

Genome-wide characterization of the bHLH gene family and developmental evaluation of anthocyanin-associated candidate genes in *Cymbidium sinense*

Xi Zhao¹, Linying Wang¹, Xiaotao Fan², Weiqiang Zhang², Yongyu Chen², Zhong-Jian Liu^{1,2,3*} and Siren Lan^{1,2*}

¹ Key Laboratory of National Forestry and Grassland Administration for Orchid Conservation and Utilization, College of Landscape Architecture and Art, Fujian Agriculture and Forestry University, Fuzhou 350002, China

² College of Forestry, Fujian Agriculture and Forestry University, Fuzhou 350002, China

³ Guangzhou Institute of Forestry and Landscape Architecture, Engineering Technology Research Center of National Forestry and Grassland Administration for Orchid Conservation and Utilization, Guangzhou 510405, China

* Correspondence: zjliu@fafu.edu.cn (Liu ZJ); lkzx@fafu.edu.cn (Lan S)

Abstract

Flower coloration is a key ornamental trait that largely determines the aesthetic and commercial value of *Cymbidium sinense*. Anthocyanins are major pigments underlying floral color variation, and their biosynthesis is regulated by transcription factors, including basic helix-loop-helix (bHLH) proteins. However, the bHLH gene family and its role in anthocyanin accumulation and flower coloration in *C. sinense* remain poorly understood. Here, 136 *CsbHLH* genes were identified and classified into 23 subfamilies. Gene duplication analysis identified 24 duplicated gene pairs, including 18 segmental and six tandem duplications. Cis-regulatory element analysis revealed diverse hormone- and stress-responsive elements in *CsbHLH* promoters. Transcriptome-based expression profiling revealed distinct expression patterns across floral organs. *CsbHLH43* and *CsbHLH110*, which belong to the IIIb subfamily, were selected as anthocyanin-associated candidate genes based on their phylogenetic relationships and organ-dependent expression patterns. Developmental qRT-PCR and anthocyanin measurements provided stronger correlative support for *CsbHLH43*, whose transcript abundance was significantly associated with anthocyanin content. In contrast, *CsbHLH110* exhibited a distinct stage-dependent expression pattern without a significant overall association. Both proteins were localized in the nucleus. Overall, these findings identify *CsbHLH43* as a more strongly supported anthocyanin-associated candidate, whereas the relationship between *CsbHLH110* and anthocyanin accumulation remains unclear.

Keywords: *Cymbidium sinense*, bHLH transcription factors, Anthocyanin biosynthesis, Flower coloration, Genome-wide analysis

Citation: Zhao X, Wang L, Fan X, Zhang W, Chen Y, et al. 2026. Genome-wide characterization of the bHLH gene family and developmental evaluation of anthocyanin-associated candidate genes in *Cymbidium sinense*. *Ornamental Plant Research* 6: e036 <https://doi.org/10.48130/opr-0026-0027>

Introduction

Transcription factors (TFs) are key regulatory proteins that coordinate plant adaptation to abiotic stress by regulating stress-responsive genes^[1]. The helix-loop-helix (HLH) family of transcription factors plays important roles in a range of developmental processes^[2,3]. The bHLH proteins are characterized by a conserved domain consisting of a DNA-binding basic region and a helix-loop-helix dimerization region^[4–6]. The combined effects of these elements result in the specific binding of DNA at particular sequences^[4,6–8].

Previous studies have shown that bHLH transcription factors are involved in plant stress responses, development, and signal transduction^[9–13]. Among these functions, the regulation of anthocyanin biosynthesis has attracted considerable attention^[12,14–16]. Anthocyanins are major pigments responsible for flower coloration and play important roles in plant adaptation and ornamental value.

Anthocyanin biosynthesis is regulated by a complex transcriptional network, primarily controlled by the MYB-bHLH-WD40 (MBW) regulatory complex. In this complex, bHLH transcription factors interact with MYB proteins to activate the expression of structural genes such as *CHS*, *DFR*, and *UGT*. For example, in *Arabidopsis thaliana*, the MBW complex (comprising *AtTT2/AtMYB123*, *AtTT8/AtbHLH42*, and *AtTTG1*) coordinately activates proanthocyanidin biosynthesis in seed coats by upregulating *BAN/ANR* gene expression^[17,18]. Previous studies have investigated bHLH transcription factors associated with anthocyanin accumulation in grapes^[19], plums^[20], and phalaenopsis^[21]. In

Camellia nitidissima, *bHLH79* was identified as a hub gene whose expression was associated with petal flavonoid accumulation^[22]. Thus, elucidating how bHLH-mediated networks drive floral color diversity across species remains a critical research frontier.

Orchidaceae, one of the most diverse angiosperm families, shows a cosmopolitan distribution pattern excluding Antarctica^[23]. The family encompasses in excess of 31,000 species, categorized into approximately 800 genera, and stands as the most extensive family within the monocotyledons^[24]. *C. sinense*, a traditional Chinese orchid with centuries of cultural significance, is highly valued for its elegant floral morphology, fragrance, and diverse flower coloration^[25,26]. One of the most significant ornamental features of *C. sinense* is its floral coloration. The bHLH transcription factors represent one of the largest transcription factor families in plants and play crucial roles in regulating flavonoid and anthocyanin biosynthesis^[27,28]. However, despite its horticultural importance, the bHLH gene family in *C. sinense* has not been systematically characterized, and its roles in anthocyanin biosynthesis remain largely unclear. Therefore, a systematic investigation of bHLH transcription factors is essential for understanding the molecular mechanisms underlying flower coloration in *C. sinense*.

In this study, we performed a comprehensive genome-wide analysis of the bHLH gene family in *C. sinense*. A total of 136 *CsbHLH* genes were identified and analyzed in terms of phylogenetic relationships, gene structures, conserved motifs, chromosomal distribution, and

duplication patterns. Furthermore, expression profiling combined with anthocyanin content analysis was conducted to identify candidate genes associated with floral pigmentation. These results provide a foundation for understanding bHLH genes associated with flower coloration in *C. sinense*.

Materials and methods

Plant samples

Floral organs (petals, sepals, lips, columns) were collected from *C. sinense* cultivars at the Orchid Germplasm Resource Nursery of Fujian Agriculture and Forestry University, in Fuzhou, Fujian Province, China. Samples were collected at three floral developmental stages: S1, closed floral buds with visible pigmentation; S3, half-open flowers with partially expanded floral organs; and S5, fully open flowers with fully expanded floral organs. For each developmental stage and floral organ, three independent biological replicates were collected, immediately frozen in liquid nitrogen, and stored at -80°C until analysis. Each biological sample was divided for RNA extraction and anthocyanin determination.

Characterization and profiling of *CsbHLHs*

The complete genome sequence information of *C. sinense* was originally obtained by Yang et al.^[25] and was registered in the NCBI database (NCBI: PRJNA743748, 2021). A comprehensive collection of 162 *AtbHLHs* was obtained using the TAIR database (www.arabidopsis.org, accessed on 1 March 2025). Using the conserved domains of *AtbHLHs* as a benchmark, we performed bidirectional BLAST comparisons with TBtools and removed any redundant sequences identified^[29]. Subsequently, the curated bHLH gene family sequences underwent further verification and confirmation using the Protein BLAST functionality provided by NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>, accessed on 4 April 2025). The physicochemical properties associated with the *CsbHLHs* were meticulously evaluated through the utilization of an online tool from ExPASy (<https://web.expasy.org/protparam>, accessed on 24 May 2025).

CsbHLHs phylogeny and molecular evolution

The amino acid sequences were aligned using MUSCLE in MEGA12^[30]. A phylogenetic tree was constructed, with *CsbHLHs*, *CebHLHs*, and *AtbHLHs* selected as the subjects of study. Phylogenetic reconstruction was performed under the ML criterion with 1,000 bootstrap replicates. iTOL (<https://itol.embl.de>, accessed on 12 June 2025) was further employed to refine the visualization of the evolutionary tree.

Conserved motif, domain, and sequence conservation analyses of *CsbHLHs*

Predictions regarding the protein motifs conserved in *CsbHLHs* were conducted using the MEME server (<https://meme-suite.org/meme>, accessed on 3 April 2025), while the protein domains of *CsbHLHs* were predicted using the CDD function in NCBI (www.ncbi.nlm.nih.gov/cdd, accessed on 22 June 2025). TBtools was used to integrate and visualize the protein conserved motifs, conserved domains, and their phylogenetic trees.

Multiple sequence alignment of *CsbHLH* proteins and representative bHLH homologs was visualized using Jalview software^[31]. Conserved amino acid residues, especially within the bHLH domain,

were identified based on the alignment. Sequence logo analysis was subsequently performed to evaluate the conservation of amino acid residues across aligned sequences, including both conserved motifs and the bHLH domain. The sequence logos provided a graphical representation of residue conservation at each position.

Promoter analysis of *CsbHLHs*

Through TBtools analysis, 2,000 bp upstream sequences from 136 *CsbHLHs* were identified as promoter regions using genome structure annotations. The prediction of regulatory elements within the promoter regions was conducted using PlantCARE (<https://bioinformatics.psb.ugent.be/webtools/plantcare/html>, accessed on 25 April 2025), and the results were visualized for graphical representation. The image was created using Excel 2021.

Chromosome mapping and collinearity patterns of *CsbHLHs*

Based on the GFF annotation file of *C. sinense*, the chromosomal positions of *CsbHLH* genes were obtained and visualized using TBtools. The distribution of *CsbHLH* genes on different chromosomes was illustrated to analyze their genomic organization.

Intraspecies collinearity analysis was performed to identify duplicated gene pairs within the *C. sinense* genome. The collinearity relationships were analyzed and visualized using the Advanced Circos function in TBtools. Ka and Ks values of duplicated gene pairs were obtained from the collinearity analysis output, and Ka/Ks ratios were subsequently calculated in Microsoft Excel.

To further investigate the evolutionary relationships of *CsbHLH* genes, interspecies collinearity analysis was conducted among *C. sinense*, *A. thaliana*, and *Cymbidium ensifolium*. Syntenic gene pairs were identified, and the collinearity relationships were visualized using TBtools.

Transcriptional patterns of *CsbHLHs*

Total RNA was extracted from different floral organs at S5, including petals, sepals, lips, and columns, using the RaPure Total RNA Plus Kit (Magen, China) following the manufacturer's instructions. Raw sequencing data were processed using SOAPnuke v1.4.0 (BGI-Shenzhen, China) to remove low-quality reads, adapter sequences, and reads containing more than 5% unknown bases (poly-N). Clean reads were then aligned to the reference genome using HISAT2 v2.1.0.

Gene expression levels were calculated as FPKM (Fragments Per Kilobase of exon per Million mapped reads) using RSEM v1.2.8. (Supplementary Table S1). The expression profiles were visualized using Chiplot (www.chiplot.online, accessed on 20 March 2025).

To identify anthocyanin biosynthesis-related genes in *C. sinense*, homologous genes were screened based on known anthocyanin structural genes from *A. thaliana*. Expression profiles were generated using log₂-transformed FPKM values to facilitate comparison among samples (Supplementary Table S2).

Determination of anthocyanin content

Anthocyanin content in different floral organs, including petals, sepals, lips, and columns at S1, S3, and S5, was determined using an anthocyanin assay kit based on the pH differential method. Briefly, approximately 0.05 g of fresh tissue from each sample was extracted, and the absorbance of the extracts was measured at 530 nm and 700 nm using a microplate reader. Anthocyanin content was calculated based on the absorbance difference ($A_{530} - A_{700}$), and anthocyanin content was expressed as mg g⁻¹ FW. All measurements were

performed with three biological replicates. Data are presented as the mean $\pm$ SD of three biological replicates.

Reverse transcription quantitative PCR (qRT-PCR) profiling of *CsbHLHs*

To evaluate the developmental expression patterns of *CsbHLH43* and *CsbHLH110*, qRT-PCR was performed using samples from the four floral organs collected at S1, S3, and S5. A stably expressed actin gene (*Mol022529*) from *C. sinense* was used as the internal reference gene^[32]. Using Primer Premier 5.0^[33], we developed primers targeting both the constitutive actin control and two candidate genes (Supplementary Table S3).

Reverse transcription was performed with 1 μ g of total RNA using the HiScript II 1st Strand cDNA Synthesis Kit (with gDNA wiper). The synthesized cDNA was diluted to 10 ng/ μ L, and 2 μ L of this dilution was used as the template in each qPCR reaction. Amplifications were carried out in 20 μ L volumes containing ChamQ Blue Universal SYBR qPCR Master Mix on a Bio-Rad CFX96 system arranged in a 96-well format. Fluorescence was recorded at the end of the extension step in each amplification cycle.

Quality control comprised: (i) melting-curve analysis to confirm amplicon specificity; and (ii) extraction of threshold cycle (Ct) values. Transcript abundance was normalized against the actin reference gene and expressed as $-\Delta$ Ct values, calculated as Ct_{reference} – Ct_{target}^[34]. Higher $-\Delta$ Ct values indicate greater transcript abundance. Associations between $-\Delta$ Ct values and total anthocyanin content across the 36 paired biological samples were evaluated using Spearman's rank correlation in R. $p < 0.05$ was considered statistically significant.

Subcellular localization analysis

The coding sequences of *CsbHLH43* and *CsbHLH110* without the stop codon were amplified and cloned into the Super1300-GFP vector via homologous recombination after linearization with XbaI and HindIII, generating the constructs *CsbHLH43*-GFP and *CsbHLH110*-GFP. The primers used for subcellular localization analysis are listed in Supplementary Table S4. The recombinant plasmids, together with the empty vector control (Super1300-GFP), were transformed into *Agrobacterium tumefaciens* strain GV3101 and transiently expressed in leaves of *Nicotiana benthamiana*. After infiltration, plants were incubated under dark conditions for 48 h. Leaf samples were collected and stained with DAPI to visualize nuclei. Fluorescence signals were observed using a laser scanning confocal microscope, and GFP and DAPI channels were used to determine the subcellular localization of the fusion proteins.

Results

Phylogenetic analysis and physicochemical characterization of *CsbHLHs*

A total of 136 *CsbHLH* proteins were identified in the *C. sinense* genome, and their physicochemical properties were subsequently predicted. As shown in Supplementary Table S5, the amino acid lengths of *CsbHLH* proteins ranged from 99 aa (*CsbHLH56*) to 999 aa (*CsbHLH87*), with a median length of 328 aa. The molecular weights (MWs) of the *CsbHLHs* exhibited substantial variation, ranging from 10.90 kDa (*CsbHLH56*) to 111.68 kDa (*CsbHLH87*), with an observed mean MW of 36.39 kDa. Predicted pIs varied substantially across the family (4.39–11.63), with *CsbHLH112* being the most acidic (4.39) and *CsbHLH73* the most basic (11.63).

To investigate the evolutionary relationships of *CsbHLHs*, a phylogenetic tree was constructed on the basis of 136 *CsbHLHs*, 94 *CebHLHs*, and 162 *AtbHLHs*. Phylogenetic reconstruction was performed using the Maximum Likelihood method, yielding the evolutionary relationship depicted in Fig. 1. Based on the classification of *AtbHLHs*^[35], all members were grouped into 23 subfamilies. Furthermore, the XII subfamily exhibited the maximum number of *CsbHLHs* (14 members), with the VIII (b + c) subfamily following closely behind (13 members). Conversely, the Ib, IVb, and XIV subfamilies demonstrated the lowest number of *CsbHLHs*, with each subfamily having only two such gene members. Notably, several *CsbHLH* proteins were clustered within the IIIf subfamily, which includes well-characterized bHLH regulators involved in anthocyanin biosynthesis in *A. thaliana*.

Gene structure, conserved motifs, and sequence logo analysis of *CsbHLHs*

To investigate the structural features of *CsbHLH* proteins, conserved motifs were identified using the MEME program, and a total of 15 motifs were detected (Fig. 2). Among these, motifs 1, 2, and 3 were widely distributed among most *CsbHLH* proteins, suggesting that they represent core conserved elements of the bHLH family.

Sequence logo analysis revealed that motifs 1 and 2 contained highly conserved basic amino acid residues, which are characteristic of the DNA-binding region of bHLH transcription factors, whereas motif 3 corresponded to the helix-loop-helix (HLH) region associated with protein dimerization (Fig. 3a). These conserved sequence features indicate that *CsbHLHs* possess the typical structural characteristics of bHLH proteins (Fig. 3b). In addition, *CsbHLH* members within the same subfamily exhibited similar motif compositions, while certain motifs were specific to particular subfamilies, suggesting potential functional divergence during evolution.

Gene structure analysis showed that all *CsbHLHs* contained introns, with intron numbers ranging from 1 to 12. Notably, 11 *CsbHLHs* contained only a single intron, indicating structural variation within the gene family.

Chromosomal distribution, gene duplication, and synteny analysis of *CsbHLHs*

The chromosomal distribution of 136 *CsbHLH* genes in *C. sinense* was analyzed to investigate their genomic organization. These genes were unevenly distributed across 20 chromosomes (Fig. 4). The number of *CsbHLHs* on individual chromosomes ranged from 1 to 15, indicating substantial variation. Chr 2 harbored the highest number of *CsbHLHs* (15), followed by Chr 5 (14), whereas Chr 15 contained only one gene.

In addition, except for chromosomes 3, 15, 19, and 20, most chromosomes exhibited clustered distributions of *CsbHLHs*, suggesting possible gene duplication events. Intrasynteny analysis identified a total of 18 segmentally duplicated gene pairs in *C. sinense* (Fig. 5a). Duplication analysis further revealed 24 duplicated gene pairs, including 18 segmental duplication events and six tandem duplication events.

Consistent with these results, several clusters of *CsbHLH* genes were observed on multiple chromosomes in the chromosomal distribution map (Fig. 4), supporting the presence of tandem duplication events. In contrast, segmental duplication events were primarily inferred from collinearity relationships across different chromosomal regions. These findings indicate that both tandem and segmental duplication contributed to the expansion of the *CsbHLH* gene family, with segmental duplication playing a major

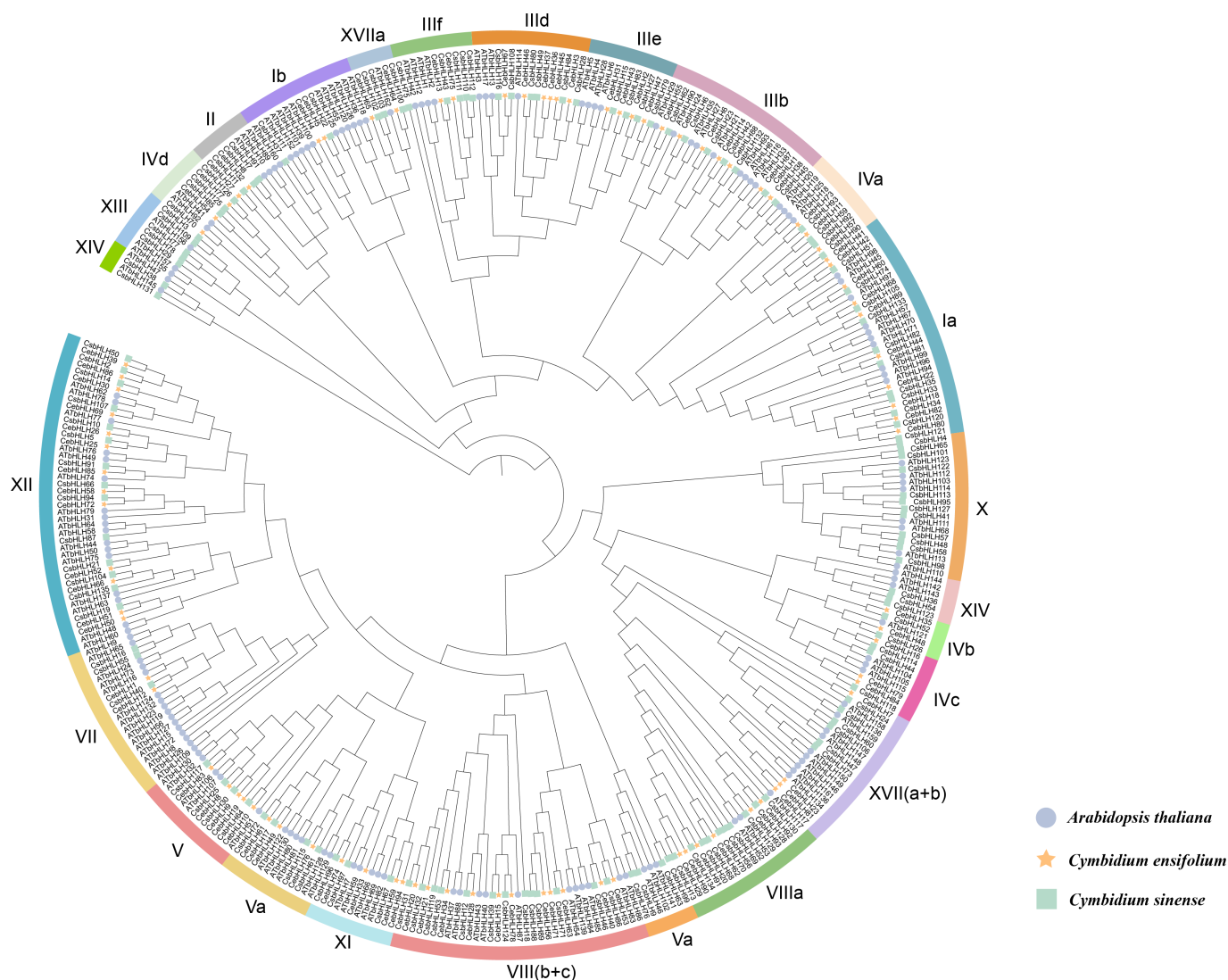


Fig. 1 Comparative phylogeny of bHLH protein families across *C. sinense* (136), *C. ensifolium* (94), and *A. thaliana* (162). Circle: *AtbHLHs*. Pentagram: *CebHLHs*. Square: *CsbHLHs*. Color-coded bands represent the 23 subfamilies.

role. To further evaluate the evolutionary constraints on duplicated genes, the Ka/Ks ratios of duplicated *CsbHLH* gene pairs were calculated. All Ka/Ks values were less than 1, indicating that these genes have undergone strong purifying selection during evolution (Supplementary Table S6).

To further explore the evolutionary conservation of *CsbHLH* genes, synteny analysis was performed between *C. sinense* and other plant species, including *C. ensifolium* and *A. thaliana* (Fig. 5b). Numerous collinear gene pairs were identified between *C. sinense* and *C. ensifolium*, indicating a high level of conservation between these closely related species. In contrast, fewer collinear gene pairs were detected between *C. sinense* and *A. thaliana*, suggesting a more distant evolutionary relationship. These results suggest that *CsbHLH* genes are relatively conserved during evolution, particularly among closely related species.

Cis-acting element analysis of *CsbHLH* promoters

To explore the potential regulatory features of *CsbHLH* genes, cis-acting elements in their promoter regions were analyzed (Fig. 6). The identified elements were classified into four categories: light-responsive elements, growth and development-related elements, stress-responsive elements, and hormone-responsive elements.

Among these, hormone-responsive elements were the most abundant (1,463), followed by light-responsive elements (910), indicating that *CsbHLHs* may be widely involved in hormone signaling and light response pathways. MBSI elements were identified in the promoters of *CsbHLH3* and *CsbHLH55*, indicating the presence of potential MYB-binding sites. Overall, the diversity of cis-acting elements in *CsbHLH* promoters suggests complex regulatory patterns that may contribute to the functional diversification of this gene family.

Expression profiling of *CsbHLHs* and prioritization of anthocyanin-associated candidate genes

To identify candidate *CsbHLH* genes potentially associated with floral anthocyanin accumulation, transcriptome analysis was performed across four floral organs, including columns, petals, sepals, and lips, at S5 (Fig. 7a). The results showed that *CsbHLHs* exhibited diverse expression patterns, indicating potential functional divergence among family members. Among these genes, *CsbHLH43* and *CsbHLH110* showed organ-dependent expression patterns in floral tissues. Based on their phylogenetic placement

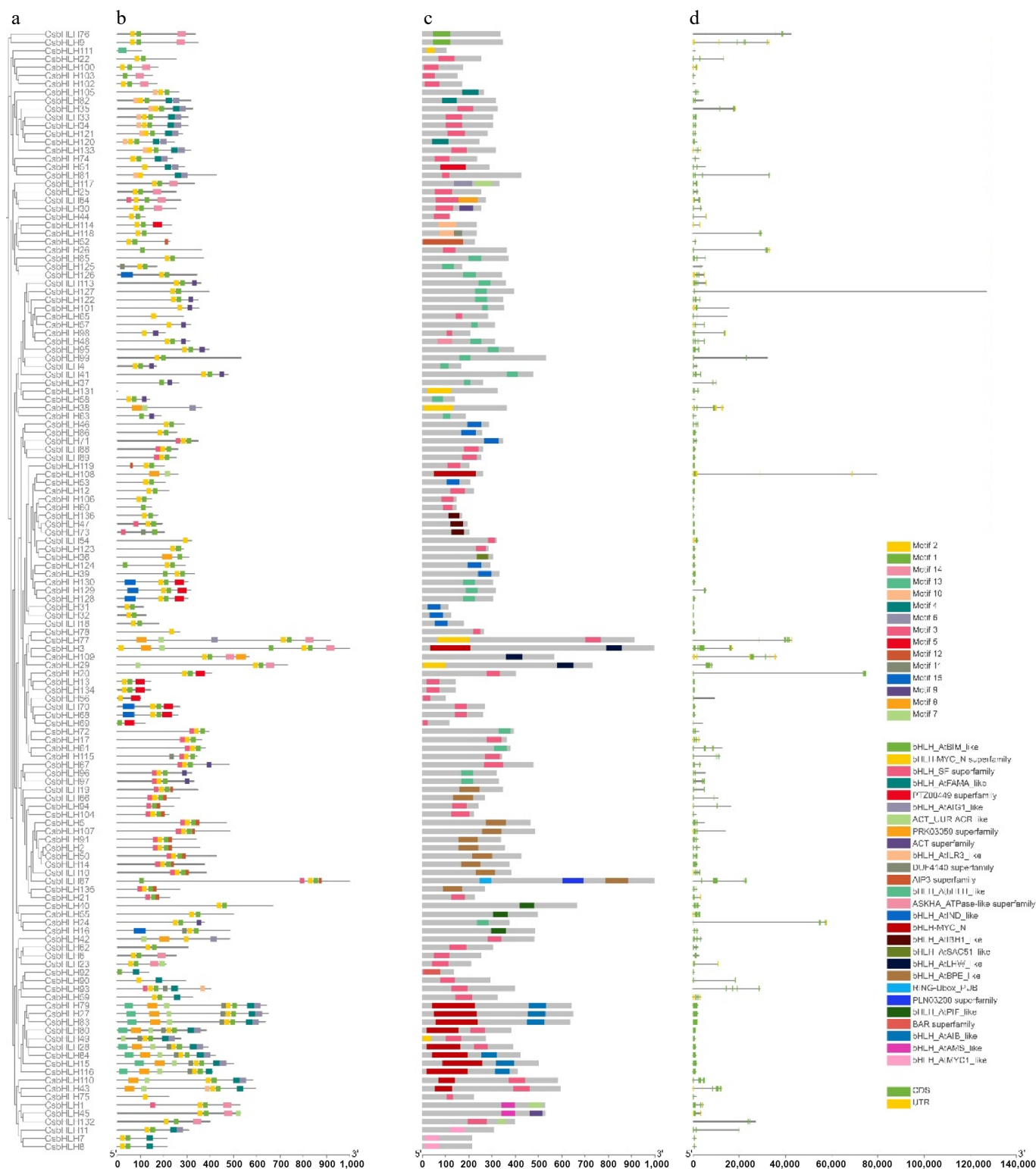


Fig. 2 Motif map, gene structure, and conserved motifs of bHLH gene family members in *C. sinense*. (a) Motifs of *CsbHLHs*; (b) domains of *CsbHLHs*; (c) gene structure of *CsbHLHs*. The legend is on the right.

within subgroup IIIf and their floral-organ expression patterns, these genes were prioritized for further evaluation. Their expression patterns partially overlapped with those of representative anthocyanin biosynthesis-related genes, including *UGT* (particularly *Mol014948*), *CHI*, and *F3'H* (Fig. 7b), further supporting their candidacy as anthocyanin-associated bHLH genes.

Developmental expression patterns of candidate genes and their associations with anthocyanin accumulation

Total anthocyanin content increased progressively from S1 to S5 in all four floral organs. Sepals consistently showed the highest anthocyanin content, followed by petals, lips, and columns (Fig. 8a).

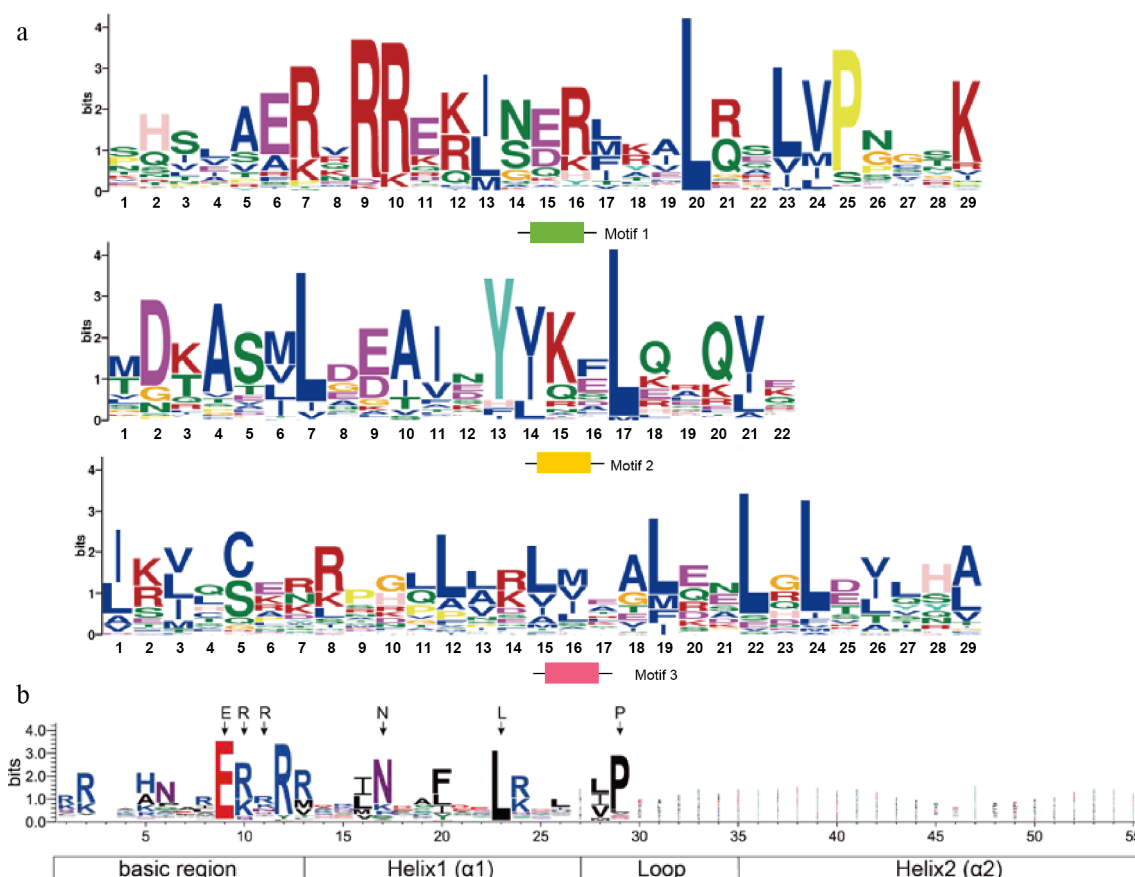


Fig. 3 Sequence logo analysis of conserved motifs in *C. sinense* bHLH proteins. (a) Sequence logos of the major conserved motifs identified by MEME analysis, highlighting conserved residues within the bHLH domain. (b) Sequence logo of the representative bHLH domain, showing the characteristic basic region and helix-loop-helix structure.

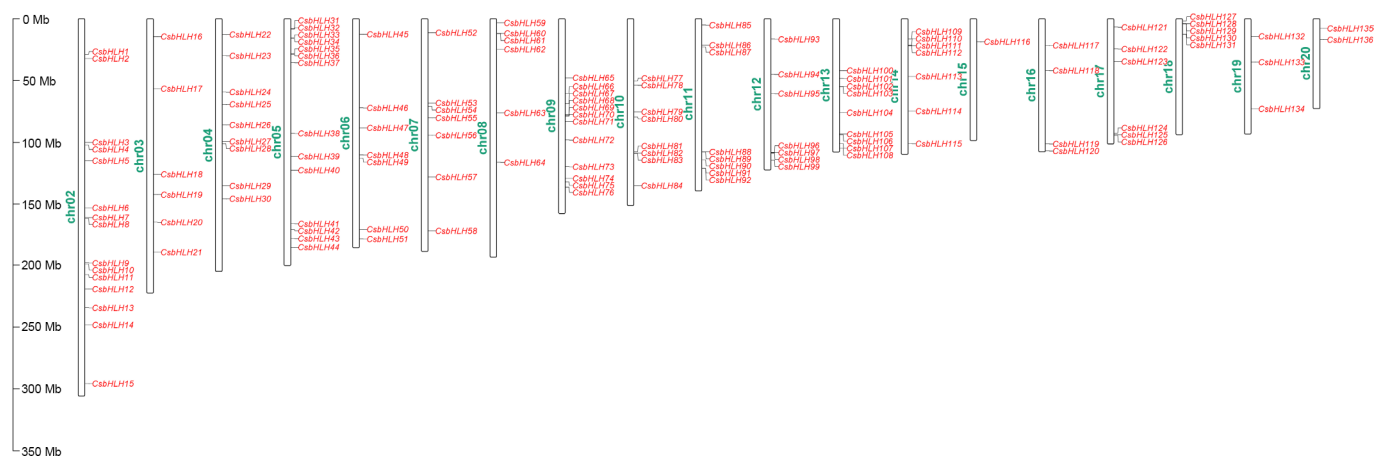


Fig. 4 Distributions of 136 *CsbHLHs* on chromosomes.

CsbHLH43 and *CsbHLH110* exhibited distinct developmental expression patterns. Transcript abundance of *CsbHLH43* increased at S5 in the lip, petal, and column, whereas relatively limited developmental variation was observed in the sepal. In contrast, *CsbHLH110* generally showed increased transcript abundance at S3 followed by a reduction at S5 (Fig. 8b, c).

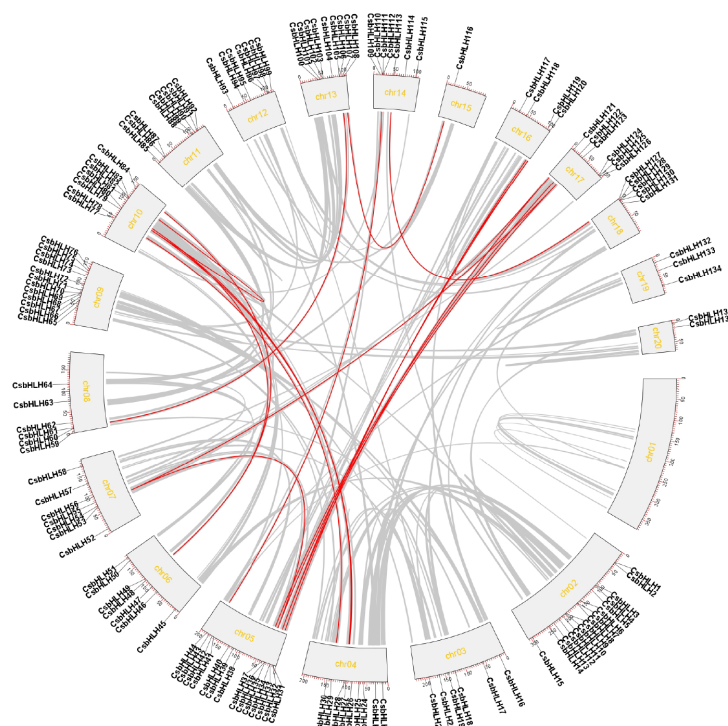
Across the 36 pooled biological samples, *CsbHLH43* transcript abundance was positively associated with total anthocyanin content (Spearman's $\rho = 0.695$, $p < 0.001$). In contrast, the association between *CsbHLH110* transcript abundance and anthocyanin content

was not significant ($p = 0.250$, $p = 0.142$; Fig. 8d, e). These results provided stronger developmental correlative support for *CsbHLH43* than for *CsbHLH110*.

Sequence features and subcellular localization of candidate *CsbHLHs*

Phylogenetic analysis showed that *CsbHLH43* and *CsbHLH110* clustered within the IIIf subfamily, which contains several previously characterized bHLH proteins associated with anthocyanin

a



b

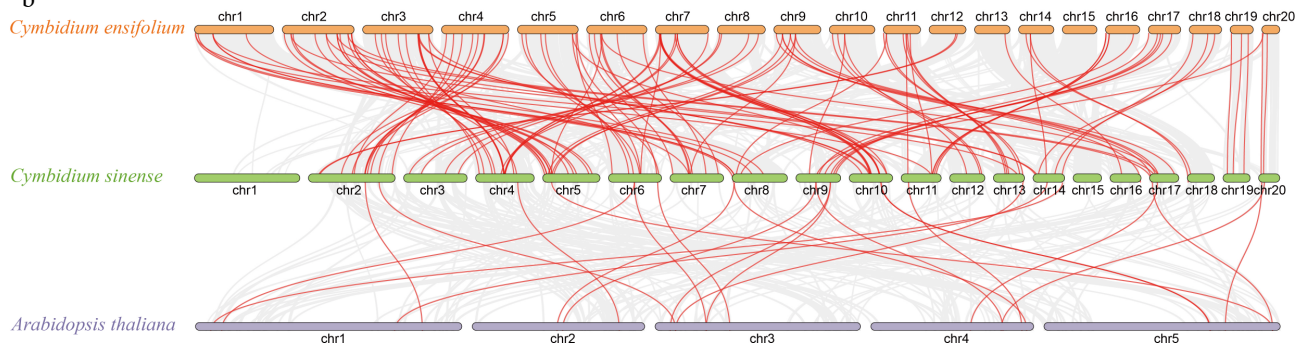


Fig. 5 Synteny analysis of CsbHLH genes within *C. sinense* and among *C. sinense*, *C. ensifolium*, and *A. thaliana*. (a) Intraspecific synteny analysis of CsbHLH genes in *C. sinense*. (b) Interspecific synteny analysis of bHLH genes among *C. sinense*, *C. ensifolium*, and *A. thaliana*.

regulation in other plant species. Multiple sequence alignment further showed that both proteins contained the conserved basic region and helix-loop-helix domain characteristic of bHLH transcription factors, together with conserved amino acid residues shared with representative anthocyanin-associated homologs (Fig. 9).

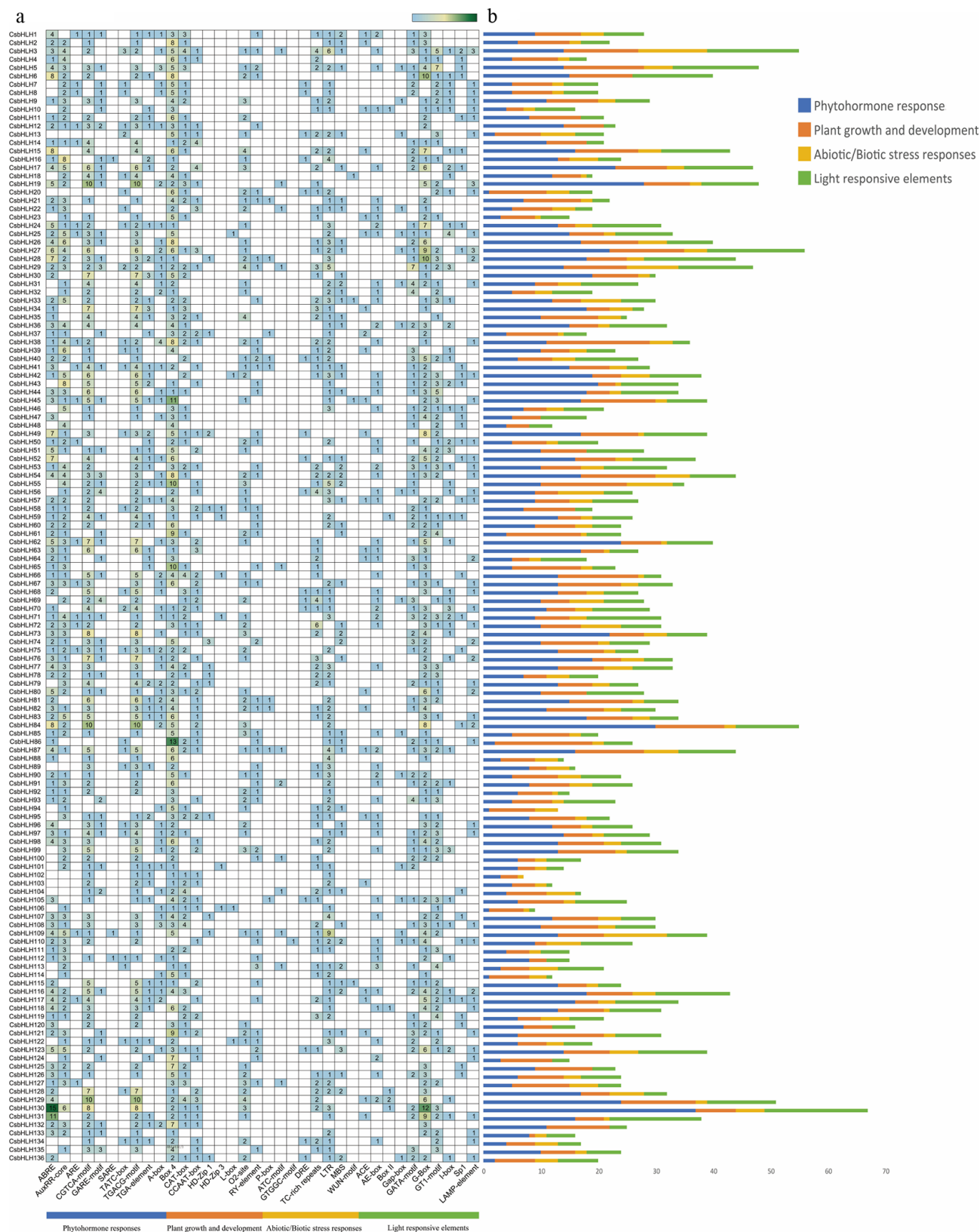
To further characterize *CsbHLH43* and *CsbHLH110*, their subcellular localization was investigated using a transient expression system in *N. benthamiana* (Fig. 10). In leaves expressing the empty vector Super1300-GFP, GFP fluorescence was observed in both the nucleus and cytoplasm. In contrast, GFP signals from *CsbHLH43*-GFP and *CsbHLH110*-GFP were exclusively detected in the nucleus and showed strong colocalization with DAPI-stained nuclear signals. No obvious GFP fluorescence was observed in the cytoplasm or plasma membrane. These results indicate that *CsbHLH43* and *CsbHLH110* are nucleus-localized proteins, consistent with their predicted roles as transcription factors.

Discussion

The bHLH transcription factor family plays important roles in plant development, stress responses, and secondary metabolism^[27,36].

Previous studies have mainly focused on the roles of bHLH transcription factors in stress responses, plant development^[37,38], and signal transduction. Within these areas, the regulatory mechanisms governing anthocyanin synthesis, as exemplified by the function of bHLH, have emerged as a pivotal subject of study^[19]. However, the composition, evolution, and potential roles of bHLH transcription factors in *C. sinense*, particularly in floral pigmentation, remain poorly understood.

Phylogenetic analysis showed that CsbHLH proteins were distributed across multiple bHLH subfamilies, indicating broad diversification of this gene family in *C. sinense*. Members within the same subfamily generally exhibited similar structural features, supporting their evolutionary conservation. Notably, *CsbHLH43* and *CsbHLH110* clustered within the IIIf subfamily, which includes several anthocyanin-associated bHLH regulators reported in other plant species. In *A. thaliana*, IIIf members such as *AtGL3*, *AtEGL3*, and *AtTT8* have been functionally characterized as regulators of anthocyanin accumulation. The clustering of *CsbHLH43* and *CsbHLH110* with these proteins suggests that they share evolutionary and sequence features with anthocyanin-associated bHLHs, making them suitable candidates for further functional investigation. Duplication and



Ka/Ks analyses further indicated that gene duplication contributed to the expansion of the *CsbHLH* family and that duplicated genes were mainly subject to purifying selection, suggesting functional conservation during evolution [39].

Structural analysis showed that *CsbHLHs* within the same sub-family shared similar motif compositions and exon-intron structures, indicating conserved sequence features. Conserved motifs corresponding to the basic region and helix-loop-helix domain

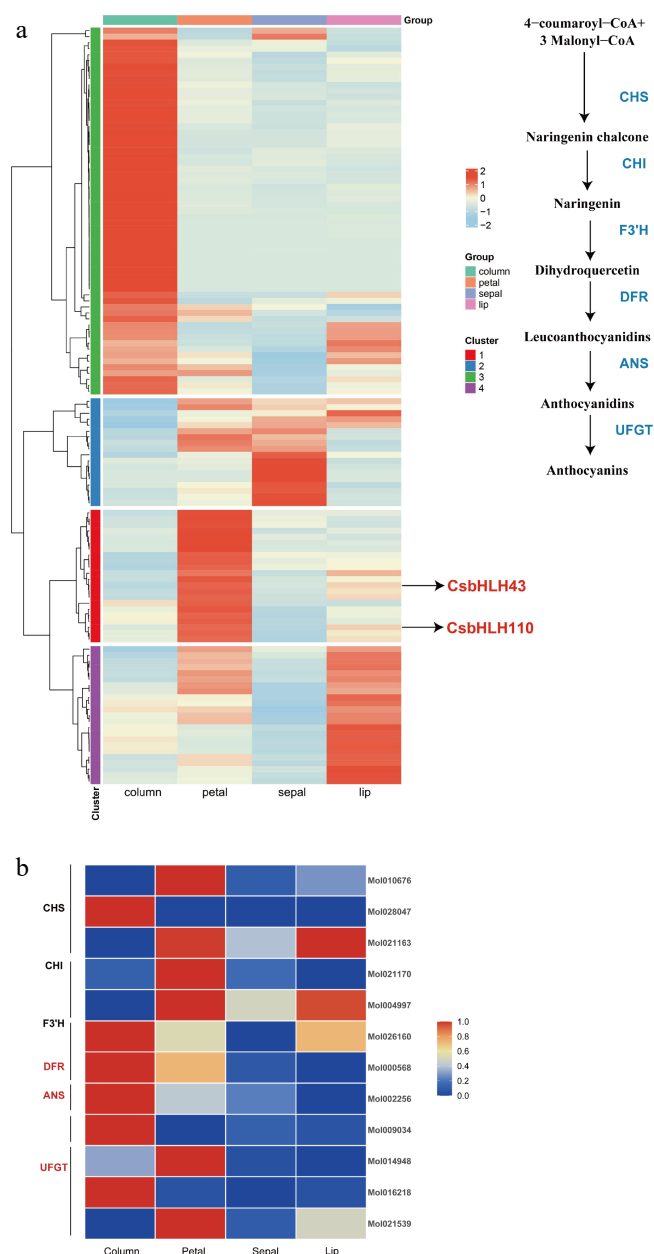


Fig. 7 Expression analysis of bHLH transcription factors and representative flavonoid biosynthesis-related genes. (a) Expression heatmap of bHLH transcription factors across different floral organs. (b) Expression of flavonoid biosynthesis genes.

highlight their roles in DNA binding and protein dimerization. Variations in exon–intron organization may contribute to functional divergence among *CsbHLH* members.

Cis-element analysis revealed that the promoters of *CsbHLHs* contain abundant hormone-responsive and stress-related elements, suggesting their involvement in complex regulatory networks. The presence of MYB-related cis-elements suggests that some *CsbHLH* genes may be connected with MYB-mediated regulatory pathways. In other species, bHLH proteins participate in anthocyanin regulation through MBW complexes^[40]. The phylogenetic placement of *CsbHLH43* and *CsbHLH110* within subgroup IIIf provides a basis for future examination of their potential interactions with MYB and WD40 proteins. In maize, petunia, and grape, bHLH proteins have been shown to regulate anthocyanin biosynthesis through

interactions with MYB and/or WD40 partners, thereby activating structural genes such as *DFR*, *ANS*, and *UFGT*^[41–43]. These studies provide useful references for understanding bHLH-mediated anthocyanin regulation in plants. However, whether *CsbHLH43* and *CsbHLH110* interact with MYB or WD40 proteins or participate in an MBW complex in *C. sinense* remains to be experimentally validated.

In the expression analysis, *CsbHLH43* and *CsbHLH110* showed organ-dependent expression patterns that partially overlapped with those of representative anthocyanin biosynthesis-related genes, including *UFGT*, *CHI*, and *F3'H*^[44].

Interestingly, although anthocyanin content was highest in sepals, the expression levels of these genes were not strictly proportional to pigment accumulation. This discrepancy suggests that anthocyanin biosynthesis is regulated by complex mechanisms beyond transcriptional control, such as temporal gene expression, metabolite transport, and coordinated regulation by multiple transcription factors. Developmental profiling further distinguished the evidence supporting the two candidates. *CsbHLH43* transcript abundance showed a significant positive association with total anthocyanin content across the pooled floral samples, providing developmental correlative support for its association with pigment accumulation. However, the clear clustering of samples according to floral organ and developmental stage indicates that this overall association may partly reflect differences among tissues and stages and should not be interpreted as evidence of direct regulation. In contrast, *CsbHLH110* exhibited a distinct stage-dependent expression pattern, with relatively high transcript abundance at the half-open stage followed by reduced expression at full opening. Its transcript abundance was not significantly associated with total anthocyanin content in the pooled contemporaneous analysis. Because anthocyanin content reflects accumulated pigment, whereas qRT-PCR measures transcript abundance at the sampling time, a simple contemporaneous correlation may not capture all temporal relationships. Nevertheless, the present results do not establish a regulatory role for *CsbHLH110* in anthocyanin biosynthesis.

In addition, subcellular localization analysis showed that *CsbHLH43* and *CsbHLH110* are localized in the nucleus, which is consistent with their predicted roles as transcription factors. Combined with phylogenetic and expression evidence, these results support *CsbHLH43* and *CsbHLH110* as candidate anthocyanin-associated bHLH genes for future functional investigation.

Overall, this study provides a systematic overview of the *CsbHLH* gene family in *C. sinense* and identifies candidate genes potentially associated with floral anthocyanin biosynthesis. Although phylogenetic analysis, expression profiling, qRT-PCR validation, and nuclear localization support *CsbHLH43* and *CsbHLH110* as anthocyanin-associated candidate bHLH genes, these results do not demonstrate direct regulatory functions. Further functional studies are required to clarify the biological roles of *CsbHLH43* and *CsbHLH110* in floral pigmentation.

Conclusions

A total of 136 *CsbHLH* genes were identified and systematically characterized in *C. sinense*. Duplication and Ka/Ks analyses indicated that segmental duplication contributed substantially to the expansion of the gene family and that the duplicated genes were mainly subjected to purifying selection. Phylogenetic, sequence, expression, and subcellular localization analyses prioritized *CsbHLH43* and

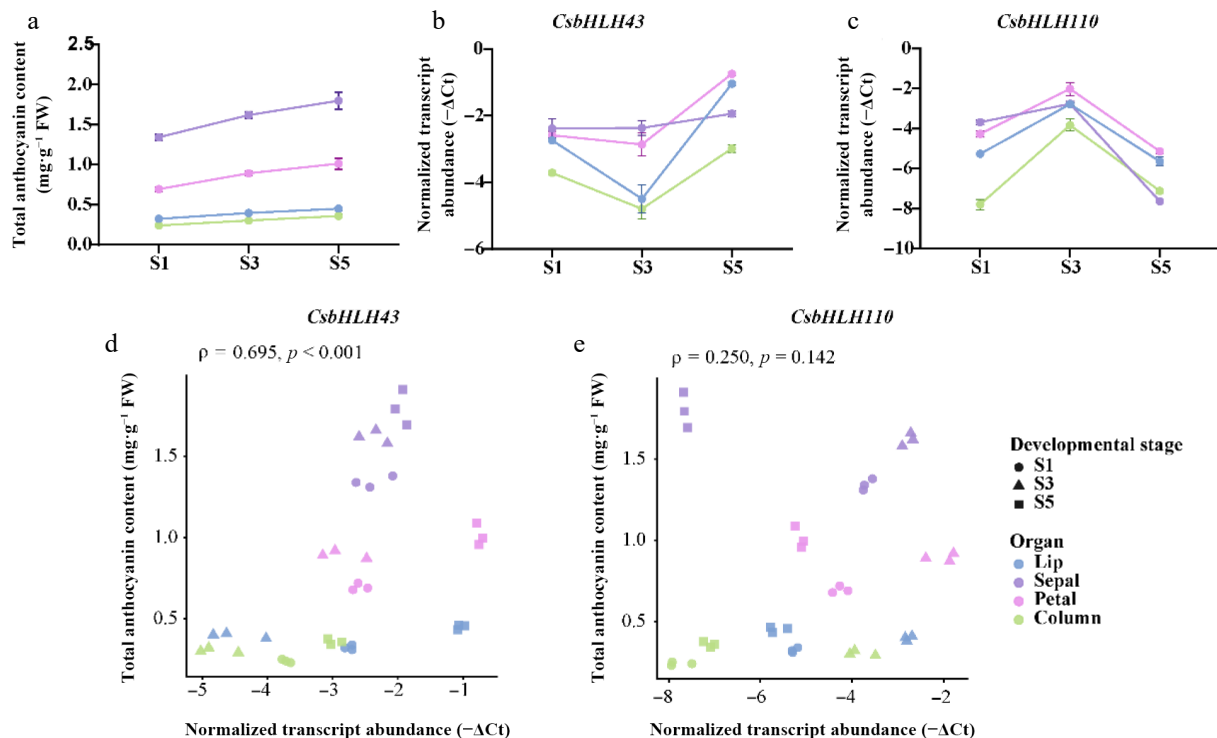


Fig. 8 Developmental changes in anthocyanin content and expression of candidate *CsbHLH* genes. (a) Total anthocyanin content in four floral organs at S1, S3, and S5. Transcript abundance of (b) *CsbHLH43* and (c) *CsbHLH110*, expressed as $-\Delta Ct$ values, with higher values indicating greater transcript abundance. Values in (a)–(c) are means $\pm$ SD of three biological replicates. (d), (e) Spearman correlations between anthocyanin content and transcript abundance across 36 biological samples.

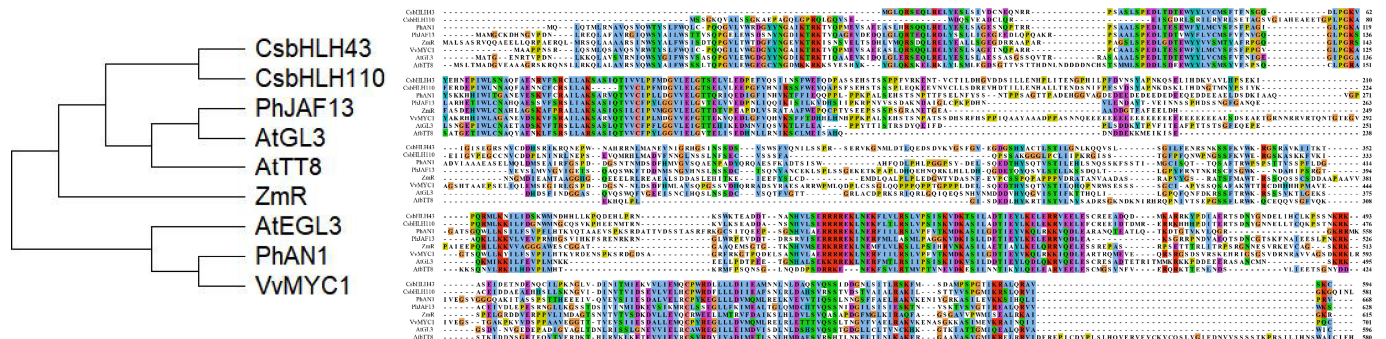


Fig. 9 Phylogenetic analysis and conserved domain alignment of candidate *CsbHLH*s. Multiple sequence alignment of *CsbHLH43* and *CsbHLH110* with bHLH amino acid sequences from other species: At: *A. thaliana* (AtGL3, AT5G41315; AtEGL3, AT1G63650; AtTT8, AT4G09820), Ph: *Petunia hybrida* (PhJAF13, AAC39455.1; PhAN1, AAG25928.1), Zm: *Zea mays* (ZmR, NP_001105339.2), Vv: *Vitis vinifera* (VvMYC1, ACC68685.1).

CsbHLH110 as candidate genes associated with floral anthocyanin accumulation. Developmental expression and correlation analyses provided stronger correlative support for *CsbHLH43*, whereas the association between *CsbHLH110* and anthocyanin accumulation remained unclear. These findings expand our understanding of bHLH genes associated with flower coloration in *C. sinense* and provide candidate genetic resources for future studies of ornamental color traits, germplasm evaluation, and flower-color improvement in *Cymbidium* breeding.

Author contributions

The authors confirm their contributions to this study as follows: conceptualization: Lan S and Liu ZJ; writing of the original draft preparation: Zhao X; formal analysis: Zhao X and Fan X;

methodology: Zhang W, Chen Y, and Wang L; writing, review, and editing: Zhao X and Wang L. All authors reviewed the results and approved the final version of the manuscript.

Data availability

The genomic data of *Cymbidium sinense* are derived from NCBI: PRJNA743748. Additional data supporting the article can be found in the supplementary materials.

Acknowledgments

This work was supported by the Conservation and Utilization of Germplasm Resources and Genetic Improvement (304-7220220306).

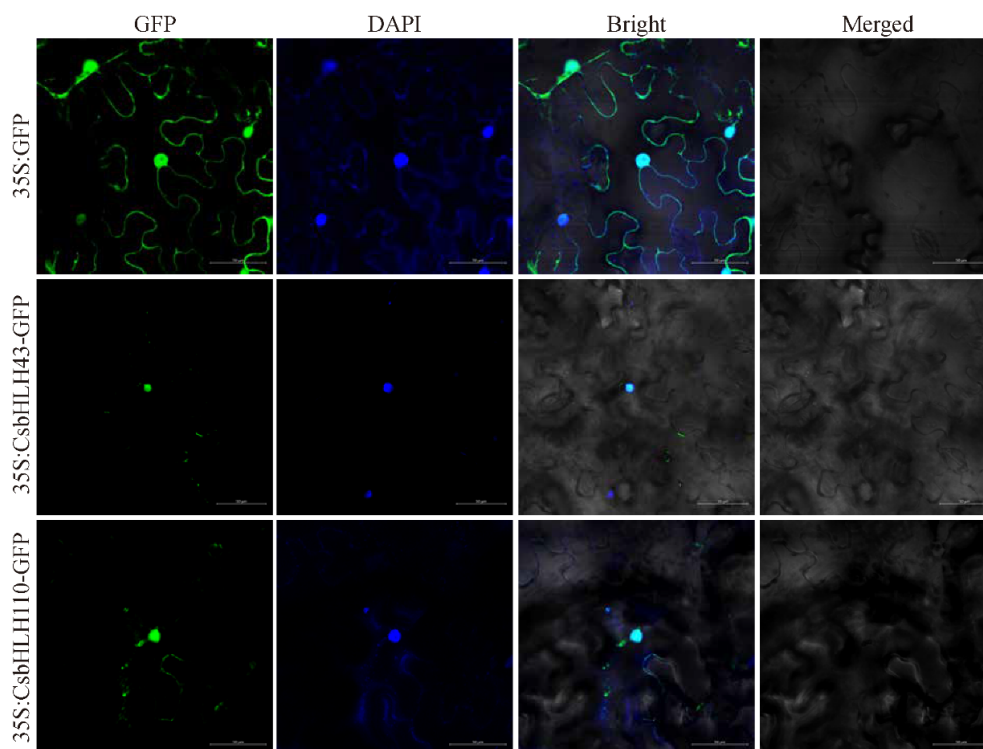


Fig. 10 Subcellular localization of CsbHLH43 and CsbHLH110. CsbHLH43-GFP and CsbHLH110-GFP fusion proteins were transiently expressed in *N. benthamiana* leaves. GFP signals were predominantly detected in the nucleus, as evidenced by their colocalization with DAPI staining. Bright-field images provide cellular context, and merged panels integrate the GFP, DAPI, and bright-field channels. Scale bar = 50 μ m.

Conflict of interest

The authors declare that they have no conflict of interest.

Supplementary information accompanies this paper online at: <https://doi.org/10.48130/opr-0026-0027>.

Dates

Received 10 April 2026; Revised 5 July 2026; Accepted 10 August 2026; Published online 14 October 2026

References

- [1] Wang K, Liu H, Mei Q, Yang J, Ma F, et al. 2023. Characteristics of bHLH transcription factors and their roles in the abiotic stress responses of horticultural crops. *Scientia Horticulturae* 310:111710
- [2] Massari ME, Murre C. 2000. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Molecular and Cellular Biology* 20:429–440
- [3] Zhang C, Jiang L, Qian J, Yu G, Qing H, et al. 2025. Genome-wide analysis of basic helix-loop-helix (bHLH) transcription factors in *petunia* and identification of the putative candidate member involved in floral volatile benzenoids/phenylpropanoids metabolism. *Gene* 938:149150
- [4] 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. *Frontiers in Plant Science* 8:480
- [5] Toledo-Ortiz G, Huq E, Quail PH. 2003. The *Arabidopsis* basic/helix-loop-helix transcription factor family. *The Plant Cell* 15:1749–1770
- [6] Feller A, Machemer K, Braun EL, Grotewold E. 2011. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *The Plant Journal* 66:94–116
- [7] Kavas M, Baloğlu MC, Atabay ES, Ziplar UT, Daşgan HY, et al. 2016. Genome-wide characterization and expression analysis of common bean bHLH transcription factors in response to excess salt concentration. *Molecular Genetics and Genomics* 291:129–143
- [8] Zhao Q, Ren Y, Wang Q, Yao Y, You C, et al. 2016. Overexpression of *MdbHLH104* gene enhances the tolerance to iron deficiency in apple. *Plant Biotechnology Journal* 14:1633–1645
- [9] Huq E, Quail PH. 2002. PIF4, a phytochrome-interacting bHLH factor, functions as a negative regulator of phytochrome B signaling in *Arabidopsis*. *The EMBO Journal* 21:2441–2450
- [10] Jiang Y, Yang B, Deyholos MK. 2009. Functional characterization of the *Arabidopsis* bHLH92 transcription factor in abiotic stress. *Molecular Genetics and Genomics* 282:503–516
- [11] Qi T, Wang J, Huang H, Liu B, Gao H, et al. 2015. Regulation of jasmonate-induced leaf senescence by antagonism between bHLH subgroup IIIe and IIId factors in *Arabidopsis*. *The Plant Cell* 27:1634–1649
- [12] Goossens J, Mertens J, Goossens A. 2017. Role and functioning of bHLH transcription factors in jasmonate signalling. *Journal of Experimental Botany* 68:1333–1347
- [13] Khan RA, Abbas N. 2023. Role of epigenetic and post-translational modifications in anthocyanin biosynthesis: a review. *Gene* 887:147694
- [14] Xie X, Li S, Zhang R, Zhao J, Chen Y, et al. 2012. The bHLH transcription factor *MdbHLH3* promotes anthocyanin accumulation and fruit colouration in response to low temperature in apples. *Plant, Cell & Environment* 35:1884–1897
- [15] Lim SH, Kim DH, Jung JA, Lee JY. 2021. Alternative splicing of the basic helix-loop-helix transcription factor gene *CmbHLH2* affects anthocyanin biosynthesis in ray florets of *Chrysanthemum morifolium*. *Frontiers in Plant Science* 12:669315
- [16] Zhao W, Liu Y, Li L, Meng H, Yang Y, et al. 2021. Genome-wide identification and characterization of bHLH transcription factors related to anthocyanin biosynthesis in red walnut (*Juglans regia* L.). *Frontiers in Genetics* 12:632509

- [17] Lepiniec L, Debeaujon I, Routaboul JM, Baudry A, Pourcel L, et al. 2006. Genetics and biochemistry of seed flavonoids. *Annual Review of Plant Biology* 57:405–430
- [18] Baudry A, Heim MA, Dubreucq B, Caboche M, Weissshaar B, et al. 2004. TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. *The Plant Journal* 39:366–380
- [19] Li M, Sun L, Gu H, Cheng D, Guo X, et al. 2021. Genome-wide characterization and analysis of bHLH transcription factors related to anthocyanin biosynthesis in spine grapes (*Vitis davidii*). *Scientific Reports* 11:6863
- [20] Ji X, Yang F, Zhou X, Jia W, Zhu X, et al. 2025. Genome-wide identification of the bHLH gene family and the mechanism regulation of anthocyanin biosynthesis by ChEGL1 in *Cerasus humilis*. *International Journal of Biological Macromolecules* 288:138783
- [21] Wang R, Mao C, Ming F. 2022. PeMYB4L interacts with PeMYC4 to regulate anthocyanin biosynthesis in *Phalaenopsis* orchid. *Plant Science* 324:111423
- [22] Liu H, Liu Q, Chen Y, Zhu Y, Zhou X, et al. 2023. Full-length transcriptome sequencing provides insights into flavonoid biosynthesis in *Camellia nitidissima* Petals. *Gene* 850:146924
- [23] Chase MW, Cameron KM, Freudenstein JV, Pridgeon AM, Salazar G, et al. 2015. An updated classification of Orchidaceae. *Botanical Journal of the Linnean Society* 177:151–174
- [24] Govaerts R, Campacci MA, Holland A, Pridgeon AM. 2024. *World Checklist of Orchidaceae*. Royal Botanic Gardens, Kew. <http://apps.kew.org/wcsp/>.
- [25] Yang F, Gao J, Wei Y, Ren R, Zhang G, et al. 2021. The genome of *Cymbidium sinense* revealed the evolution of orchid traits. *Plant Biotechnology Journal* 19:2501–2516
- [26] Wang Y, Li S, Zhang Q, Shi Z, Zhai G, et al. 2026. Genome-wide analysis of MYB transcription factors in *Cymbidium* species reveals their roles in anthocyanin synthesis. *Ornamental Plant Research* 6:e013
- [27] Gao F, Dubos C. 2024. The *Arabidopsis* bHLH transcription factor family. *Trends in Plant Science* 29:668–680
- [28] Wu Y, Wu S, Wang X, Mao T, Bao M, et al. 2022. Genome-wide identification and characterization of the bHLH gene family in an ornamental woody plant *Prunus mume*. *Horticultural Plant Journal* 8:531–544
- [29] Chen C, Wu Y, Li J, Wang X, Zeng Z, et al. 2023. TBtools-II: a 'one for all, all for one' bioinformatics platform for biological big-data mining. *Molecular Plant* 16:1733–1742
- [30] Kumar S, Stecher G, Suleski M, Sanderford M, Sharma S, et al. 2024. MEGA12: molecular evolutionary genetic analysis version 12 for adaptive and green computing. *Molecular Biology and Evolution* 41:msae263
- [31] Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189–1191
- [32] Zhu M, Wang Q, Tu S, Ke S, Bi Y, et al. 2023. Genome-wide identification analysis of the R2R3-MYB transcription factor family in *Cymbidium sinense* for insights into drought stress responses. *International Journal of Molecular Sciences* 24:3235
- [33] Lalitha S. 2000. Primer premier 5. *Biotech Software & Internet Report* 1:270–272
- [34] 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
- [35] Carretero-Paulet L, Galstyan A, Roig-Villanova I, Martínez-García JF, Bilbao-Castro JR, et al. 2010. Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in *Arabidopsis*, poplar, rice, moss, and algae. *Plant Physiology* 153:1398–1412
- [36] Wang M, Ou Y, Li Z, Zheng Q, Ke Y, et al. 2023. Genome-wide identification and analysis of bHLH transcription factors related to anthocyanin biosynthesis in *Cymbidium ensifolium*. *International Journal of Molecular Sciences* 24:3825
- [37] Zhai X, Wang X, Yang X, Huang Q, Wu D, et al. 2024. Genome-wide identification of bHLH transcription factors and expression analysis under drought stress in *Pseudoroegneria libanotica* at germination. *Physiology and Molecular Biology of Plants* 30:467–481
- [38] Yang T, Zhu R, Li J, Wang Y, Zhou L, et al. 2026. A bHLH transcription factor negatively regulates effective panicle number and grain yield by modulating auxin transport and distribution in rice. *Molecular Plant* 19:278–294
- [39] Shen J, Su T, Yu D, Tan S, Zhang Y. 2025. Evolution by gene duplication: in the era of genomics. *Yi Chuan* 47:147–171
- [40] Gong Y, Chen X, Qian Z, Pan P, Sun Z, et al. 2026. Evolutionary and functional divergence of MYB regulatory networks in flavonoid biosynthesis across Poaceae, non-Poaceae monocots, and eudicot horticultural plants. *Horticulture Research* 13:uhag169
- [41] 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. *Proceedings of the National Academy of Sciences of the United States of America* 86:7092–7096
- [42] Albert NW, Davies KM, Lewis DH, Zhang H, Montefiori M, et al. 2014. A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. *The Plant Cell* 26:962–980
- [43] Hichri I, Heppel SC, Pillet J, Léon C, Czemmel S, et al. 2010. The basic helix-loop-helix transcription factor MYC1 is involved in the regulation of the flavonoid biosynthesis pathway in grapevine. *Molecular Plant* 3:509–523
- [44] Zhao K, Zhang Q, Wang Y, Wei Q, Wang Y. 2025. Advances in the biosynthesis and regulatory mechanisms of anthocyanins in horticultural plants: a comprehensive review. *Tropical Plants* 4:e010



Copyright: © 2026 by the author(s). Published by Maximum Academic Press, Fayetteville, GA. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit <https://creativecommons.org/licenses/by/4.0/>.