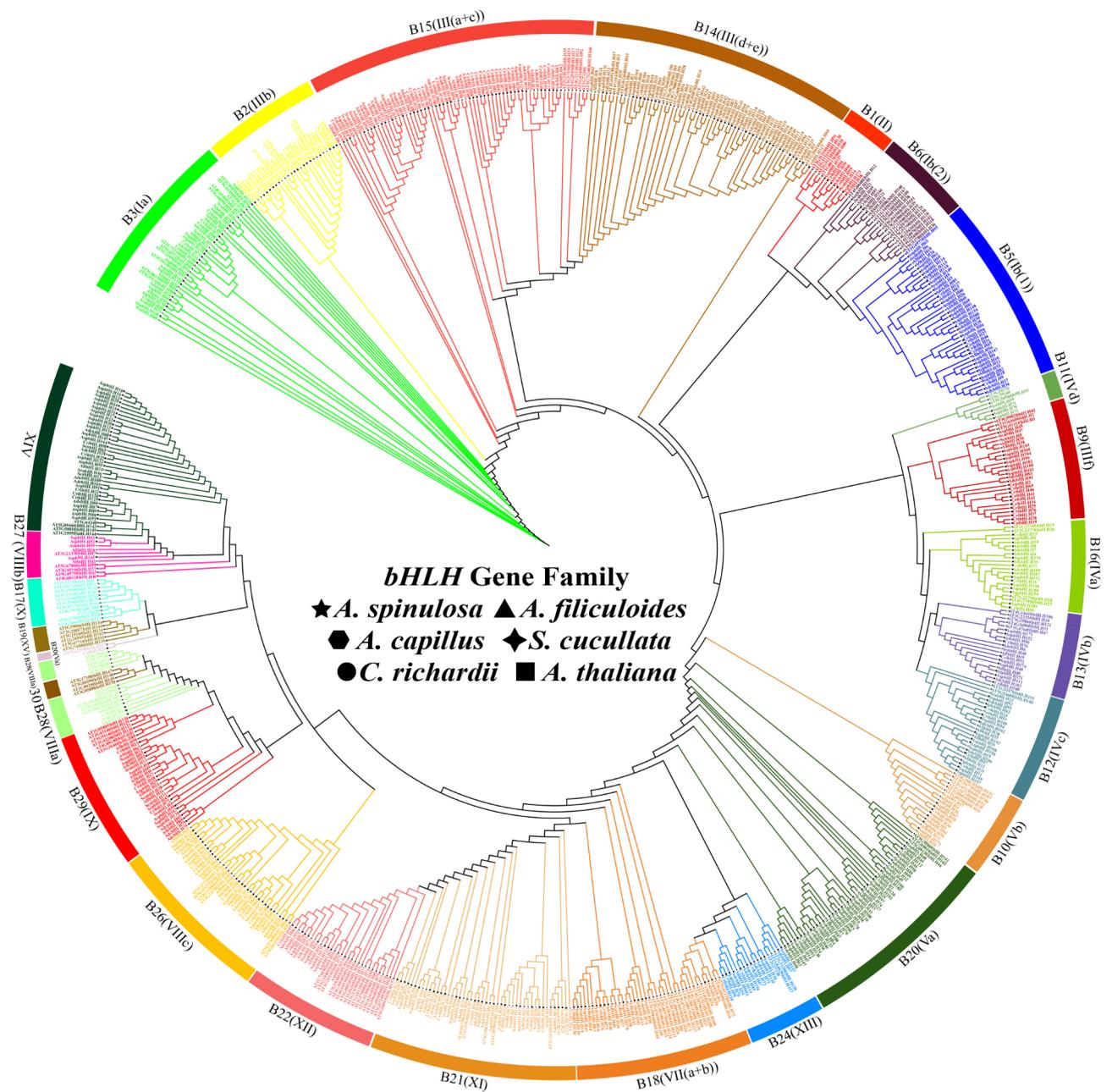
**Table 1** Number of *bHLH* genes in five ferns species

Species name	Chromosome number	Genome gene number	Identified <i>bHLH</i> genes	Proportion of <i>bHLHs</i> genes (%)	Average Molecular Weight (KD)	Average iso-electric point
<i>Alsophila spinulosa</i>	69	72,971	186	0.25	54.83	6.96
<i>Ceratopteris richardii</i>	39	75,253	130	0.17	45.01	6.75
<i>Adiantum capillus</i>	30	31,244	107	0.34	56.16	6.48
<i>Azolla filiculoides</i>	\	20,203	71	0.35	49.16	6.67
<i>Salvinia cucullata</i>	\	19,780	67	0.34	53.7	6.47

isoelectric point of these proteins ranged from 59.16 to 45.01 and 6.47 to 6.96, respectively. Subcellular localization predictions suggest that most of these proteins function in the cell nucleus, with a smaller fraction predicted to be located in other cellular compartments such as chloroplasts. Additional details and physical properties can be found in Supplementary Table 3.

**Phylogenetic analysis reveals evolutionary divergence and subfamily distribution of bHLH proteins in five fern species**

To further investigate the evolutionary relationships between the 173 *bHLH* family members in *A. thaliana* and the 561 *bHLH* proteins in five ferns, a rooted phylogenetic tree was constructed using the neighbor-joining method (Fig. 1). *Arabidopsis* was used as an outgroup. The 561 *bHLH* gene family members were classified into 30 subfamilies, with detailed information available in Supplementary Table 4. It is worth noting that none



**Fig. 1** Systematic phylogenetic analysis of bHLH proteins in five ferns and *A. thaliana*. Different colors represent different groups, and all bHLH proteins are clustered into subclades based on the priority classification rule of *A. thaliana* bHLH proteins. Different shapes along the phylogenetic tree represent bHLH proteins from different species

of the five ferns exhibited clustering of *bHLH* genes in subfamilies B17(X), B19(XV), and 30. Notably, subfamily B15(III(a+c)) stands out with the highest number of *bHLH* genes (60), while subfamily B23(VIIIa) is characterized by the lowest number of *bHLH* genes (1). Within the same subfamily, there is a substantial difference in the number of *bHLH* genes among the five ferns. For instance, in subfamily XIV, *A. spinulosa* possesses 21 members, while *C. richardii* and *A. capillus* have 5 members each, and *S. cucullata* and *A. filiculoides* exhibit only 3 and 1 member(s), respectively. Similar variations are observed in other subfamilies. The above results indicate that *bHLH* family proteins in the five ferns exhibit significant diversity and differentiation throughout the evolutionary process, providing valuable insights for future studies aiming at uncover their specific biological functions.

#### Gene structure and protein motif diversity among *bHLH* genes in five ferns

The results showed that the structures of the 561 *bHLH* genes differ significantly (Supplementary Fig. 1A-E). Some of the genes possess multiple UTR regions, while a few lack this structure. This structural diversity may reflect evolutionary differences between homologous *bHLH* genes. In addition, *bHLH* genes clustered in the same group tend to have similar exon/intron structures, exon numbers, and gene lengths. The distribution of *bHLH* protein motifs in the same subgroups was also broadly similar, suggesting that motif distribution may be functionally related. The motifs of *bHLH* proteins in five ferns were analyzed using the MEME online software tool, and the 20 motifs obtained were compared and named motifs 1 to 20 (Supplementary Table 5). Motif 1 is the basic region and the first helix, and motif 2 contains the second helix, both of which are *bHLH*-specific structural domains, and both motifs are identified in all *bHLH* proteins.

#### Cis-regulatory element diversity in the promoter regions of *bHLH* genes in five fern species

Cis-regulatory elements (CREs) in promoter regions are typically located upstream of the 5' end of genes, where they bind to transcription factors (TFs) and play a crucial role in regulating gene expression. In this study, we predicted 561 CREs in the promoter region of the *bHLH* gene. Based on their function of cis-regulatory elements and response to the environment, we classified all CREs into 15 categories. Further details on the cis-regulatory elements of the five ferns can be found in Supplementary Tables 6–1. The results showed that light-responsive elements accounted for most of the regulatory elements identified, followed by phytohormone-responsive and abiotic stress-responsive elements

(Supplementary Tables 6–2). In addition, the number of MeJA and ABA response elements was much higher than that of other phytohormone response elements (Supplementary Fig. 2A-D). Subfamily members such as B5(Ib(1)), B14(III(d+e)), and XIV possessed a large number of MeJA and ABA response elements in *A. spinulosa* (Fig. 2A). The results also revealed that 13 *AspbHLH* family members had five phytohormone-responsive cis-regulatory progenitors in their promoter sequences (Fig. 2B). These results suggest that phytohormones, particularly MeJA and ABA, play important roles in regulating the expression of *bHLH* genes in the five fern species.

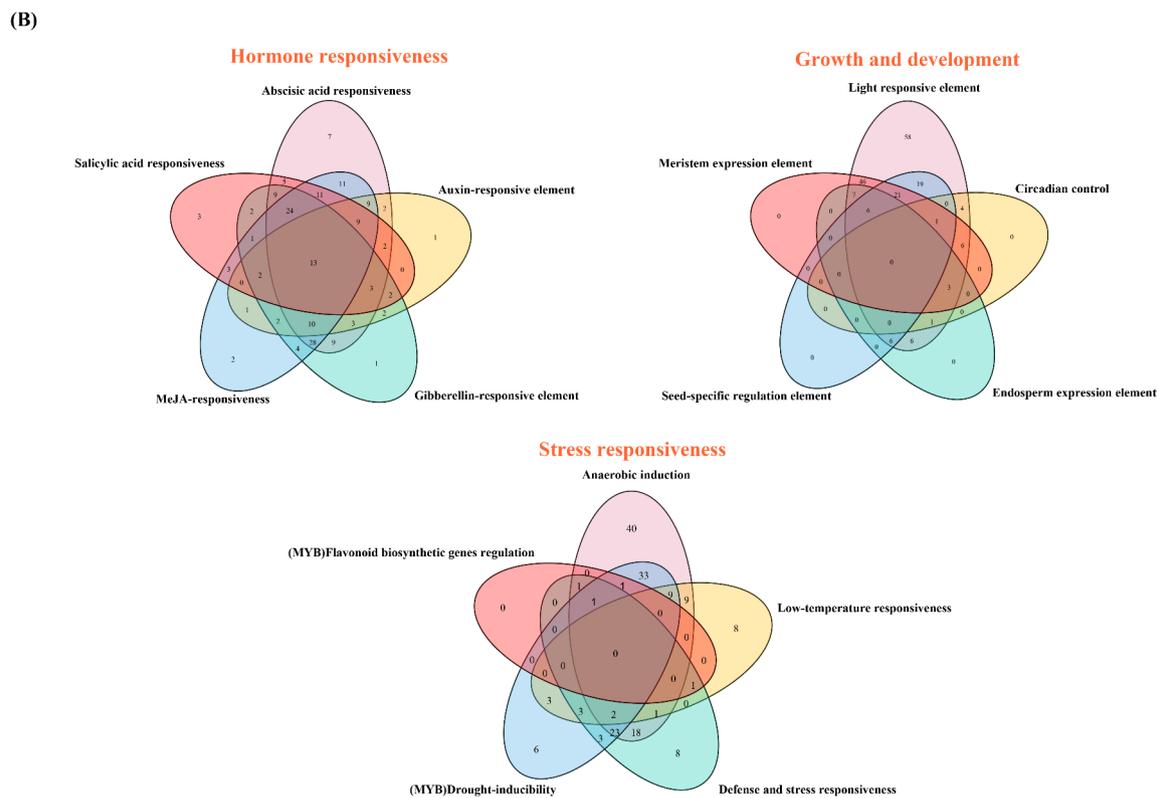
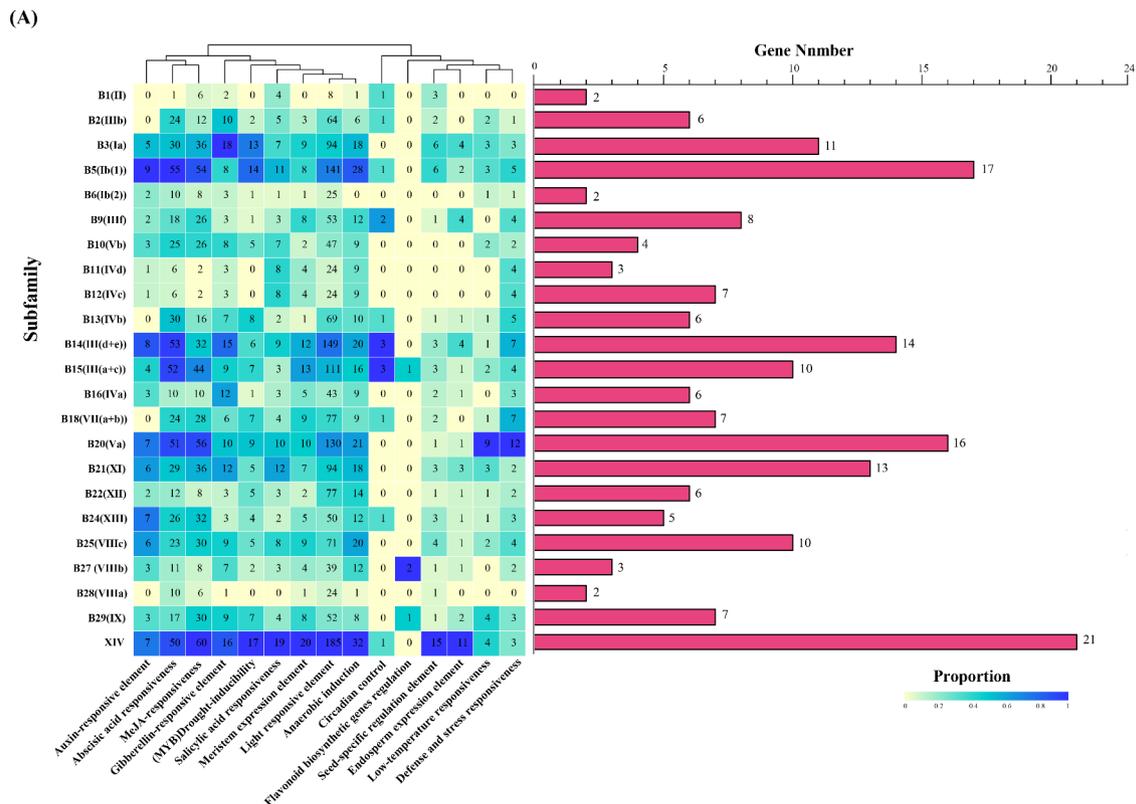
#### Diverse chromosomal distribution and evolutionary divergence of *bHLH* genes across three fern species

Since the genomes of *A. filiculoides* and *S. cucullata* have not been assembled to the chromosome level, chromosomal location information of *bHLH* genes was only available for the remaining three fern species. The results showed considerable variation in the distribution of *bHLH* genes on chromosomes. *A. spinulosa* has 69 chromosomes, with the *AspbHLH* genes distributed across only 28 of these chromosomes (Fig. 3A). Chromosome 2 has the most *AspbHLH* members, while chromosomes 12, 24, 25, and 26 have no *AspbHLH* genes. Each of the remaining chromosomes contains at least one *AspbHLH* gene. *A. capillus* has 30 chromosomes, with *AdcbHLH* genes distributed on 29 of them; only chromosome 20 lacks of *AdcbHLH* genes. *A. capillus* has 39 chromosomes, and the *AdcbHLH* genes are distributed across 34 of these, with chromosome 20 having the most *AspbHLH* members.

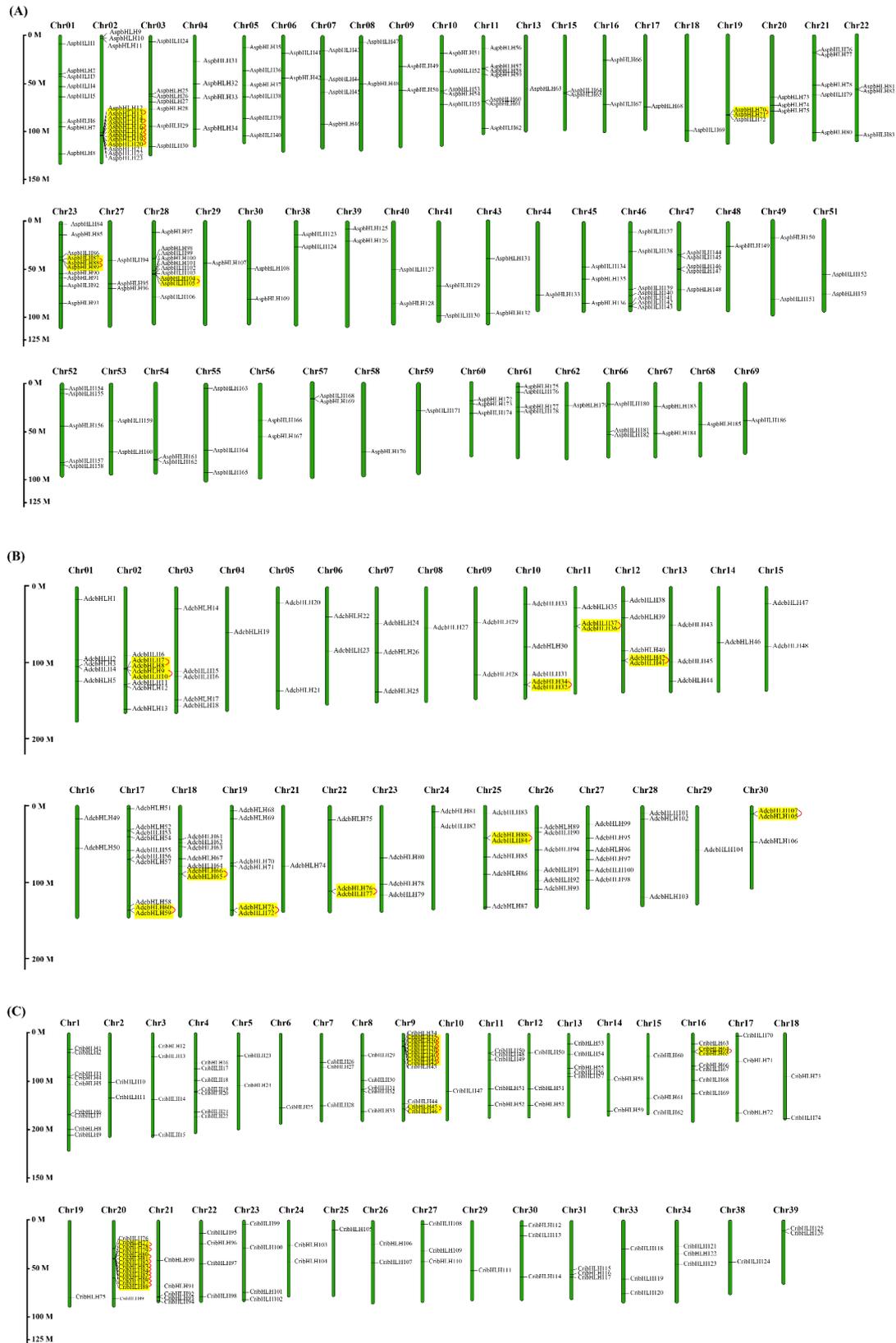
Tandem duplications of *bHLH* genes in *A. spinulosa* (11), *C. richardii* (19), and *A. capillus* (11) were identified using MCScanX (Fig. 3A-C), and several tandemly duplicated *bHLH* genes were found on the chromosomes of the three ferns. By calculating the Ka/Ks ratios of tandemly duplicated gene pairs, we gained insights into the evolutionary constraints on the *bHLH* gene family members. It was found that *AspbHLH19* & *AspbHLH20* and *CribHLH45* & *CribHLH46* have Ka/Ks values greater than 1, indicating that they evolved under positive selection pressure (Supplementary Table 7). The remaining tandemly duplicated genes have Ka/Ks values below 1, indicating that they have undergone purifying selection during their evolution. In contrast, tandem duplicate genes with Ka/Ks ratios close to 1 have undergone neutral evolution.

#### Gene duplication and synteny analysis of *AspbHLHs*

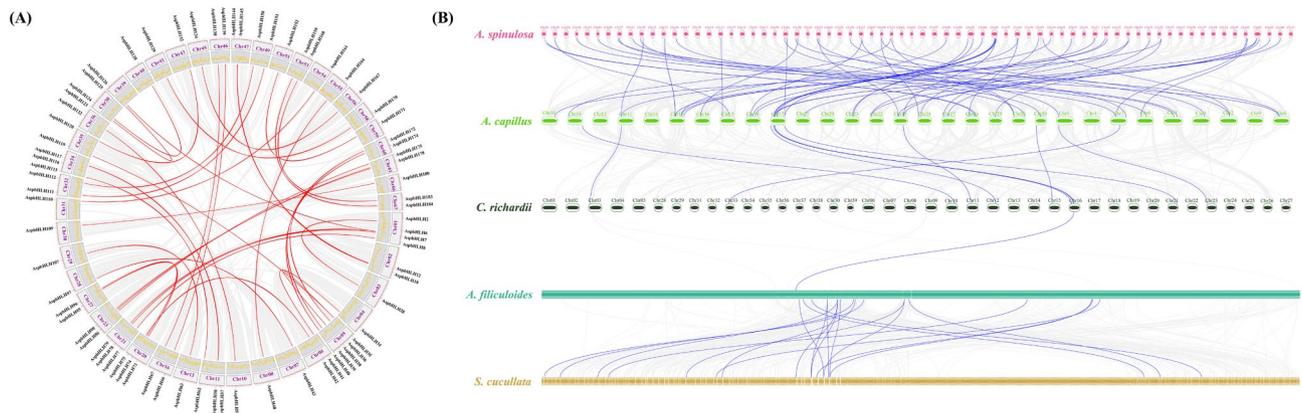
To further investigate the amplification mechanism of the *AspbHLH* gene family, a total of 76 gene duplication events were identified in this study (Fig. 4A), leading to the formation of gene pairs within *AspbHLHs*. Gene



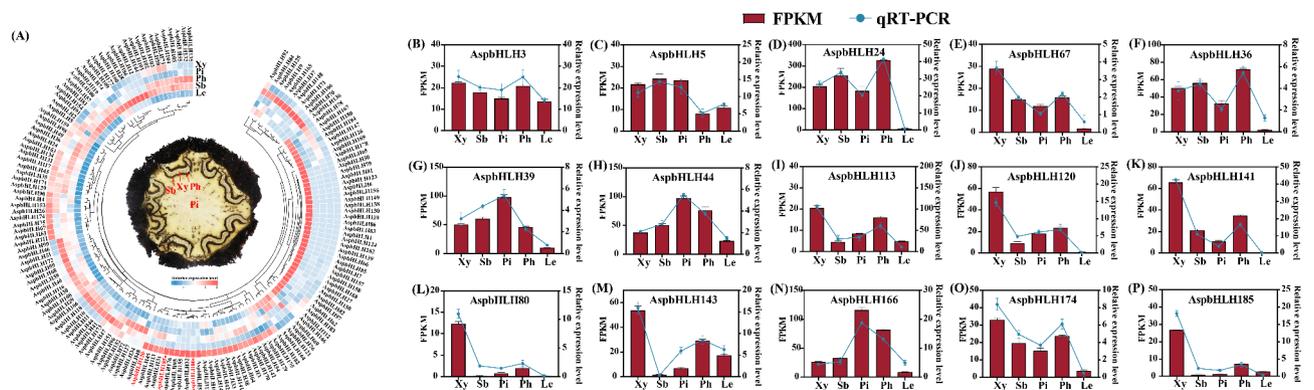
**Fig. 2** The cis-regulatory elements analysis of *bHLH* promoter regions in *A. spinulosa*. Promoter cis-acting elements in *A. spinulosa*. Different colors represent different types of cis-acting elements. (A) The number of CREs identified in the putative promoters of *AspbHLHs*. The bar chart shows the frequency of each type of CRE in the promoters of all *AspbHLHs*; (B) Venn diagram of various CREs. Venn diagram illustrating the overlap and distribution of different CREs across the promoter regions of *AspbHLHs*. This diagram highlights shared and unique cis-regulatory elements across the promoters



**Fig. 3** Chromosomal locations and Segmental duplication of *bHLH* genes. **(A)** *A. spinulosa*. **(B)** *A. capillus*. **(C)** *C. richardii*. The red arcs indicate regions of tandem duplication, while the yellow squares highlight the specific chromosomal regions where tandem duplications occur. The numbers represent the positions of the *bHLH* genes on the respective chromosomes (M), illustrating the distribution of *bHLH* genes and their duplication events



**Fig. 4** The Segmental duplication in *A. spinulosa* and collinearity analysis of *bHLH* genes in different species. **(A)** Red lines connect fragments of duplicated gene pairs, while grey lines represent co-linear blocks within *A. spinulosa*. **(B)** Multi-collinearity analysis of *bHLH* genes across five fern species. A blue line between two species indicates collinearity of genes



**Fig. 5** Expression patterns of *bHLH* genes in different organs based on transcriptome datasets. **(A)** Expression heatmap of *AspBHLHs* in xylem (Xy), phloem (Ph), sclerenchymatic belt (Sb) in the vascular bundle, pith (Pi) and leaf (Le). The center of the heat map is a cross-section of the *A. spinulosa* stem, with red arrows marking the locations of the different tissues. Genes that were not expressed in any tissue were excluded from the figures. The color bars represent the relative expression levels after column scale, where red indicates high expression levels and blue represents low expression levels. The data in the heatmap were the original FPKM values of the samples. **(B-P)** Analysis of *AspBHLHs* transcript levels in xylem (Xy), phloem (Ph), sclerenchymatic belt (Sb) in the vascular bundle, pith (Pi) and leaf (Le). Blue line represents the transcript levels of *AspBHLHs* determined by RT-qPCR, and red bar graph indicates the FPKM value. Values are from three technical replicates and three biological replicates, with error bars indicating standard deviations

duplication events were not detected in *C. richardii* and *A. capillus* (Supplementary Fig. 3A-B). To explore the evolutionary relationship of *bHLH* genes among different ferns, we subjected two terrestrial ferns, *C. richardii* and *A. capillus*, and two aquatic ferns, *A. filiculoides* and *S. cucullata* with *A. spinulosa*, to covariance analysis. It was found that the number of covariate genes was closely related to the living environment of different ferns. Specifically, 53 covariant genes were found in *A. spinulosa* and *A. capillus*, while 4, 7, and 7 covariant genes were found in *C. richardii*, *A. filiculoides*, and *S. cucullata*, respectively (Supplementary Fig. 3C-E, Supplementary Table 8). This suggests that ferns from similar habitats share more covariant genes. In addition, the number of *bHLH* gene covariates differed significantly between aquatic and terrestrial ferns (Fig. 4B). *C. richardii* exhibited less covariance with *A. filiculoides*, possibly due to the more distant evolutionary relationship between these

two species. We also observed that some genes have multiple homologs in the other two species. For example, the gene in Chr01 of *A. spinulosa* had two collinearity genes in Chr17 and Chr25 of *A. capillus*. The above results suggest that the fern *bHLH* gene family may have expanded during evolution.

**Expression patterns of *AspBHLH* genes across five tissues**

To deeply uncover the role of *bHLH* family genes and their expression patterns in the developmental process, the expression of *AspBHLH* gene was extracted from the transcriptome data in five tissues: xylem (Xy), phloem (Ph), sclerenchymatic belt (Sb) in the vascular bundle, pith (Pi) and leaf (Le) expression in five tissues (Supplementary Table 9). A circular heatmap (Fig. 5A) was created, with a cross-sectional view of the *A. spinulosa* stem at the center of the heatmap. The locations of the different tissues are marked with arrows in the figure. The



(See figure on previous page.)

**Fig. 6** Results of WGCNA analysis and KEGG, GO analysis of yellow module. **(A)** Hierarchical clustering of gene modules in 10 tissues of *A. spinulosa*, including leaves, stems at three different developmental periods, xylem, phloem, pith, and sclerenchymatic belt. **(B)** Heatmap showing the gene co-expression network across different modules in different tissues. **(C)** GO analysis for the yellow module. **(D)** KEGG analysis for the yellow module. **(E)** Co-expression network of three differentially expressed *bHLH* genes (*AspbHLH80*, *AspbHLH120*, and *AspbHLH185*) in xylem, co-expressed with genes related to lignin biosynthesis. Different colors represent the number of co-expressed genes, with redder colors indicating a higher number of co-expressed genes. Genes highlighted in red indicate co-expression with the three *bHLH* genes. **(F)** Expression patterns of lignin biosynthesis-related genes co-expressed with *AspbHLH80*, *AspbHLH120*, and *AspbHLH185* in various tissues. The heatmap data represents the original FPKM values of the samples

results showed that *AspbHLH* genes were expressed in five different groups, but some genes showed tissue-specific expression patterns. For instance, 26 genes, including *AspbHLH65*, *AspbHLH50*, *AspbHLH79*, etc., were highly expressed only in leaves (Fig. 5A). *AspbHLH12*, *AspbHLH28*, *AspbHLH44*, *AspbHLH86*, and *AspbHLH160* were highly expressed only in the sclerenchymatic belt. It is noteworthy that more than one-third of *AspbHLHs* genes were highly expressed in xylem, especially *AspbHLH80*, *AspbHLH120*, and *AspbHLH185* were significantly higher than the other four tissues (Supplementary Table 10). Prediction of target genes downstream of these three genes revealed that the target genes contained many transcription factors that regulate lignin synthesis, such as NAC and MYB (Supplementary Fig. 4A-C), suggesting that *AspbHLH80*, *AspbHLH120* and *AspbHLH185* may play roles in the transcriptional regulation of xylem development and lignin synthesis. Although most *AspbHLHs* showed differential expression in different tissues, indicating their functional diversity in different tissues, these inferences require further validation. To validate the reliability of the transcriptome data, a random selection of 15 *AspbHLH* genes was subjected to further analysis of their expression patterns using qRT-PCR. The results demonstrated consistent expression patterns of these 15 genes across five different tissues, consistent with the RNA-Seq (Fig. 5B-P).

#### Identification of key *AspbHLH* genes involved in xylem development and lignin biosynthesis

Previous results indicated that many *AspbHLH* genes were highly expressed in xylem. To further understand the potential functions of *AspbHLH* genes during xylem development or lignin synthesis, we performed WGCNA analysis using RNA-Seq data from five tissues of *A. spinulosa*, yielding a total of 14 clustered dendrogram modules (Fig. 6A). Among them, the network in the yellow module had a strong correlation with xylem (Fig. 6B). GO and KEGG enrichment analysis of the genes in the yellow module revealed that many of them were enriched in the secondary metabolite synthesis pathway and the phenylpropanoid biosynthesis pathway, among others (Fig. 6C-D, Supplementary Tables 11–12). Notably, some of the genes were significantly enriched in xylem development and lignin biosynthesis. In-depth analysis revealed that *AspbHLH80*, *AspbHLH120*, and *AspbHLH185* were also located in the yellow module, and they were potentially

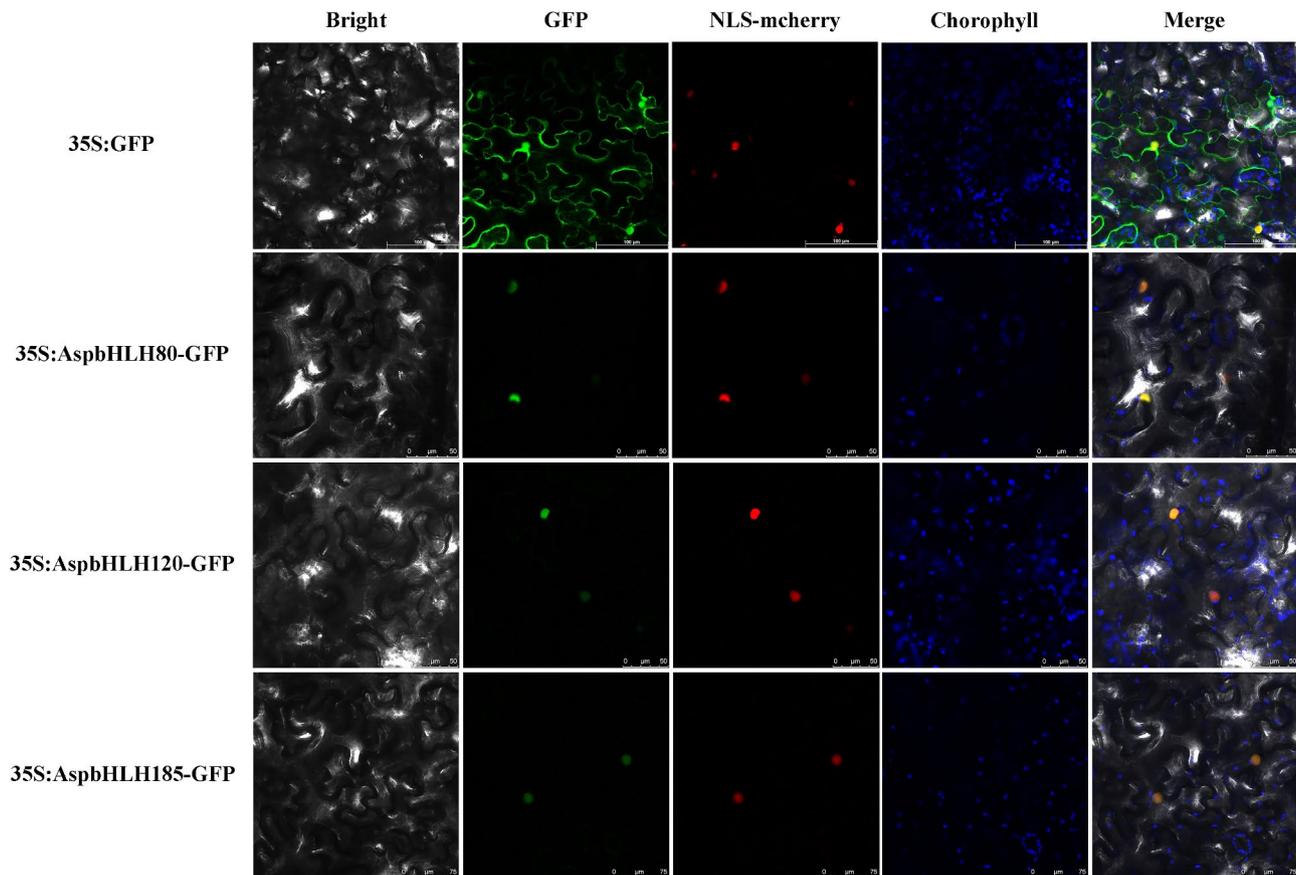
associated with several key enzyme genes related to lignin synthesis, such as *AspiHCT45*, *Aspi4CL5*, and *AspiCOMT2* (Fig. 6E), which were also highly expressed in xylem (Fig. 6F). These findings further support the potential role of the *AspbHLH* genes in lignin biosynthesis and xylem development.

#### Subcellular localization of *AspbHLHs*

Normally, transcription factors (TFs) play a role in transcriptional regulation in the nucleus [40]. To further explore the subcellular localization of the highly expressed *AspbHLH80*, *AspbHLH120*, and *AspbHLH185* genes involved in transcriptional regulation in the xylem, we predicted their subcellular localization using the WoLFPSORT online tool. The results indicated that all three genes were located in the nucleus. To validate these predictions and further investigate their potential role in regulating gene expression, we transiently expressed 35 S: GFP-*AspbHLH80*, 35 S: GFP-*AspbHLH120*, and 35 S: GFP-*AspbHLH185* fusion proteins in tobacco. Confocal microscopy analysis confirmed that these three genes were indeed localized in the nucleus (Fig. 7), suggesting that *AspbHLH80*, *AspbHLH120*, and *AspbHLH185* might be involved in regulating gene expression.

#### Discussion

*A. spinulosa* is the only extant woody fern that distinguishes itself from other ferns by the presence of vascular tissue in its main stem, which provides a strong structural support [22]. The emergence of this organization is an important symbol of the evolution of aquatic plants to land plants [41]. However, lignin is an important component of the secondary cell wall of the xylem in vascular tissues, which provides mechanical support in fiber cells and aids in water transport [42]. Studying lignin synthesis and its transcriptional regulation in *Cyathea* is crucial for understanding trunk formation and development during the evolutionary shift from aquatic to terrestrial ferns. The *bHLH* family of transcription factors, the second-largest family of transcription factors in plants, is involved in regulating various metabolic processes [43]. Previous study have shown that *bHLH* transcription factors are associated with abiotic stress and the synthesis of secondary metabolites in species such as *L. cubeba*, *C. paliurus*, and *Artemisia argyi* [7–8, 44], with evidence suggesting their potential involvement in lignin biosynthesis [20]. Despite limited research on the *bHLH* gene family in



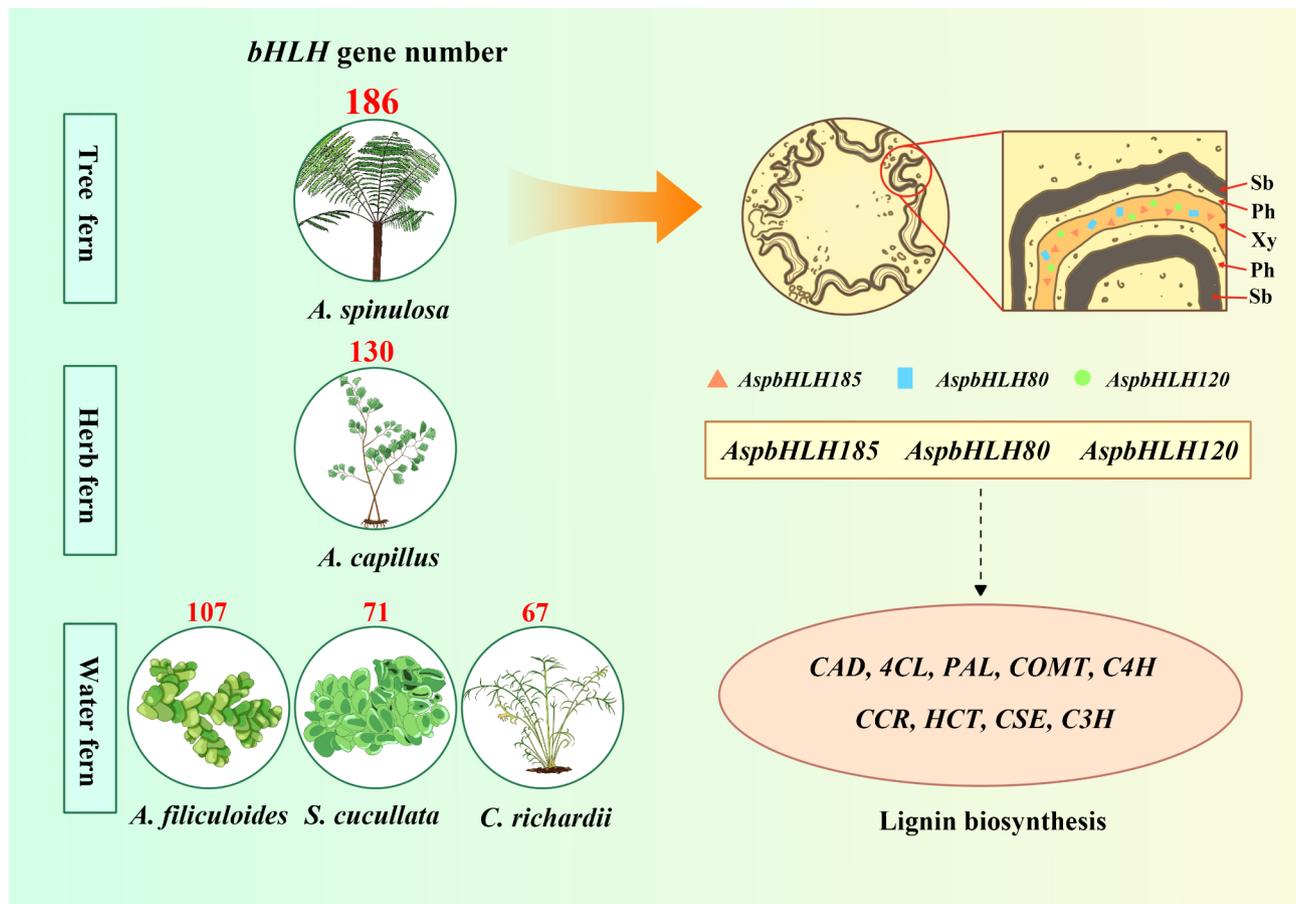
**Fig. 7** Subcellular localization of AspbHLH80, 120 and 185 proteins in *N. benthamiana* leaves

ferns, recent genomic studies of ferns provide an opportunity for a comprehensive analysis and characterization of the *bHLH* family at the genomic level [22–25].

In this study, we observe a significant correlation between the number of *bHLH* family genes identified in the five ferns and their respective habitats. Terrestrial ferns exhibited a higher abundance of *bHLH* genes compared to aquatic ferns, with this pattern being particularly evident in *A. spinulosa* (Fig. 8). This suggests that the expansion of the *bHLH* gene family may play a key role in fern growth and development and adaptation to complex environments [45]. Terrestrial ferns are exposed to more variable climatic and soil conditions and have more complex root systems and above-ground structures [46], which may require additional genes to cope with these challenges. This mechanism is prevalent in the plant kingdom; for example, a significant expansion of WRKY transcription factors has helped *Primulina huaijiensis* adapt with high salinity and water scarcity [47]. In addition, the significant expansion of *HAT* and *LIMYB* gene family members in *Tamarix chinensis* increased its disease resistance [48]. Our phylogenetic analysis suggests that *bHLH* gene diversification in ferns is shaped by ecological adaptation, with expansions in tree ferns

likely linked to structural and metabolic needs. Promoter cis-acting elements regulate gene expression, and typically, the greater the number of elements, the higher the level of transcription [49, 50]. Insights into promoter structures can provide clues about possible functions and regulatory mechanisms [51]. Research has found that the light responsive elements in the promoter region of the *bHLH* gene are the most abundant regulatory elements. The second most abundant responsive element is the hormone-related element, with the ABA responsive element having the highest number. ABA is closely related to stress and can increase the plant's resistance to stress [52], indicating that some *bHLH* genes may be induced by ABA to affect the development or stress of ferns. These findings align with research on *Prunus sibirica*, which also suggests ABA's role in regulating stress responses [53].

Genomic evolution of the *bHLH* gene in *A. spinulosa* is largely driven by gene duplication events, which include tandem and segmental duplications [54, 55]. Through gene duplication, plants can acquire new functions and expand their adaptations [56]. MCScanX analysis revealed the identification of many tandemly duplicated genes in the fern genome, most of which were



**Fig. 8** Model of *bHLH* gene family evolution and regulation in ferns: insights into lignin biosynthesis. Xy: xylem, Ph: phloem, Sb: sclerenchymatic belt, Pi: pith. The numbers on the species indicate the number of *bHLH* family members, and the dashed lines represent potential regulatory relationships

found under purifying selection (Supplementary Table 5), although some genes are under positive selection. Based on these findings, we hypothesize that gene duplication events are the main factors driving the evolution of fern *bHLH* genes [57]. In addition to segmental and tandem duplications, gene families also undergo gene loss events during evolution, which result in subgroups with varying numbers of genes per subfamily across different species [58]. Phylogenetic tree analysis showed that the 561 *bHLH* genes could be classified into 28 subfamilies, which is consistent with the results of *Carthamus tinctorius* [59]. The absence of *bHLH* genes in three subfamilies (X, XV, and 30) compared to *Arabidopsis* may be attributed to gene loss or natural selection, which led to the elimination of these genes during fern evolution.

Gene expression patterns reflect their functional roles to some extent [37]. *A. spinulosa* *bHLH* family members exhibit differential expression across various tissues (Fig. 5). This study identified certain *AspbHLH* genes that display tissue-specific expression patterns, such as *AspbHLH50* and *AspbHLH65*, which are predominantly expressed in leaves. These genes may be associated with

leaf development and stress resistance [60]. Genes such as *AspbHLH70* and *AspbHLH136* exhibit high expression in the pith. Previous research has found that sassafras pith is rich in various secondary metabolites such as starch, phenols, and terpenoids [22]. Moreover, substantial evidence suggests a correlation between *bHLH* genes and the synthesis of secondary metabolites, including terpenoids and alkaloids [44, 61]. Therefore, it is speculated that these *bHLH* genes highly expressed in the pith may be associated with the synthesis of secondary metabolites. Notably, *AspbHLH80*, *120*, and *185* are exclusively highly expressed in the xylem, with no detection in other tissues (Fig. 8). Although *bHLH* genes are not direct regulatory factors of lignin biosynthetic enzymes, studies have demonstrated their ability to indirectly regulate lignin biosynthesis by binding to other transcription factors. For example, in *G. hirsutum*, *GhbHLH18* binds to the E-box in the *GhPER8* promoter, regulating its expression, and *GhPER8* subsequently catalyzes lignin biosynthesis, which is crucial for fiber strength and elongation [19]. Overexpression of *PtrbHLH186* in *P. trichocarpa* resulted in retarded growth, increased guaiacyl lignin, a

higher proportion of smaller stem vessels, and enhanced drought tolerance, with *PtrMYB074* and *PtrWRKY19* transactivating *PtrbHLH186* through dimer formation [20]. A similar study found that the *Sorghum* transcription factor *SbbHLH1*, when heterologously expressed in *Arabidopsis*, downregulated lignin biosynthesis [21]. Overexpression of *PagLINE12* in *P. alba* resulted in significantly reduced plant height, shorter internodes, and curled leaves. Additionally, it promoted the development of secondary xylem, leading to thicker secondary cell walls compared to wild-type *P. alba* [62]. Based on these findings, we speculate that the three genes may be involved in the transcriptional regulation of lignin biosynthesis, but their specific functions require further validation.

WGCNA can cluster genes with similar expression patterns and identify interactions among these genes [38]. These genes may exhibit similar functions or participate in the same signaling pathways or physiological processes [63]. In this study, WGCNA analysis focused on studying a key module associated with wood. Enrichment analysis revealed that genes in the yellow module are enriched in various pathways, including the lignin biosynthesis pathway. Notably, the wood-specifically expressed genes *AspbHLH80*, *AspbHLH120*, and *AspbHLH185* are also found within this module. Further analysis showed that these three genes exhibit co-expression relationships with other genes related to wood formation. Additionally, downstream target gene prediction suggests their potential interaction with key transcription factors regulating lignin biosynthesis, such as NAC and MYB [64]. Therefore, we speculate that these three genes may be involved in regulating lignin synthesis. However, the exact mechanism by which *AspbHLH80*, *AspbHLH120*, and *AspbHLH185* regulate lignin synthesis remains to be further investigated. Future studies should focus on functional analysis through overexpression or silencing of these genes in *A. spinulosa*, coupled with metabolite profiling of lignin biosynthesis intermediates. Additionally, identifying their target genes using chromatin immunoprecipitation (ChIP) assays and exploring their interactions with key enzymes in the lignin biosynthetic pathway will provide a deeper insight into their regulatory roles.

The findings of this study hold significant potential for advancing plant biotechnology and genetic improvement, particularly in agricultural and industrial applications. The *bHLH* genes identified in this research are likely involved in key processes like growth regulation and lignin biosynthesis, which are crucial for enhancing plant productivity and biomass quality. For example, manipulating genes involved in lignin biosynthesis can improve the quality of biomass for bioenergy production by optimizing the lignin content for better degradation or enhancing lignin deposition for structural strength [65].

These strategies can be particularly valuable in industries such as biofuels and paper production [66]. Additionally, understanding the function of *bHLH* genes in lignin biosynthesis offers opportunities to improve wood quality in forestry. By controlling the expression of *bHLH* genes, it may be possible to create plants with enhanced or reduced lignin content, optimizing wood for various industrial applications [21]. This has broad implications, not only in agriculture but also in the development of sustainable materials and energy sources [67].

## Conclusion

In this study, we systematically investigated the *bHLH* gene families of five fern species at the genomic level and conducted a series of analyses and validations. We identified 561 members of the *bHLH* gene family, which were classified into 28 subfamilies. Our results revealed a close correlation between the number of *bHLH* members in ferns and their growth morphology and lifestyle, with tree ferns exhibiting a significantly higher number of *bHLH* members compared to other ferns. Tandem duplication was found to be a key driver of *bHLH* gene diversity in *C. richardii*, *A. capillus*, and *A. spinulosa*, while segmental duplication contributed more significantly to the expansion and evolution of *bHLH* genes in *A. spinulosa*. During evolution, most *bHLH* genes in ferns were under purifying selection. Additionally, tissue-specific expression patterns of *AspbHLH* genes were observed, indicating their diverse roles in plant growth, development, and metabolite synthesis. We also highlighted the potential functions of three genes specifically expressed in the wood, with WGCNA and downstream target gene prediction analyses suggesting their potential regulatory roles in lignin synthesis. In conclusion, this research provides valuable insights into the mechanisms and regulatory networks underlying wood formation in ferns and contributes to our understanding of plant secondary metabolism.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-11522-z>.

**Supplementary Material 1: Supplementary Fig. 1.** Evolutionary tree, motifs and gene structures of the five ferns *bHLH*. Different colors represent different motifs, and gene structures. (A) *A. spinulosa*. (B) *A. filiculoides*, (C) *S. cucullata* (D) *A. capillus*. (E) *C. richardii*.

**Supplementary Material 2: Supplementary Fig. 2.** The cis-regulatory elements analysis of *bHLH* promoter regions in four ferns. (A) *A. capillus*, (B) *A. filiculoides*, (C) *C. richardii*. (D) *S. cucullata*.

**Supplementary Material 3: Supplementary Fig. 3.** The synteny analysis of *bHLH* genes between *A. spinulosa* and other four ferns and Segmental duplication of *A. capillus* and *C. richardii* *bHLH* genes. (A-B) the grey lines are co-linear blocks in the *A. capillus* and *C. richardii*. (C-F) The gray lines were the co-linear blocks among the species, and the red lines highlighted the syntenic *bHLH* pairs between *A. spinulosa* and the species.

**Supplementary Material 4: Supplementary Fig.4.** Predictive analysis of downstream target genes for *AspbHLH80*, *AspbHLH120*, and *AspbHLH185*. (A) *AspbHLH80*, (B) *AspbHLH185*, (C) *AspbHLH120*.

**Supplementary Material 5: Supplementary Table 1.** The *AspbHLHs* primers of qRT-PCR used for detecting the expression patterns. **Supplementary Table 2.** Nucleotide sequences of primers used for gene cloning and vector construction **Supplementary Table 3.** Detailed information regarding *bHLH* gene family in five ferns. **Supplementary Table 4.** *bHLH* subfamily gene members in six species. **Supplementary Table 5.** The amino acid sequence of each conserved motif from each protein. **Supplementary Table 6–1.** Prediction of cis-regulatory elements in promoter regions of *bHLH* genes. **Supplementary Table 6–2.** Statistics on the quantitative relationship between cis-regulatory elements and *bHLH* genes. **Supplementary Table 7.** Ka, Ks and Ka/Ks of replication pairs of *bHLH* gene family in three ferns. **Supplementary Table 8.** Collinearity result among four ferns species. **Supplementary Table 9.** Expression of the *AspbHLH* family in different tissues (FPKM). (Pi: pith; Sb: sclerenchymatic belt; Xy: xylem; Ph: phloem; Le: leaf) **Supplementary Table 10.** Differential expression analysis of *AspbHLH* among five tissues (Pi: pith; Sb: sclerenchymatic belt; Xy: xylem; Ph: phloem; Le: leaf). **Supplementary Table 11.** GO enrichment analysis results of yellow module. **Supplementary Table 12.** KEGG enrichment analysis results of yellow module

#### Acknowledgements

The author thanks the lab members for assistance. This study was carried out with the support of molecular laboratory of college of forestry, sichuan agricultural university.

#### Author contributions

XC and GW formulated and designed the experiments; CL and PW collected the materials; JF and XH performed the experiments; JF and HL drawn the figures, analyzed the data and wrote the paper; XC and GW revised and proofread the paper. All authors read and approved the final manuscript.

#### Funding

This research was funded by Sichuan Province Innovative Talent Funding Project for Postdoctoral Fellows (X H, 2322999070), National Natural Science Foundation of China (X H, 32301612) and Natural Science Foundation of Sichuan Province (X C, 2022NSFSC0135; X H, 2025ZNSFSC0994) and China Postdoctoral Science Foundation (X H, 2023M742510). The funding bodies had no role in the experimental design, sample collection, data analysis and interpretation, and manuscript writing.

#### Data availability

The genome sequence, protein sequences, and annotation file of *A. spinulosa* and *A. capillus* were obtained from the FigShare database (<https://doi.org/10.6084/m9.figshare.19075346>; BioProject number: PRJCA006485) (<https://figshare.com/s/47be9fe90124b22d3c0e>; BioProject number: PRJNA593372). The genome data of *A. filiculoides* and *S. cucullata* were obtained from the FernBase database (<https://www.fernbase.org>; BioProject number: PRJNA430527 and PRJNA430459). The genome data for *C. richardii* were obtained from the Phytozome database ([https://phytozome-next.jgi.doe.gov/info/Crichardii\\_v2\\_1](https://phytozome-next.jgi.doe.gov/info/Crichardii_v2_1); BioProject number: PRJNA729743). The transcriptome data of 10 tissues (leaves and stems at three different developmental periods, and xylem, phloem, pith, and sclerenchymatic belt) of *A. spinulosa*, were retrieved from the National Center for Bioinformation database under the BioProject no. PRJCA006485.

#### Declarations

##### Ethics approval and consent to participate

The *A. spinulosa* materials were obtained from MeiShan City, Sichuan Province, Professor Xiong Huang ensures that we have permission to collect and culture the *A. spinulosa*. Meanwhile, Professor Xiaohong Chen, undertook the formal identification of the plant material used in the present study. The study was conducted in accordance with the relevant guidelines in Sichuan Agricultural University.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare no competing interests.

#### Author details

<sup>1</sup>Key Laboratory of Ecological Forestry Engineering of Sichuan Province, College of Forestry, Sichuan Agricultural University, Chengdu 611130, China

<sup>2</sup>National Forestry and Grassland Southwest Engineering Technology Research Centre of Taxus, Sichuan Agricultural University, Dujiangyan 611800, China

<sup>3</sup>Sichuan Forestry and Grassland Science and Technology Extension Station, Chengdu 610081, China

<sup>4</sup>Sichuan Provincial Forestry Station General, Chengdu 610081, China

Received: 24 October 2024 / Accepted: 24 March 2025

Published online: 09 April 2025

#### References

- An F, Xiao X, Chen T, Xue J, Luo X, Ou W, Li K, Cai J, Chen S. Systematic analysis of bHLH transcription factors in cassava uncovers their roles in postharvest physiological deterioration and cyanogenic glycosides biosynthesis. *Front Plant Sci.* 2022;13:901128.
- Massari ME, Murre C. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol Cell Biol.* 2000;20(2):429–40.
- Wu Y, Wu S, Wang X, Mao T, Bao M, Zhang J, Zhang J. Genome-wide identification and characterization of the bHLH gene family in an ornamental Woody plant *Prunus Mume*. *Hortic Plant J.* 2022;8(4):531–44.
- Buck MJ, Atchley WR. Phylogenetic analysis of plant basic helix-loop-helix proteins. *J Mol Evol.* 2003;56(6):742–50.
- Carretero-Paulet L, Galstyan A, Roig-Villanova I, Martínez-García JF, Billbao-Castro JR, Robertson DL. Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in *Arabidopsis*, Poplar, rice, Moss, and algae. *Plant Physiol.* 2010;153(3):1398–412.
- Ludwig SR, Habera LF, Dellaporta SL, Wessler SR. 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. *PNAS.* 1989;86(18):7092–6.
- Yang J, Chen Y, Gao M, Wu L, Xiong S, Wang S, Gao J, Zhao Y, Wang Y. Comprehensive identification of bHLH transcription factors in *Litsea Cubeba* reveals candidate gene involved in the monoterpene biosynthesis pathway. *Front Plant Sci.* 2022;13:1081335.
- Zhang Z, Fang J, Zhang L, Jin H, Fang S. Genome-wide identification of bHLH transcription factors and their response to salt stress in *Cyclacarya Paliurus*. *Front Plant Sci.* 2023;14:1117246.
- Li T, Shi Y, Zhu B, Zhang T, Feng Z, Wang X, Li X, You C. Genome-wide identification of Apple atypical bHLH subfamily PRE members and functional characterization of MdPRE4.3 in response to abiotic stress. *Front Genet.* 2022;13:846559.
- Hou XJ, Li JM, Liu BL, Wei L. Co-expression of basic helix-loop-helix protein (bHLH) and transcriptional activator-MYB genes induced anthocyanin biosynthesis in hairy root culture of *Nicotiana tabacum* L and *Ipomea tricolor*. *Acta Physiol Plant.* 2017;39(2):1–7.
- Goossens J, Mertens J, Goossens A. Role and functioning of bHLH transcription factors in jasmonate signalling. *J Exp Bot.* 2017;68(6):1333–47.
- Sorensen AM, Krober S, Unte US, Huijser P, Dekker K, Saedler H. *The Arabidopsis ABORTED MICROSPORES (AMS)* gene encodes a MYC class transcription factor. *Plant J.* 2003;33(2):413–23.
- Gao M, Zhu Y, Yang J, Zhang H, Cheng C, Zhang Y, Wan R, Fei Z, Wang X. Identification of the grape basic helix-loop-helix transcription factor family and characterization of expression patterns in response to different stresses. *Plant Growth Regul.* 2019;88(1):19–39.
- Gen J, Wei T, Wang Y, Huang X, Liu JH. Overexpression of PtrbHLH, a basic helix-loop-helix transcription factor from *Poncirus trifoliata*, confers enhanced cold tolerance in Pummelo (*Citrus grandis*) by modulation of H<sub>2</sub>O<sub>2</sub> level via regulating a CAT gene. *Tree Physiol.* 2019;39(12):2045–54.

15. De Rybel B, Mähönen AP, Helariutta Y, Weijers D. Plant vascular development: from early specification to differentiation. *Nat Rev Mol Cell Bio.* 2016;17(1):30–40.
16. Sakai K, Citerne S, Antelme S, Le Bris P, Daniel S, Boudier A, D'Orlando A, Cartwright A, Tellier F, Pateyron S. BdERECTA controls vasculature patterning and phloem-xylem organization in *Brachypodium distachyon*. *BMC Plant Biol.* 2021;21(1):196.
17. Pittermann J, Limm E, Rico C, Christman MA. Structure-function constraints of tracheid-based Xylem: a comparison of conifers and ferns. *New Phytol.* 2011;192(2):449–61.
18. Hiraide H, Tobimatsu Y, Yoshinaga A, Lam PY, Kobayashi M, Matsushita Y, Fukushima K, Takabe K. Localised laccase activity modulates distribution of lignin polymers in gymnosperm compression wood. *New Phytol.* 2021;230(6):2186–99.
19. Gao Z, Sun W, Wang J, Zhao C, Zuo K. *GhbHLH18* negatively regulates fiber strength and length by enhancing lignin biosynthesis in cotton fibers. *Plant Sci.* 2019;286:7–16.
20. Liu H, Gao J, Sun J, Li S, Zhang B, Wang Z, Zhou C, Sulis DB, Wang JP, Chiang VL. Dimerization of PtrMYB074 and PtrWRKY19 mediates transcriptional activation of PtrbHLH186 for secondary xylem development in *Populus trichocarpa*. *New Phytol.* 2022;234(3):918–33.
21. Yan L, Xu C, Kang Y, Gu T, Wang D, Zhao S, Xia G. The heterologous expression in *Arabidopsis Thaliana* of sorghum transcription factor *SbbHLH1* downregulates lignin synthesis. *J Exp Bot.* 2013;64(10):3021–32.
22. Huang X, Wang W, Gong T, Wickell D, Kuo L, Zhang X, Wen J, Kim H, Lu F, Zhao H, et al. The flying spider-monkey tree fern genome provides insights into fern evolution and arborescence. *Nat Plants.* 2022;8(5):500–12.
23. Li F, Brouwer P, Carretero-Paulet L, Cheng S, de Vries J, Delaux P, Eily A, Koppers N, Kuo L, Li Z, et al. Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat Plants.* 2018;4(7):460–72.
24. Marchant DB, Chen G, Cai S, Chen F, Schafran P, Jenkins J, Shu S, Plott C, Webber J, Lovell JT, et al. Dynamic genome evolution in a model fern. *Nat Plants.* 2022;8(9):1038–51.
25. Fang Y, Qin X, Liao Q, Du R, Luo X, Zhou Q, Li Z, Chen H, Jin W, Yuan Y, et al. The genome of homosporous maidenhair fern sheds light on the euphyllophyte evolution and defences. *Nat Plants.* 2022;8(9):1024–37.
26. Simon CP, Aurélien L, Sean RE, Youngmi P, Rodrigo L, Robert DF. HMMER web server: 2018 update. *Nucleic Acids Res.* 2018;46(W1):W200–4.
27. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF. Protein identification and analysis tools in the expasy server. *Methods Mol Biol.* 1999;112(112):531–52.
28. Kumar S, Stecher G, Tamura K. MEGA7. Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33(7):1870–4.
29. Subramanian B, Gao S, Lercher MJ, Hu S, Chen WH. Evolvview v3: a webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res.* 2019;47(W1):W270–5.
30. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009;37(suppl\_2):W202–8.
31. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant.* 2020;13(8):1194–202.
32. Voorrips RE. MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered.* 2002;93(1):77–8.
33. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, et al. MCSScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 2012;40(7):e49.
34. Huo SJ, Li YF, Li RP, Chen RH, Xing HT, Wang J, Zhao Y, Song XQ. Genome-wide analysis of the MADS-box gene family in *Rhododendron Hainanense* Merr. And expression analysis under heat And waterlogging stresses. *Ind Crop Prod.* 2021, 172.
35. Zhang Z, Li J, Zhao X, Wang J, Wong GK, Yu J. KaKs\_Calculator: calculating Ka and Ks through model selection and model averaging. *Genom Proteom Bioinf.* 2006;4(4):259–63.
36. Sasi S, Krishnan S, Kodackattumannil P, Shamisi AA, Aldarmaki M, Lekshmi G, Kottackal M, Amiri K. DNA-free high-quality RNA extraction from 39 difficult-to-extract plant species (representing seasonal tissues and tissue types) of 32 families, and its validation for downstream molecular applications. *Plant Methods.* 2023;19(1):84.
37. Hui W, Zheng H, Fan J, Wang J, Saba T, Wang K, Wu J, Wu H, Zhong Y, Chen G, et al. Genome-wide characterization of the MBF1 gene family and its expression pattern in different tissues and stresses in *Zanthoxylum armatum*. *BMC Genomics.* 2022;23(1):652.
38. Li H, Wen X, Wei M, Huang X, Dai S, Ruan L, Yu Y. Genome-wide identification, characterization, and expression pattern of MYB gene family in *Melastoma candidum*. *Horticulturae.* 2023;9(6):708.
39. Li B, Zhao Y, Wang S, Zhang X, Wang Y, Shen Y, Yuan Z. Genome-wide identification, gene cloning, subcellular location and expression analysis of SPL gene family in *P. granatum* L. *BMC Plant Biol.* 2021;21(1):400.
40. Bian J, Cui Y, Li J, Guan Y, Tian S, Liu X. Genome-wide analysis of PIN genes in cultivated peanuts (*Arachis Hypogaea* L.): identification, subcellular localization, evolution, and expression patterns. *BMC Genomics.* 2023;24(1):1–629.
41. Delwiche CF, Cooper ED. The evolutionary origin of a terrestrial flora. *Curr Biol.* 2015;25(19):R899–910.
42. Choi SJ, Lee Z, Kim S, Jeong E, Shim JS. Modulation of lignin biosynthesis for drought tolerance in plants. *Front Plant Sci.* 2023;14:1116426.
43. Song M, Wang H, Wang Z, Huang H, Chen S, Ma H. Genome-wide characterization and analysis of bHLH transcription factors related to anthocyanin biosynthesis in *Fig (Ficus carica)* L. *Front Plant Sci.* 2021;12:730692.
44. Yi X, Wang X, Wu L, Wang M, Yang L, Liu X, Chen S, Shi Y. Integrated analysis of basic helix loop helix transcription factor family and targeted terpenoids reveals candidate *AarbHLH* genes involved in terpenoid biosynthesis in *Artemisia argyi*. *Front Plant Sci.* 2021;12:811166.
45. Li S, Wei L, Gao Q, Xu M, Wang Y, Lin Z, Holford P, Chen Z, Zhang L. Molecular and phylogenetic evidence of parallel expansion of anion channels in plants. *Plant Physiol.* 2024;194(4):2533–48.
46. Hernandez-Rojas AC, Kluge J, Noben S, Reyes Chávez JD, Krömer T, Carvajal-Hernández CI, Salazar L, Kessler M. Phylogenetic diversity of ferns reveals different patterns of niche conservatism and habitat filtering between epiphytic and terrestrial assemblages. *Front Biogeogr.* 2021;13(3).
47. Feng C, Wang J, Wu L, Kong H, Yang L, Feng C, Wang K, Rausher M, Kang M. The genome of a cave plant, *Primulina huaijiensis*, provides insights into adaptation to limestone karst habitats. *New Phytol.* 2020;227(4):1249–63.
48. Liu JN, Fang H, Liang Q, Dong Y, Wang C, Yan L, Ma X, Zhou R, Lang X, Gai S, et al. Genomic analyses provide insights into the evolution and salinity adaptation of halophyte *Tamarix chinensis*. *GigaScience.* 2022;12:giad053.
49. Reddy PS, Rao TSB, Sharma KK, Vadez V. Genome-wide identification and characterization of the Aquaporin gene family in *Sorghum bicolor* (L). *Plant Gene.* 2015;1:18–28.
50. Chuang YC, Hung YC, Hsu CY, Yeh CM, Mitsuda N, Ohme-Takagi M, Tsai WC, Chen WH, Chen HH. A dual repeat cis-element determines expression of Geranyl Diphosphate Synthase for monoterpene production in *Phalaenopsis orchids*. *Front Plant Sci.* 2018;9:765.
51. Li J, Chen X, Zhou X, Huang H, Wu D, Shao J, Zhan R, Chen L. Identification of trihelix transcription factors in *Pogostemon Cablin* reveals PatGT-1 negatively regulates Patchoulol biosynthesis. *Ind Crop Prod.* 2021;161:113182.
52. Wang W, Qiu X, Yang Y, Kim HS, Jia X, Yu H, Kwak SS. Sweetpotato bZIP transcription factor IbABF4 confers tolerance to multiple abiotic stresses. *Front Plant Sci.* 2019;10:630.
53. Liu Q, Wen J, Wang S, Chen J, Sun Y, Liu Q, Li X, Dong S. Genome-wide identification, expression analysis, and potential roles under low-temperature stress of bHLH gene family in *Prunus sibirica*. *Front Plant Sci.* 2023;14:1267107.
54. Li G, Jin L, Sheng S. Genome-wide identification of bHLH transcription factor in medicago sativa in response to cold stress. *Genes.* 2022;13(12):2371.
55. Han W, Zhang Q, Suo Y, Li H, Diao S, Sun P, Huang L, Fu J. Identification and expression analysis of the bHLH gene family members in *Diospyros Kaki*. *Horticulturae.* 2023;9(3):380.
56. Ma R, Liu W, Li S, Zhu X, Yang J, Zhang N, Si H. Genome-wide identification, characterization and expression analysis of the *CIPK* gene family in potato (*Solanum tuberosum* L.) and the role of *StCIPK10* in response to drought and osmotic stress. *Int J Mol Sci.* 2021;22(24):13535.
57. Fan Y, Yang H, Lai D, He A, Xue G, Feng L, Chen L, Cheng X, Ruan J, Yan J, et al. Genome-wide identification and expression analysis of the bHLH transcription factor family and its response to abiotic stress in sorghum [*Sorghum bicolor* (L) Moench]. *BMC Genomics.* 2021;22(1):1–415.
58. Sheng Y, Yu H, Pan H, Qiu K, Xie Q, Chen H, Fu S, Zhang J, Zhou H. Genome-wide analysis of the gene structure, expression and protein interactions of the Peach (*Prunus persica*) *TIFY* gene family. *Front Plant Sci.* 2022;13:792802.
59. Yingqi H, Ahmad N, Yuanyuan T, Jianyu L, Liyan W, Gang W, Xiuming L, Yuanyuan D, Fawei W, Weican L, et al. Genome-wide identification, expression analysis, and subcellular localization of *Carthamus tinctorius* bHLH transcription factors. *Int J Mol Sci.* 2019;20(12):3044.

60. Wang X, Peng X, Shu X, Li Y, Wang Z, Zhuang W. Genome-wide identification and characterization of PdbHLH transcription factors related to anthocyanin biosynthesis in colored-leaf Poplar (*Populus deltoids*). *BMC Genomics*. 2022;23(1):244.
61. Liu W, Tian X, Feng Y, Hu J, Wang B, Chen S, Liu D, Liu Y. Genome-wide analysis of bHLH gene family in *Coptis chinensis* provides insights into the regulatory role in benzyloquinoline alkaloid biosynthesis. *Plant Physiol Bioch*. 2023;201:107846.
62. Song C, Guo Y, Shen W, Yao X, Xu H, Zhao Y, Li R, Lin J. PagUNE12 encodes a basic helix-loop-helix transcription factor that regulates the development of secondary vascular tissue in Poplar. *Plant Physiol*. 2023;192(2):1046–62.
63. Wei M, Li H, Wang Q, Liu R, Yang L, Li Q. Genome-wide identification and expression profiling of B3 transcription factor genes in *Populus Alba* × *Populus glandulosa*. *Front Plant Sci*. 2023;14:1193065.
64. Ohtani M, Demura T. The quest for transcriptional hubs of lignin biosynthesis: beyond the NAC-MYB-gene regulatory network model. *Curr Opin Biotech*. 2019;56:82–7.
65. Zhao S, Deng D, Wan T, Feng J, Deng L, Tian Q, Wang J, Aiman UE, Mukhaddi B, Hu X, Chen S, Qiu L, Huang L, Wei Y. Lignin bioconversion based on genome mining for ligninolytic genes in *Erwinia billingiae* QL-Z3. *Biotechnol Biofuels Bioprod*. 2024;17(1):25.
66. Yao X, Zhang G, Zhang G, Sun Q, Liu C, Chu J, Jing Y, Niu S, Fu C, Lew T, Lin J, Li X. PagARGOS promotes low-lignin wood formation in Poplar. *Plant Biotechnol J*. 2024;22(8):2201–15.
67. Iram A, Berenjian A, Demirci A. A review on the utilization of lignin as a fermentation substrate to produce lignin-modifying enzymes and other value-added products. *Molecules*. 2021;26(10):2960.

### Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.