



Article

# Genome-Wide Identification and Analysis of bHLH Transcription Factors Related to Anthocyanin Biosynthesis in *Cymbidium ensifolium*

Meng-Jie Wang<sup>1</sup>, Yue Ou<sup>1</sup>, Zuo Li<sup>2</sup>, Qing-Dong Zheng<sup>1</sup>, Yu-Jie Ke<sup>1</sup>, Hui-Ping Lai<sup>1</sup>, Si-Ren Lan<sup>1</sup>, Dong-Hui Peng<sup>1</sup>, Zhong-Jian Liu<sup>1</sup>  and Ye Ai<sup>1,\*</sup>

<sup>1</sup> 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

<sup>2</sup> Guangdong Key Laboratory of Ornamental Plant Germplasm Innovation and Utilization, Environmental Horticulture Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China

\* Correspondence: aiye@fafu.edu.cn

**Abstract:** The basic helix-loop-helix (bHLH) transcription factors are widely distributed across eukaryotic kingdoms and participate in various physiological processes. To date, the bHLH family has been identified and functionally analyzed in many plants. However, systematic identification of bHLH transcription factors has yet to be reported in orchids. Here, 94 bHLH transcription factors were identified from the *Cymbidium ensifolium* genome and divided into 18 subfamilies. Most *CebHLHs* contain numerous *cis-acting* elements associated with abiotic stress responses and phytohormone responses. A total of 19 pairs of duplicated genes were found in the *CebHLHs*, of which 13 pairs were segmentally duplicated genes and six pairs were tandemly duplicated genes. Expression pattern analysis based on transcriptome data revealed that 84 *CebHLHs* were differentially expressed in four different color sepals, especially *CebHLH13* and *CebHLH75* of the *S7* subfamily. The expression profiles of *CebHLH13* and *CebHLH75* in sepals, which are considered potential genes regulating anthocyanin biosynthesis, were confirmed through the qRT-PCR technique. Furthermore, subcellular localization results showed that *CebHLH13* and *CebHLH75* were located in the nucleus. This research lays a foundation for further exploration of the mechanism of *CebHLHs* in flower color formation.

**Keywords:** *Cymbidium ensifolium*; bHLH transcription factor; anthocyanin biosynthesis; expression analysis



**Citation:** Wang, M.-J.; Ou, Y.; Li, Z.; Zheng, Q.-D.; Ke, Y.-J.; Lai, H.-P.; Lan, S.-R.; Peng, D.-H.; Liu, Z.-J.; Ai, Y. Genome-Wide Identification and Analysis of bHLH Transcription Factors Related to Anthocyanin Biosynthesis in *Cymbidium ensifolium*.

*Int. J. Mol. Sci.* **2023**, *24*, 3825.

<https://doi.org/10.3390/ijms24043825>

ijms24043825

Academic Editor: Manzhou Bao

Received: 17 January 2023

Revised: 5 February 2023

Accepted: 9 February 2023

Published: 14 February 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Transcription factors are a type of regulatory proteins that are mainly combined with the measuring component on gene promoters to activate or repress the expression of downstream target genes [1,2]. bHLH transcription factors are widely distributed in eukaryotes and are named for their highly conserved basic helix-loop-helix (bHLH) domain [3]. The bHLH domain consists of  $\approx 60$  amino acids with two functionally distinct regions, a basic region and a helix-loop-helix (HLH) region [4,5]. The basic region is predicted to recognize and bind DNA located at the N-terminus and contains 13–17 amino acids [2,6], typically with a highly conserved His5-Glu9-Arg13 motif [2,7,8]. The HLH region comprises two  $\alpha$ -helices linked by a loop of variable length, located at the C-terminus, and consists of  $\approx 40$  amino acids. The bHLH proteins containing the HLH region are generally able to form homodimers or heterodimers with other proteins [9,10], which is a prerequisite for DNA recognition and DNA-binding specificity and is also an important factor affecting the expression pattern of bHLH transcription factors [2,7].

Numerous studies have shown that bHLH transcription factors play essential roles in the physiological process of plants, such as plant growth and development [11–14], abiotic stress response [15–18], and anthocyanin biosynthesis [19–21]. Lc was the first

bHLH protein identified in plants, and it was shown to regulate anthocyanin biosynthesis in maize [22]. In orchids, the role of bHLH proteins in anthocyanin biosynthesis has been continuously explored. *DhbHLH1* interacts with *DhMYB2* to regulate the synthesis of anthocyanins in the petals of *Dendrobium* hybrids, while *DhbHLH1* can also independently regulate the synthesis of anthocyanins in the lips [23]. In *Paphiopedilum hirsutissimum*, *bHLH14* and *bHLH106* were found to be involved in the formation of flower color [24]. *DcTT8* regulates anthocyanin biosynthesis in the stems of *D. candidum*, which causes them to exhibit a red phenotype [25]. Three anthocyanin biosynthesis-regulating genes, *CybHLH1*, *CybHLH2*, and *CyMYB1*, were identified in *Cymbidium* 'Mystique' [26]. In addition, bHLH transcription factors can also interact with MYB and WD40 transcription factors to form an MBW complex to regulate the biosynthesis of anthocyanins. For example, simultaneous overexpression of *Lc* (a bHLH transcription factor) and *C1* (a MYB transcription factor) resulted in scarlet spots on the white petals of *C. hybrid* 'Jung Frau dos Pueblos' [27]. It was also found that *PlbHLH20*, *PlbHLH26*, *PIMYB10*, *PIWD40-1*, and their MBW protein complex in *Pleione limprichtii* may be involved in anthocyanin biosynthesis [28].

*Cymbidium ensifolium* is a traditional Chinese orchid with a long history of cultivation and high ornamental value [29]. Flower color is one of the main ornamental characteristics of *C. ensifolium*, which contains purple, dark red, yellow, green, white, and so on. However, the molecular mechanism of its formation remains unclear. Here, a total of 94 CebHLH transcription factors were identified, distributed in 18 subfamilies. They were named according to their location on the chromosomes, and a series of analysis of their characteristics, phylogenetic relationships, gene structure, conserved motifs, *cis-acting* elements, and collinearity analysis were conducted. In addition, the expression pattern of *CebHLHs* in four different color sepals were analyzed with transcriptome sequencing and qRT-PCR, screening out the potential genes *CebHLH13* and *CebHLH75* potentially involved in anthocyanin biosynthesis in *C. ensifolium*. This study lays a theoretical foundation for further exploring the role of *CebHLHs* in flower color formation.

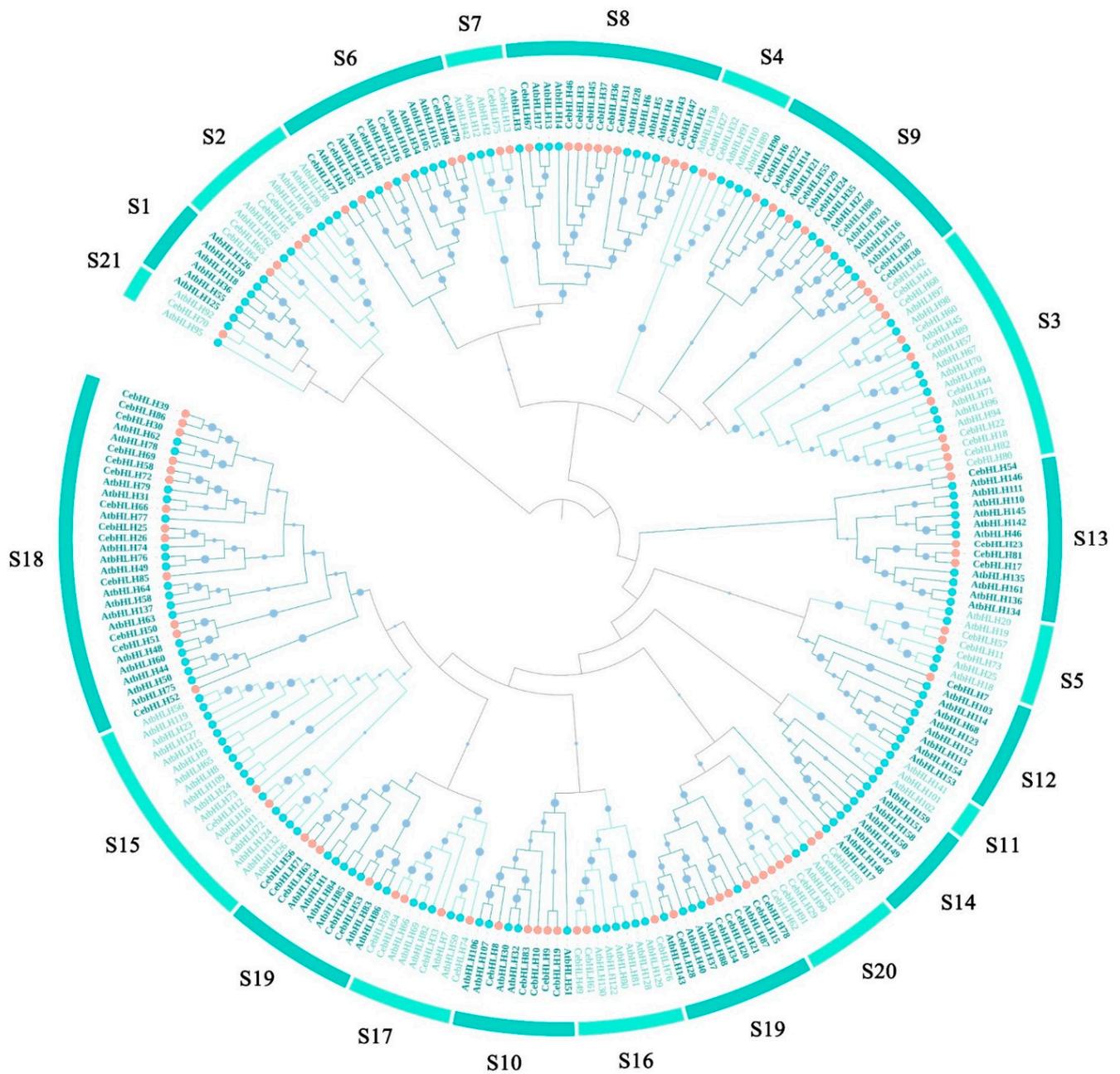
## 2. Results

### 2.1. Identification of CebHLH Transcription Factors and Analysis of Their Physicochemical Properties

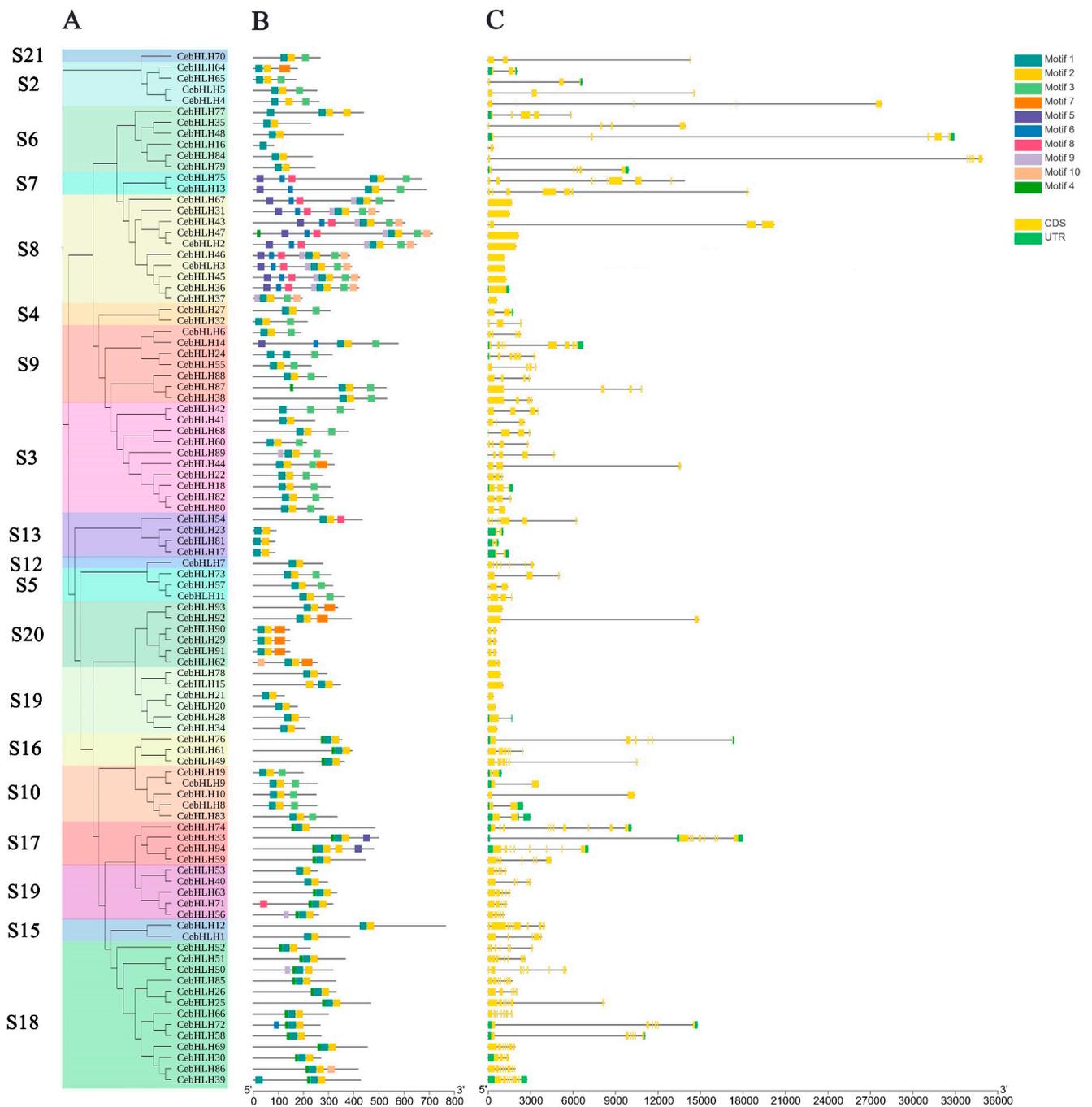
A total of 94 CebHLHs were identified from the *C. ensifolium* genome and named CebHLH1–CebHLH94 on the basis of their chromosomal locations. To further analyze the features of the 94 CebHLHs, we predicted their physicochemical properties. As shown in Table S1, the amino acid lengths of CebHLHs ranged from 79 aa (CebHLH16) to 764 aa (CebHLH12), with an average length of 331 aa. The molecular weights (MW) of CebHLHs varied from 9.10 kDa (CebHLH16) to 83.82 kDa (CebHLH12), with an average MW of 36.55 kDa. The isoelectric point (pI) of these proteins ranged from 4.79 (CebHLH90) to 10.72 (CebHLH21). Furthermore, subcellular localization prediction revealed that most CebHLHs were located in the nucleus.

### 2.2. Phylogenetic Analysis of CebHLHs

To explore the function and evolutionary relationship of CebHLHs, a phylogenetic tree was constructed on the basis of the 94 CebHLH proteins and 152 AtbHLH proteins using the neighbor-joining (NJ) method (Figure 1). The results showed that all members were divided into 21 subfamilies, but CebHLH family members were distributed in 18 subfamilies (Figure 2A), which were absent in the S1, S11, and S14 subfamilies. Moreover, the S18 subfamily had the largest number of CebHLHs (13 members), followed by the S10 (10 members) and S19 subfamily (10 members), and the S12 and S21 subfamilies had the least number (with only one member). Notably, AtbHLH2, AtbHLH42, and AtbHLH12 in the S7 subfamily were previously found to be involved in the regulation of anthocyanin biosynthesis [30], and thus we speculated that CebHLH13 and CebHLH75 in the S7 subfamily may have similar functions and should be analyzed in the subsequent analysis.



**Figure 1.** Phylogenetic tree of bHLH proteins based on 94 CebHLH proteins and 152 AtbHLH proteins. Orange circles represent CebHLHs, blue circles represent AtbHLHs. S1–S21 represent 21 subfamilies.

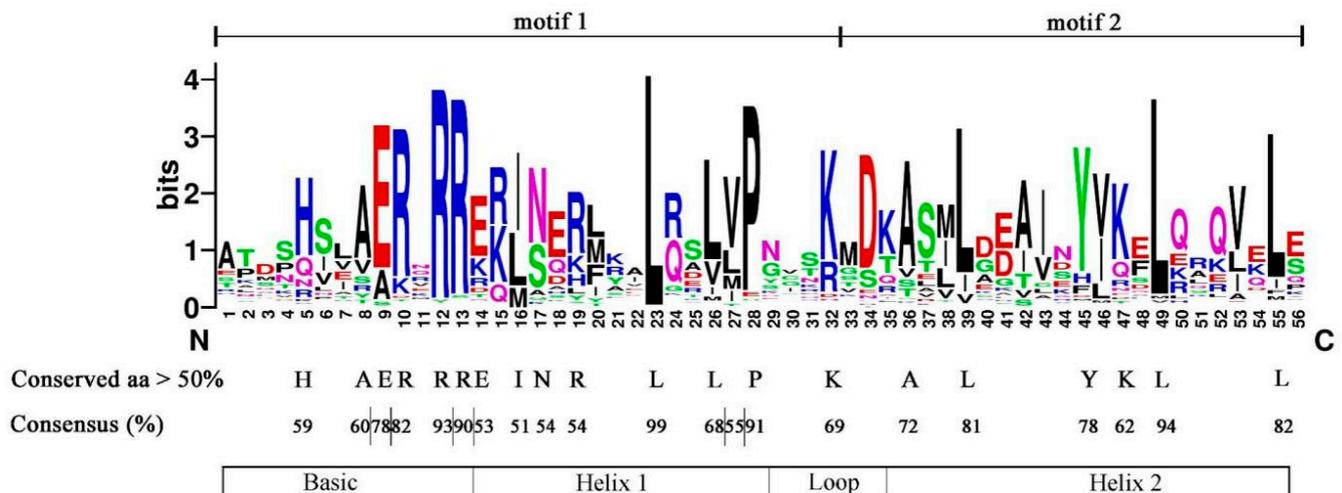


**Figure 2.** Gene structure and conserved motifs of CebHLHs. (A) The phylogenetic tree containing 94 CebHLHs. (B) Conserved motifs of CebHLHs represented by squares of different colors. (C) Intron–exon structure of *CebHLHs*. The x-axis in B is the number of amino acids (aa), and the x-axis in C is the number of base pairs (bp).

### 2.3. Gene Structure and Motif Analysis of CebHLHs

Ten conserved motifs of CebHLHs were identified by the MEME program [31] (Figure 2B). Most CebHLHs contained three motifs, and members of the S7 and S8 subfamilies contain the highest number of motifs (8–9), whereas CebHLH16 contains only one. In addition, CebHLHs members in the same subfamily usually contain similar motifs, and some motifs only exist in the individual subfamily, which may be related to the evolution of gene functional diversity.

To understand the characteristics of the conserved domains of CebHLHs, the multiple sequence alignment results of 94 CebHLHs were uploaded to Weblogo [32]. As shown in Figure 3, the conserved bHLH domains constituted by motif 1 and motif 2 were highly conserved in the sequences of CebHLHs and contained 56 amino acids. The consensus of amino acid residues at 22 positions was higher than 50%, and the consensus of Arg-12, Leu-23, Pro-28, and Leu-39 was higher than 90%.



**Figure 3.** The bHLH domain of CebHLHs, consisting of motif 1 and motif 2.

In order to further analyze the gene structure of *CebHLHs*, an intron–exon structure map of 94 *CebHLHs* was obtained (Figure 2C). There were significant differences in the number of introns or exons among *CebHLHs*. Among them, *CebHLH74* had the largest number of introns, with a total of nine introns. A total of 73.4% of *CebHLHs* contained 1–4 introns, while *CebHLH2*, *CebHLH3*, *CebHLH15*, *CebHLH20*, *CebHLH21*, *CebHLH31*, *CebHLH34*, *CebHLH36*, *CebHLH37*, *CebHLH45*, *CebHLH46*, *CebHLH47*, *CebHLH67*, *CebHLH78*, and *CebHLH93* did not contain introns, accounting for 16% of the total.

#### 2.4. Promoter Analysis of *CebHLHs*

The promoter regions of *CebHLHs* contained a large number of *cis-acting* elements (Figure 4A), including light-responsive elements (GT1-motif), MeJA-responsive elements (TGACG motif and CGTCA motif), auxin regulatory elements (TGA-element), and abscisic acid regulatory elements (ABRE), among which light-responsive elements had the largest number (365), followed by MeJA response elements (310). All *cis-acting* elements were classified into three major categories (Figure 4B), namely, plant growth and development, abiotic stress responses, and phytohormone responses. Among these three categories, the abiotic stress response category had the largest number, followed by the phytohormone regulation category, while the number related to plant growth and development was relatively small. Notably, there were *cis-acting* elements (MBSI) in *CebHLH48* and *CebHLH54* that can bind to MYB and regulate flavonoid biosynthesis, suggesting that these two genes may be involved in flavonoid biosynthesis. The results of promoter analysis suggested that there were differences in the transcriptional regulation of *CebHLHs*, which may be related to their functional diversity.

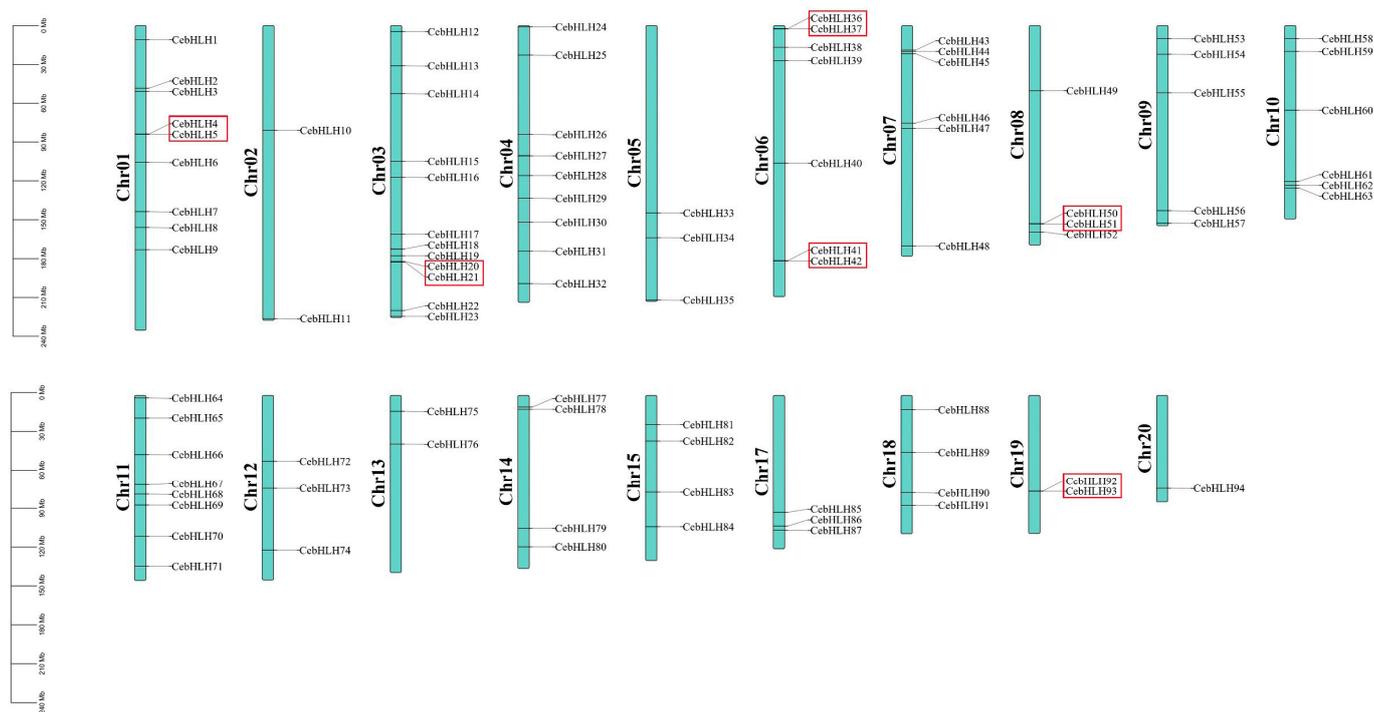


### 2.5. Chromosomal Localization and Collinearity Analysis of *CebHLHs*

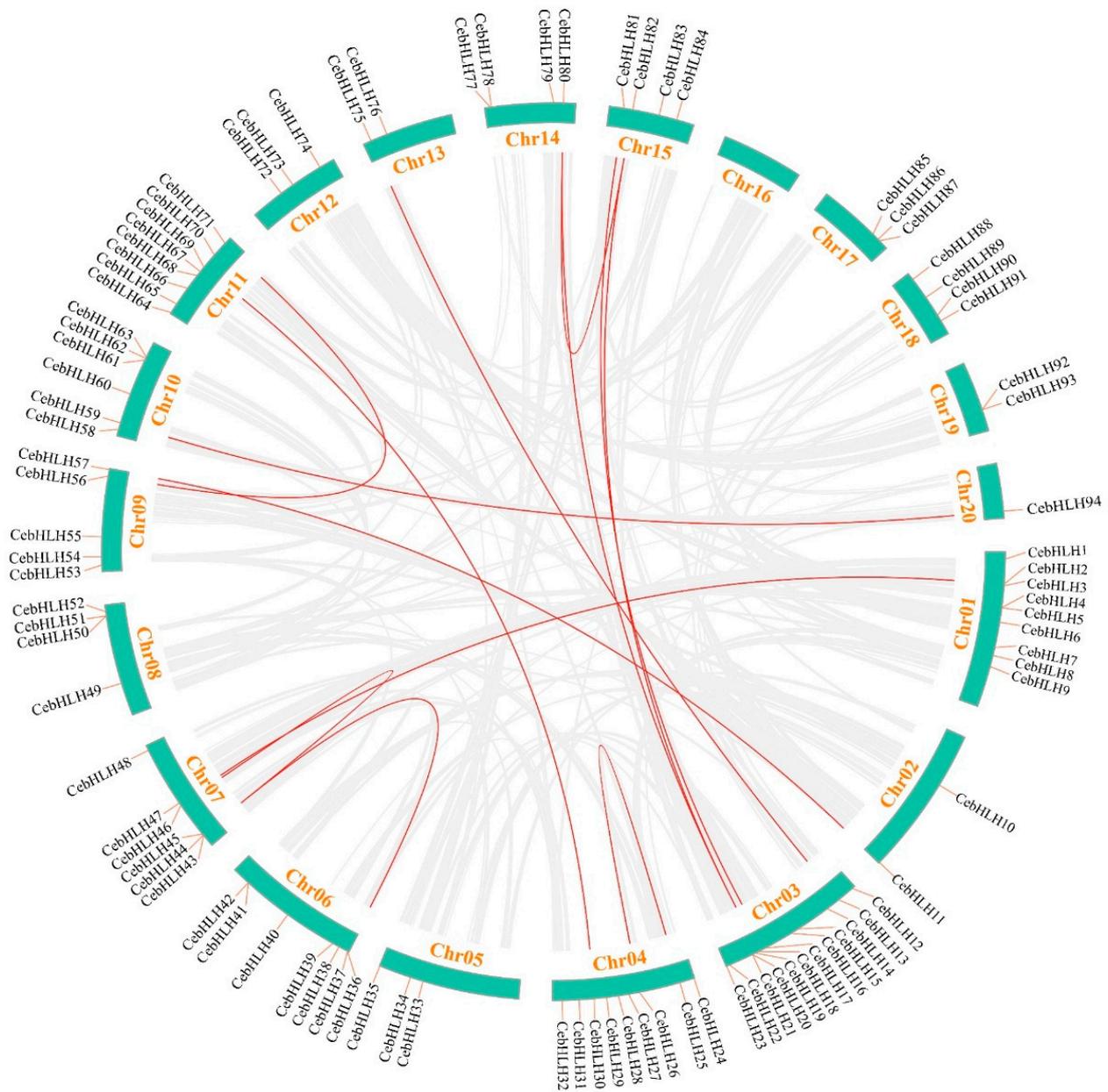
On the basis of the annotation information of the *C. ensifolium* genome, 94 *CebHLHs* were unevenly distributed on 19 chromosomes. Chromosome 3 contained the largest number of *CebHLHs* (12), followed by chromosomes 1 and 4 (9 each), while chromosome 20 had only one.

Gene duplication events are crucial evolutionary processes leading to gene structural and functional divergence [33]. A total of 19 pairs of duplicated genes were found in the *CebHLH* family, of which 6 pairs were tandemly duplicated genes (Figure 5), and 13 pairs were segmentally duplicated genes (Figure 6). Among the segmentally duplicated genes, *CebHLH25* and *CebHLH26* are located on chromosome 4, and *CebHLH45* and *CebHLH46* are located on chromosome 7. The largest number of tandemly duplicated genes was found on chromosomes 3 and 7, with three pairs, and only one was found on chromosomes 1, 6, 10, 13, and 20. Among the tandemly duplicated genes, *CebHLH20* and *CebHLH21*, *CebHLH4* and *CebHLH5*, *CebHLH50* and *CebHLH51*, and *CebHLH92* and *CebHLH93* share similar conserved motifs and gene structures.

The nonsynonymous/synonymous mutation ( $Ka/Ks$ ) ratio is essential for exploring genomic evolution [34], which can show purifying selection ( $Ka/Ks < 1$ ), neutral mutation ( $Ka/Ks = 1$ ), and positive selection ( $Ka/Ks > 1$ ). As shown in Table S2, the  $Ka/Ks$  ratios of 16 pairs of genes were between 0.11 and 0.54, indicating that these genes underwent strong purifying selection during evolution [35,36]. Notably, the  $Ka/Ks$  ratios of *CebHLH20* and *CebHLH21*, and *CebHLH4* and *CebHLH5* were greater than 1, suggesting that these two pairs of genes underwent positive selection [35,36].



**Figure 5.** Chromosomal localization and gene duplications of *CebHLHs*. The tandemly duplicated genes are represented by red boxes. The scale bars on the left are the length (Mb) of the chromosomes of *C. ensifolium*.

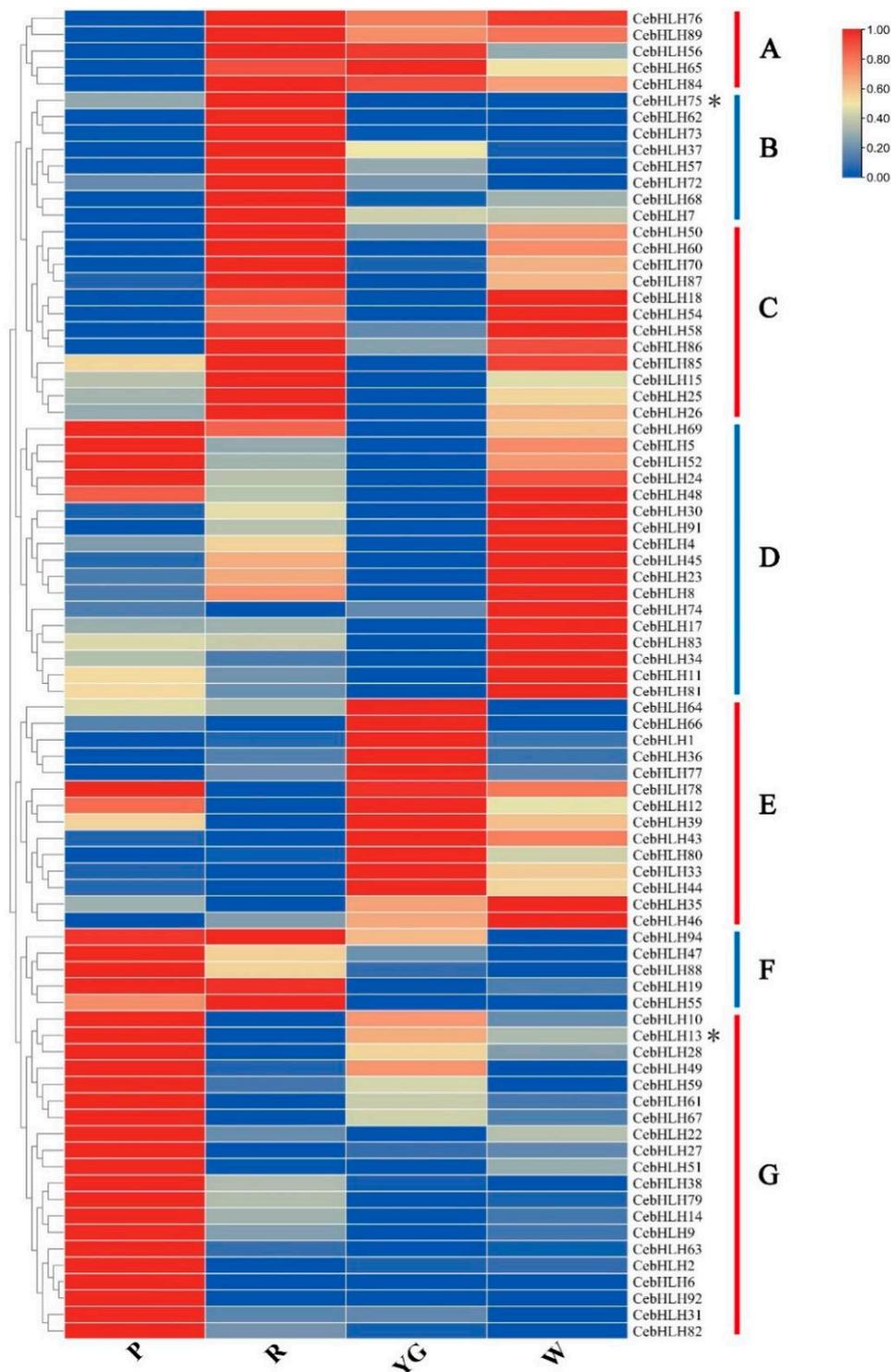


**Figure 6.** Collinearity analysis of *CebHLHs*. The duplicated gene pairs in the genome were linked by red lines.

### 2.6. Expression Pattern of *CebHLHs* in Four Different Color Sepals

To understand the expression patterns of *CebHLHs* and screen out potential genes that may be related to anthocyanin biosynthesis, we performed transcriptome sequencing using the sepals of four different colors of *C. ensifolium* (Table S3). As shown in Figure 7, 82 genes had different expression levels in sepals of different colors and were divided into seven groups (A–G). The five genes in group A were lowly expressed in purple-red sepals and highly expressed in other sepals. Eight genes in group B were highly expressed in red sepals, among which *CebHLH75* was also expressed in purple-red sepals, but not expressed in yellow-green and white sepals, indicating that *CebHLH75* may regulate the biosynthesis of anthocyanins. Most genes in groups C and D were highly expressed in white sepals, and 14 genes in group E were mainly expressed in yellow-green sepals. In group F, five genes were highly expressed in purple-red and red sepals, and lower in yellow-green and white sepals. All genes in group G were highly expressed in purple-red sepals and lower in the

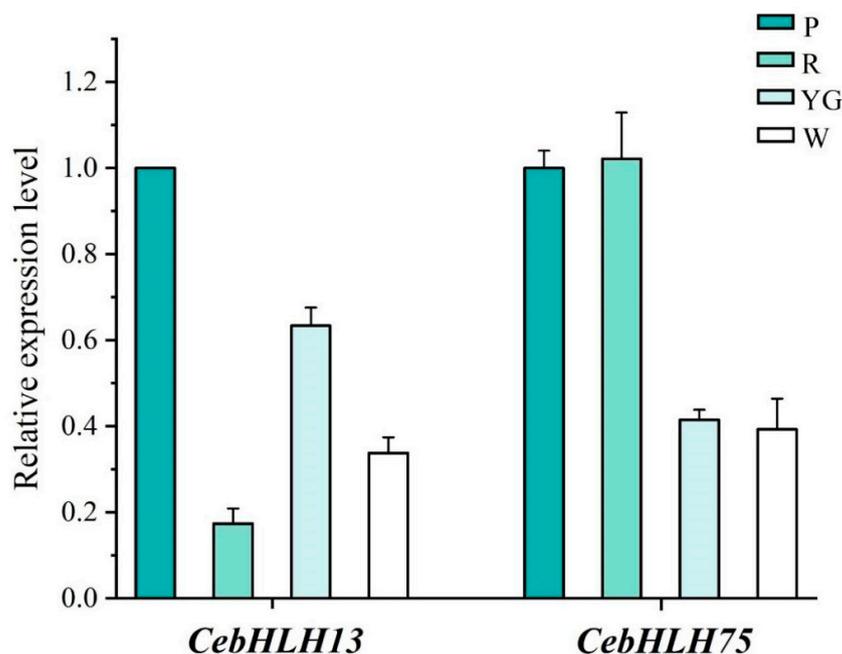
other sepals. Notably, *CebHLH75* in group B and *CebHLH13* in group G, which belonged to the *S7* subfamily, were highly expressed in red and purple-red sepals. Therefore, we speculated that *CebHLH75* and *CebHLH13* play a positive role in anthocyanin biosynthesis.



**Figure 7.** Expression patterns of 82 *CebHLHs* were divided into seven groups (A–G). P stands for purple-red sepals; R stands for red sepals; YG stands for yellow-green sepals; W stands for white sepals. \* denotes *CebHLH13* and *CebHLH75* of the *S7* subfamily.

### 2.7. qRT-PCR Analysis of *CebHLHs*

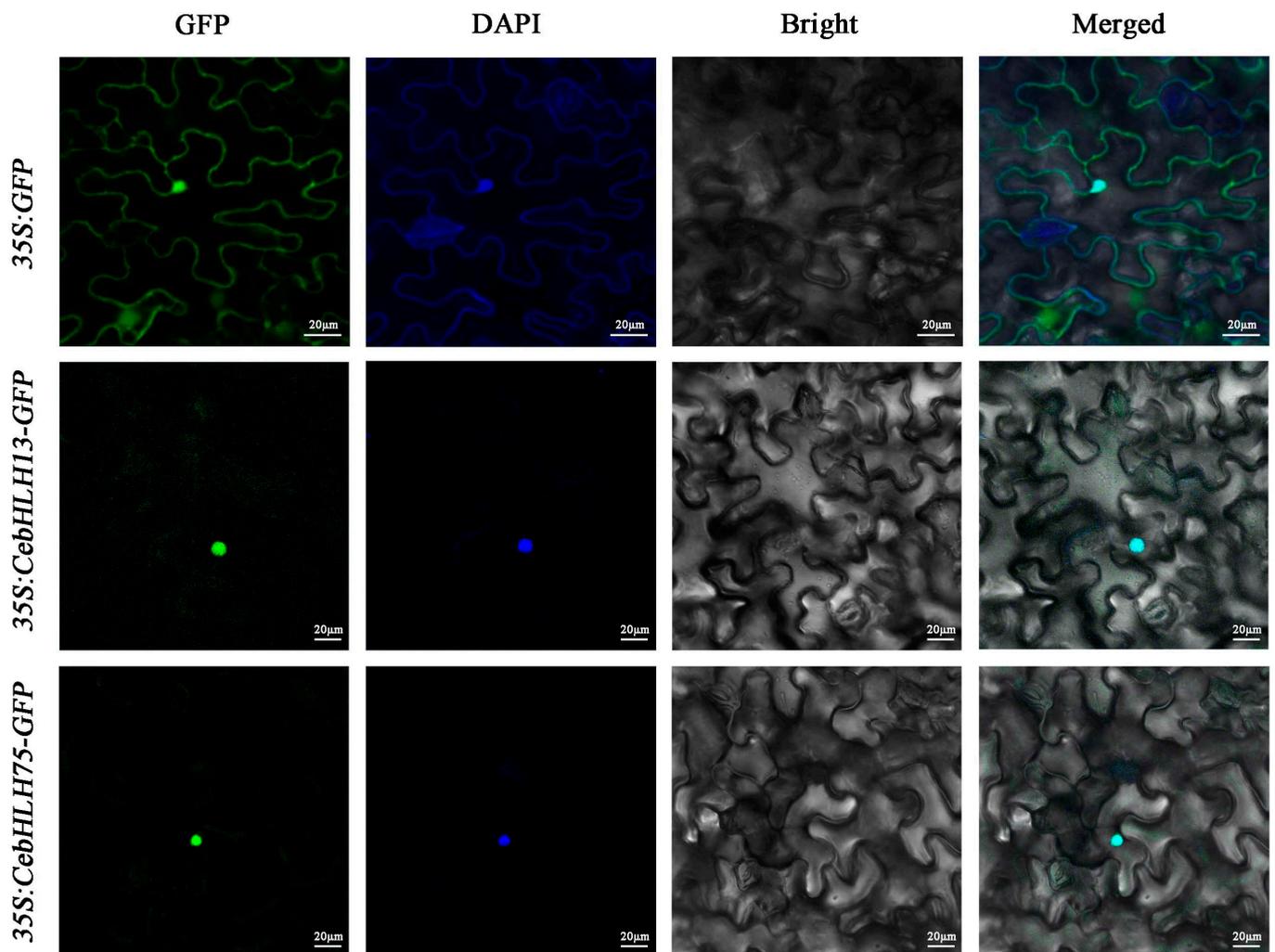
qRT-PCR was used to further verify the expression patterns of *CebHLH13* and *CebHLH75* in four colored sepals (Figure 8). The results showed that the expression of *CebHLH13* was highest in purple-red sepals. *CebHLH75* had a high expression in purple-red and red sepals and was lowly expressed in yellow-green and white sepals. The expression levels of *CebHLH13* and *CebHLH75* in the four colored sepals obtained by qRT-PCR were consistent with the transcriptome data, supporting the accuracy of the transcriptome sequencing results.



**Figure 8.** qRT-PCR Analysis of *CebHLH13* and *CebHLH75*. P stands for purple-red sepals; R stands for red sepals; YG stands for yellow-green sepals; W stands for white sepals.

### 2.8. Subcellular Localization of *CebHLH13* and *CebHLH75*

To explore the subcellular localization of *CebHLH13* and *CebHLH75* proteins, two recombinant vectors (*35S: CebHLH13-GFP* and *35S: CebHLH75-GFP*) and the control vector (*35S: GFP*) were separately introduced into *Nicotiana benthamiana* leaves. The subcellular localization of these two proteins was observed after 48 h. As shown in Figure 9, the GFP signals of *35S: CebHLH13-GFP* and *35S: CebHLH75-GFP* were significantly detected in the nucleus, demonstrating that *CebHLH13* and *CebHLH75* were localized in the nucleus.



**Figure 9.** Subcellular localization analysis of CebHLH13 and CebHLH75 proteins.

### 3. Discussion

The bHLH transcription factors are one of the largest transcription factor families in plants and play important roles in various physiological processes. However, the systematic identification of the bHLH transcription factor family in orchids has yet to be reported. Therefore, it is of great theoretical and practical significance to systematically classify and functionally study bHLH transcription factors of *C. ensifolium*. In this study, we identified 94 CebHLHs from the genome of *C. ensifolium*. The number was the same as *Vitis vinifera* (94 members) [37], less than *Arabidopsis thaliana* (162 members) [38], *Daucus carota* (146 members) [8], *Triticum aestivum* (225 members) [39], and *Solanum tuberosum* (124 members) [40], and more than *Paeonia suffruticosa* (84 members) [41] and *Citrus sinensis* (56 members) [42]. Differences in the number of bHLH family members are normal due to gene duplication, deletion, and functional diversification [43].

A total of 94 *C. ensifolium* bHLH proteins and 152 *A. thaliana* bHLH proteins were combined to construct a phylogenetic tree. According to the previous classification of *A. thaliana* [38], all members were divided into 21 subfamilies, of which the S1, S11, and S14 subfamilies did not contain CebHLHs, indicating that CebHLHs in these three subfamilies may have been differentiated or lost in the long-term evolutionary process. Proteins with similar functions tend to cluster in the same subfamily. In previous studies, AtbHLH2, AtbHLH42, and AtbHLH12 of the S7 subfamily were found to promote anthocyanin biosynthesis [30]. Therefore, it is speculated that CebHLH13 and CebHLH75 in the same subfamily also have the same function.

Previous studies have found that the bHLH domain consists of  $\approx 60$  amino acids, and its basic region has at least five highly conserved amino acids and a His5-Glu9-Arg13 conservative structure [2,6–8]; our study of CebHLHs supports this notion. The bHLH transcription factors often exert their biological functions by forming homologous or heterodimers [9,10], and Leu23 residues in the HLH region are crucial for dimer formation [8,30,44]. Sequence analysis found that the conservation of Leu23 of CebHLHs is as high as 99%, indicating that almost all CebHLHs can form homologous or heterodimers, which is of great significance for the biosynthesis of anthocyanins. In addition, gene structure and motif analysis found that the members of the same subfamily have similar gene structures. The specific conserved motifs in some subfamilies also support the reliability of the phylogenetic tree. Except for 15 CebHLHs that do not contain introns, the number of introns in most CebHLHs ranges from 1 to 9. The intron positions and numbers of CebHLHs are polymorphic, which may be the evolution of functional diversity during long-term evolution. Different *cis-acting* elements were observed in the promoter regions of CebHLHs, and the data showed that most of them were associated with abiotic stress responses, indicating that CebHLHs may be regulated by various factors and play an important role in the transcriptional regulation of abiotic stress responses.

Gene duplication events play a central role in the evolutionary process and are important for the generation of new gene members [33]. Our study revealed the gene duplication pattern of the CebHLH family and determined its relative location on the chromosomes. There were 19 pairs of repetitive genes in 94 CebHLHs, of which 13 pairs were segmentally duplicated genes, and 6 pairs were tandemly duplicated genes. In addition, 84.2% of the gene pairs had *Ka/Ks* ratios less than 1, suggesting that most CebHLHs underwent strong purifying selection during evolution [35,36]. However, we found two pairs of genes (CebHLH20 and CebHLH21, CebHLH4 and CebHLH5) with *Ka/Ks* ratios greater than 1, suggesting that these two pairs of genes underwent positive selection [35,36], which is of great significance to the study of species evolution.

Expression pattern analysis revealed that CebHLH13 was highly expressed in the purple-red sepals but very low in the red sepals, which may be related to the type and content of anthocyanins it regulates. CebHLH75 was significantly highly expressed in purple-red and red sepals, and lower in yellow-green and white sepals. Their expression levels were consistent with phenotype observation. Therefore, we speculated that CebHLH13 and CebHLH75 could promote anthocyanin biosynthesis in *C. ensifolium*. Previous research has shown that bHLH transcription factors play critical roles in anthocyanin biosynthesis. In this study, potential CebHLHs related to anthocyanin biosynthesis in *C. ensifolium* were screened out, laying a foundation for further exploring the mechanism of CebHLHs in flower color formation and also providing valuable information for orchid flower color breeding.

## 4. Materials and Methods

### 4.1. Plant Materials

*C. ensifolium* cultivars with purple-red sepals, red sepals, yellow-green sepals, and white sepals were collected from the Orchid Germplasm Resource Nursery of Fujian Agriculture and Forestry University, Fuzhou, Fujian Province, China. All materials were immediately frozen in liquid nitrogen and stored at  $-80\text{ }^{\circ}\text{C}$  for later analysis. For each sample, three replicates were obtained from different plants.

### 4.2. Identification and Sequence Analysis of CebHLHs

The genome sequence and annotation information for *C. ensifolium* was downloaded from the National Genome Data Center (NGDC) (<https://ngdc.cnbc.ac.cn/>, accessed on 25 March 2022). The HMM (Hidden Markov Model) profile of the bHLH domain (PF00010) was obtained from the Pfam database (<http://pfam.xfam.org/search>, accessed on 25 March 2022) for protein screening by TBtools software [45] (E-value  $\leq 10^{-4}$ ). All protein sequences were further confirmed using SMART (<http://smart.emblheidelberg>

de/, accessed on 25 March 2022), and proteins without the bHLH domain were deleted. The physicochemical properties of the protein amino acid sequences of CebHLHs were predicted by ExPASy (<http://www.expasy.org/tools/>, accessed on 25 March 2022) [46], and subcellular localization prediction was performed on WoLF PSORT (<https://wolfpsort.hgc.jp/>, accessed on 25 March 2022) [47]. In addition, the multiple sequence alignment results of 94 CebHLH proteins were submitted to WebLogo (<https://weblogo.berkeley.edu/logo.cgi>, accessed on 25 March 2022) to observe their conserved domain characteristics [32].

#### 4.3. Phylogenetic Analysis of CebHLHs

Multisequence alignment of 94 CebHLH proteins and 152 AtbHLH proteins was performed using MEGA11 and Jalview software [48,49], and the results were uploaded to MEGA11 to construct an unrooted neighbor-joining phylogenetic tree with 1000 bootstraps. Finally, the phylogenetic tree was imported into iTOL (<https://itol.embl.de/itol.cgi>, accessed on 25 March 2022) for modification [50].

#### 4.4. Gene Structure and Motif Analysis of CebHLHs

The intron–exon structure of *CebHLHs* was identified by the Visualize Gene Structure program of TBtools [45]. Conserved motifs of CebHLHs were analyzed by MEME (<https://meme-suite.org/meme/>, accessed on 25 March 2022) [31]. The maximum number of motifs in the parameter settings was 10, while the other parameters were default. The results were then imported into TBtools for visualization [45].

#### 4.5. Promoter Analysis of CebHLHs

The 2000 bp regions upstream of the start codon were extracted and uploaded to TBtools to identify the putative *cis-acting* elements in the promoter region [45]. Data processing was performed with Excel, and then the results were visualized with TBtools [45].

#### 4.6. Chromosomal Localization and Collinearity Analysis of CebHLHs

After obtaining information on the location of the bHLH family on 19 chromosomes from the genome annotation information of *C. ensifolium*, the chromosomal location map was obtained with TBtools [45]. The collinear relationship between the chromosome pairs was drawn and visualized by the One Step MCScanx and Advance Circos program of TBtools [45]. Moreover, the *Ka/Ks* ratios were calculated by TBtools [45].

#### 4.7. Expression Pattern of CebHLHs

Total RNA from four different color sepals was extracted using the OMEGA kit (Norgene, Georgia, USA). RNA-seq and library construction were performed by the Novogene Bioinformatics Co., Ltd. (Beijing, China) on an Illumina HiSeq 2500 platform. Expression levels of CebHLHs were represented by fragments per kilobase of exon model per million mapped reads (FPKM) values. The FPKM values of CebHLHs were imported into TBtools to generate a heat map [45].

#### 4.8. qRT-PCR Analysis of CebHLHs

The PrimerScript<sup>®</sup> RT Reagent Kit with gDNA Eraser (TaKaRa, Dalian, China) was used to reverse transcribe the RNA into cDNA. qRT-PCR was performed with three biological replicates and three technical replicates using Taq Pro Universal SYBR qPCR Master Mix (TaKaRa, Dalian, China) on an Applied Biosystems 7500 Real-Time System (Applied Biosystems, Foster City, CA, USA). All primers used for qRT-PCR are listed in Table S4. *GAPDH* was used as an internal reference for the data. Relative expression was calculated using the  $2^{-\Delta\Delta CT}$  method.

#### 4.9. Subcellular Localization of CebHLH13 and CebHLH75

The full-length ORFs of *CebHLH13* and *CebHLH75* without the termination codon were inserted into the pCAMBIA1302 vector with *NcoI* and *SpeI* restriction sites to create the 35S:

*CebHLH13-GFP* and *35S:CebHLH75-GFP* fusion construct. In addition, the recombinant plasmid and control *35S:GFP* plasmid were transferred into *Agrobacterium* strain GV3101. The *Agrobacterium* containing the target plasmid was resuspended and transiently infected into *Nicotiana benthamiana* leaves. The subcellular localization of *CebHLH13* and *CebHLH75* were observed by LSM710 confocal laser microscopy (CarlZeiss, Jena, Germany) after 48 h. All primers used in this study are listed in Table S5.

## 5. Conclusions

In this study, 94 *CebHLHs* were identified in the genome of *C. ensifolium* and subjected to classification, phylogenetic construction, gene structure analysis, conserved motif characterization, chromosomal localization, and expression pattern analysis. The expression patterns of these genes were specific in four colored sepals of *C. ensifolium*. Two potential genes, *CebHLH13* and *CebHLH75*, which may be related to anthocyanin biosynthesis, were screened. The research provided useful information for the functional analysis of bHLH transcription factors, as well as for flower color improvement and molecular breeding in orchids.

**Supplementary Materials:** The supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/ijms24043825/s1>.

**Author Contributions:** Y.A. designed and directed this research. S.-R.L., D.-H.P. and Z.-J.L. planned and coordinated the project. M.-J.W. performed the bioinformatics analysis and wrote the manuscript. Z.L. collected plant material and helped to design and advise on the research. M.-J.W., Y.O., Q.-D.Z., Y.-J.K. and H.-P.L. conducted transcriptome sequencing and analysis. M.-J.W. and Y.O. performed the qRT-PCR and subcellular localization experiments. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by The National Key Research and Development Program of China (2019YFD1000400), the Fujian Natural Science Foundation Project of China (2020J01585), and the Outstanding Young Scientific Research Talent Project of Fujian Agriculture and Forestry University (xjq201910).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All sequences of *C. ensifolium* in this study can be found at the National Genomics Data Center (NGDC). The transcriptomic data are openly available from the National Center for Biotechnology Information under the accession number PRJNA771426. Additional data supporting the article can be found in the supplementary materials.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Golldack, D.; Lüking, I.; Yang, O. Plant Tolerance to Drought and Salinity: Stress Regulating Transcription Factors and Their Functional Significance in the Cellular Transcriptional Network. *Plant Cell Rep.* **2011**, *30*, 1383–1391. [[CrossRef](#)] [[PubMed](#)]
2. Feller, A.; Machemer, K.; Braun, E.L.; Grotewold, E. Evolutionary and Comparative Analysis of MYB and bHLH Plant Transcription Factors. *Plant J.* **2011**, *66*, 94–116. [[CrossRef](#)] [[PubMed](#)]
3. Massari, M.E.; Murre, C. Helix-Loop-Helix Proteins: Regulators of Transcription in Eucaryotic Organisms. *Mol. Cell. Biol.* **2000**, *20*, 429–440. [[CrossRef](#)]
4. Murre, C.; McCaw, P.S.; Baltimore, D. A New DNA Binding and Dimerization Motif in Immunoglobulin Enhancer Binding, Daughterless, MyoD, and Myc Proteins. *Cell* **1989**, *56*, 777–783. [[CrossRef](#)]
5. Toledo-Ortiz, G.; Huq, E.; Quail, P.H. The Arabidopsis Basic/Helix-Loop-Helix Transcription Factor Family[W]. *Plant Cell* **2003**, *15*, 1749–1770. [[CrossRef](#)]
6. Atchley, W.R.; Fitch, W.M. A Natural Classification of the Basic Helix-Loop-Helix Class of Transcription Factors. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 5172–5176. [[CrossRef](#)]
7. Shimizu, T. Crystal Structure of PHO4 bHLH Domain-DNA Complex: Flanking Base Recognition. *EMBO J.* **1997**, *16*, 4689–4697. [[CrossRef](#)]
8. Chen, Y.Y.; Li, M.Y.; Wu, X.J.; Huang, Y.; Ma, J.; Xiong, A.S. Genome-Wide Analysis of Basic Helix-loop-helix Family Transcription Factors and Their Role in Responses to Abiotic Stress in Carrot. *Mol. Breeding* **2015**, *35*, 125. [[CrossRef](#)]

9. Yin, Y.; Vafeados, D.; Tao, Y.; Yoshida, S.; Asami, T.; Chory, J. A New Class of Transcription Factors Mediates Brassinosteroid-Regulated Gene Expression in Arabidopsis. *Cell* **2005**, *120*, 249–259. [[CrossRef](#)]
10. Pires, N.; Dolan, L. Origin and Diversification of Basic-Helix-Loop-Helix Proteins in Plants. *Mol. Biol. Evol.* **2010**, *27*, 862–874. [[CrossRef](#)]
11. Kanaoka, M.M.; Pillitteri, L.J.; Fujii, H.; Yoshida, Y.; Bogenschutz, N.L.; Takabayashi, J.; Zhu, J.-K.; Torii, K.U. SCREAM/ICE1 and SCREAM2 Specify Three Cell-State Transitional Steps Leading to Arabidopsis Stomatal Differentiation. *Plant Cell* **2008**, *20*, 1775–1785. [[CrossRef](#)] [[PubMed](#)]
12. Menand, B.; Yi, K.; Jouannic, S.; Hoffmann, L.; Ryan, E.; Linstead, P.; Schaefer, D.G.; Dolan, L. An Ancient Mechanism Controls the Development of Cells with a Rooting Function in Land Plants. *Science* **2007**, *316*, 1477–1480. [[CrossRef](#)] [[PubMed](#)]
13. Ito, S.; Song, Y.H.; Josephson-Day, A.R.; Miller, R.J.; Breton, G.; Olmstead, R.G.; Imaizumi, T. FLOWERING BHLH Transcriptional Activators Control Expression of the Photoperiodic Flowering Regulator CONSTANS in Arabidopsis. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 3582–3587. [[CrossRef](#)] [[PubMed](#)]
14. Chen, N.; Li, L.; Li, C.; Lin, Z.; Meng, J.; Liu, S.; Song, K.; Bao, Y. bHLH Genes Polymorphisms and Their Association with Growth Traits in the Pacific Oyster *Crassostrea Gigas*. *J. Ocean. Limnol.* **2020**, *38*, 862–868. [[CrossRef](#)]
15. Huq, E. PIF4, a Phytochrome-Interacting bHLH Factor, Functions as a Negative Regulator of Phytochrome B Signaling in Arabidopsis. *EMBO J.* **2002**, *21*, 2441–2450. [[CrossRef](#)]
16. Jiang, Y.; Yang, B.; Deyholos, M.K. Functional Characterization of the Arabidopsis bHLH92 Transcription Factor in Abiotic Stress. *Mol. Genet. Genom.* **2009**, *282*, 503–516. [[CrossRef](#)]
17. Qi, T.; Wang, J.; Huang, H.; Liu, B.; Gao, H.; Liu, Y.; Song, S.; Xie, D. Regulation of Jasmonate-Induced Leaf Senescence by Antagonism between bHLH Subgroup IIIe and IIIc Factors in Arabidopsis. *Plant Cell* **2015**, *27*, 1634–1649. [[CrossRef](#)] [[PubMed](#)]
18. Zhao, Q.; Ren, Y.R.; Wang, Q.J.; Yao, Y.X.; You, C.X.; Hao, Y.J. Overexpression of *MdbHLH104* Gene Enhances the Tolerance to Iron Deficiency in Apple. *Plant Biotechnol. J.* **2016**, *14*, 1633–1645. [[CrossRef](#)] [[PubMed](#)]
19. Xie, X.B.; Li, S.; Zhang, R.F.; Zhao, J.; Chen, Y.C.; Zhao, Q.; Yao, Y.X.; You, C.X.; Zhang, X.S.; Hao, Y.J. The bHLH Transcription Factor *MdbHLH3* Promotes Anthocyanin Accumulation and Fruit Colouration in Response to Low Temperature in Apples. *Plant Cell Environ.* **2012**, *35*, 1884–1897. [[CrossRef](#)] [[PubMed](#)]
20. Lim, S.H.; Kim, D.H.; Jung, J.A.; Lee, J.Y. Alternative Splicing of the Basic Helix–Loop–Helix Transcription Factor Gene *CmbHLH2* Affects Anthocyanin Biosynthesis in Ray Florets of Chrysanthemum (*Chrysanthemum morifolium*). *Front. Plant Sci.* **2021**, *12*, 669315. [[CrossRef](#)] [[PubMed](#)]
21. Zhao, W.; Liu, Y.; Li, L.; Meng, H.; Yang, Y.; Dong, Z.; Wang, L.; Wu, G. Genome-Wide Identification and Characterization of bHLH Transcription Factors Related to Anthocyanin Biosynthesis in Red Walnut (*Juglans regia* L.). *Front. Genet.* **2021**, *12*, 632509. [[CrossRef](#)] [[PubMed](#)]
22. Chandler, V.L.; Radicella, J.P.; Robbins, T.P.; Chen, J.; Turks, D. Two Regulatory Genes of the Maize Anthocyanin Pathway Are Homologous: Isolation of B Utilizing R Genomic Sequences. *Plant Cell* **1989**, *1*, 1175–1183. [[CrossRef](#)] [[PubMed](#)]
23. Li, C.; Qiu, J.; Ding, L.; Huang, M.; Huang, S.; Yang, G.; Yin, J. Anthocyanin Biosynthesis Regulation of *DhMYB2* and *DhbHLH1* in Dendrobium Hybrids Petals. *Plant Physiol. Biochem.* **2017**, *112*, 335–345. [[CrossRef](#)]
24. Li, X.; Fan, J.; Luo, S.; Yin, L.; Liao, H.; Cui, X.; He, J.; Zeng, Y.; Qu, J.; Bu, Z. Comparative Transcriptome Analysis Identified Important Genes and Regulatory Pathways for Flower Color Variation in *Paphiopedilum Hirsutissimum*. *BMC Plant Biol.* **2021**, *21*, 495. [[CrossRef](#)]
25. Jia, N.; Wang, J.J.; Liu, J.; Jiang, J.; Sun, J.; Yan, P.; Sun, Y.; Wan, P.; Ye, W.; Fan, B. DcTT8, a bHLH Transcription Factor, Regulates Anthocyanin Biosynthesis in Dendrobium Candidum. *Plant Physiol. Biochem.* **2021**, *162*, 603–612. [[CrossRef](#)] [[PubMed](#)]
26. Nakatsuka, T.; Suzuki, T.; Harada, K.; Kobayashi, Y.; Dohra, H.; Ohno, H. Floral Organ- and Temperature-Dependent Regulation of Anthocyanin Biosynthesis in Cymbidium Hybrid Flowers. *Plant Sci.* **2019**, *287*, 110173. [[CrossRef](#)] [[PubMed](#)]
27. Albert, N.W.; Arathoon, S.; Collette, V.E.; Schwinn, K.E.; Jameson, P.E.; Lewis, D.H.; Zhang, H.; Davies, K.M. Activation of Anthocyanin Synthesis in Cymbidium Orchids: Variability between Known Regulators. *Plant Cell Tiss. Organ. Cult.* **2010**, *100*, 355–360. [[CrossRef](#)]
28. Zhang, Y.; Zhou, T.; Dai, Z.; Dai, X.; Li, W.; Cao, M.; Li, C.; Tsai, W.C.; Wu, X.; Zhai, J.; et al. Comparative Transcriptomics Provides Insight into Floral Color Polymorphism in a *Pleione Limprichtii* Orchid Population. *Int. J. Mol. Sci.* **2019**, *21*, 247. [[CrossRef](#)] [[PubMed](#)]
29. Ke, Y.J.; Zheng, Q.D.; Yao, Y.H.; Ou, Y.; Chen, J.Y.; Wang, M.J.; Lai, H.P.; Yan, L.; Liu, Z.J.; Ai, Y. Genome-Wide Identification of the MYB Gene Family in *Cymbidium ensifolium* and Its Expression Analysis in Different Flower Colors. *Int. J. Mol. Sci.* **2021**, *22*, 13245. [[CrossRef](#)]
30. Heim, M.A. The Basic Helix-Loop-Helix Transcription Factor Family in Plants: A Genome-Wide Study of Protein Structure and Functional Diversity. *Mol. Biol. Evol.* **2003**, *20*, 735–747. [[CrossRef](#)]
31. Bailey, T.L.; Williams, N.; Misleh, C.; Li, W.W. MEME: Discovering and Analyzing DNA and Protein Sequence Motifs. *Nucleic Acids Res.* **2006**, *34*, W369–W373. [[CrossRef](#)]
32. Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. WebLogo: A Sequence Logo Generator: Figure 1. *Genome Res.* **2004**, *14*, 1188–1190. [[CrossRef](#)] [[PubMed](#)]
33. Flagel, L.E.; Wendel, J.F. Gene Duplication and Evolutionary Novelty in Plants. *New Phytologist* **2009**, *183*, 557–564. [[CrossRef](#)] [[PubMed](#)]

34. Fay, J.C.; Wu, C.I. Sequence Divergence, Functional Constraint, and Selection in Protein Evolution. *Annu. Rev. Genom. Hum. Genet.* **2003**, *4*, 213–235. [[CrossRef](#)]
35. Zhang, Z.; Li, J.; Zhao, X.Q.; Wang, J.; Wong, G.K.-S.; Yu, J. KaKs\_Calculator: Calculating Ka and Ks Through Model Selection and Model Averaging. *Genom. Proteom. Bioinform.* **2006**, *4*, 259–263. [[CrossRef](#)]
36. Yang, Z.; Bielawski, J.P. Statistical Methods for Detecting Molecular Adaptation. *Trends Ecol. Evol.* **2000**, *15*, 496–503. [[CrossRef](#)]
37. Wang, P.; Su, L.; Gao, H.; Jiang, X.; Wu, X.; Li, Y.; Zhang, Q.; Wang, Y.; Ren, F. Genome-Wide Characterization of bHLH Genes in Grape and Analysis of Their Potential Relevance to Abiotic Stress Tolerance and Secondary Metabolite Biosynthesis. *Front. Plant Sci.* **2018**, *9*, 64. [[CrossRef](#)]
38. Bailey, P.C.; Martin, C.; Toledo-Ortiz, G.; Quail, P.H.; Huq, E.; Heim, M.A.; Jakoby, M.; Werber, M.; Weisshaar, B. Update on the Basic Helix-Loop-Helix Transcription Factor Gene Family in Arabidopsis Thaliana. *Plant Cell* **2003**, *15*, 2497–2502. [[CrossRef](#)]
39. Guo, X.J.; Wang, J.R. Global Identification, Structural Analysis and Expression Characterization of bHLH Transcription Factors in Wheat. *BMC Plant Biol.* **2017**, *17*, 90. [[CrossRef](#)]
40. Wang, R.; Zhao, P.; Kong, N.; Lu, R.; Pei, Y.; Huang, C.; Ma, H.; Chen, Q. Genome-Wide Identification and Characterization of the Potato bHLH Transcription Factor Family. *Genes* **2018**, *9*, 54. [[CrossRef](#)]
41. Qi, Y.; Zhou, L.; Han, L.; Zou, H.; Miao, K.; Wang, Y. PsbHLH1, a Novel Transcription Factor Involved in Regulating Anthocyanin Biosynthesis in Tree Peony (*Paeonia suffruticosa*). *Plant Physiol. Biochem.* **2020**, *154*, 396–408. [[CrossRef](#)]
42. Geng, J.; Liu, J.H. The Transcription Factor CsbHLH18 of Sweet Orange Functions in Modulation of Cold Tolerance and Homeostasis of Reactive Oxygen Species by Regulating the Antioxidant Gene. *J. Exp. Bot.* **2018**, *69*, 2677–2692. [[CrossRef](#)]
43. Roelofs, J.; Van Haastert, P.J.M. Genes Lost during Evolution. *Nature* **2001**, *411*, 1013–1014. [[CrossRef](#)]
44. Brownlie, P.; Ceska, T.; Lamers, M.; Romier, C.; Stier, G.; Teo, H.; Suck, D. The Crystal Structure of an Intact Human Max–DNA Complex: New Insights into Mechanisms of Transcriptional Control. *Structure* **1997**, *5*, 509–520. [[CrossRef](#)]
45. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant* **2020**, *13*, 1194–1202. [[CrossRef](#)]
46. Artimo, P.; Jonnalagedda, M.; Arnold, K.; Baratin, D.; Csardi, G.; de Castro, E.; Duvaud, S.; Flegel, V.; Fortier, A.; Gasteiger, E.; et al. ExPASy: SIB Bioinformatics Resource Portal. *Nucleic Acids Res.* **2012**, *40*, W597–W603. [[CrossRef](#)]
47. Horton, P.; Park, K.J.; Obayashi, T.; Fujita, N.; Harada, H.; Adams-Collier, C.J.; Nakai, K. WoLF PSORT: Protein Localization Predictor. *Nucleic Acids Res.* **2007**, *35*, W585–W587. [[CrossRef](#)]
48. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027. [[CrossRef](#)]
49. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.A.; Clamp, M.; Barton, G.J. Jalview Version 2—a Multiple Sequence Alignment Editor and Analysis Workbench. *Bioinformatics* **2009**, *25*, 1189–1191. [[CrossRef](#)]
50. Letunic, I.; Bork, P. Interactive Tree Of Life (ITOL): An Online Tool for Phylogenetic Tree Display and Annotation. *Bioinformatics* **2007**, *23*, 127–128. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.