Available online: www.notulaebotanicae.ro

Print ISSN 0255-965X; Electronic 1842-4309

Not Bot Horti Agrobo, 2019, 47(3):903-912. DOI:10.15835/nbha47311531

Original Article

Characterization and Tissue-specific Expression of *bHLH* Genes in *Dimocarpus longan*

Wei ZHENG*, Xueming DONG, Xuefei YU, Qiuying ZHANG, Ning CHEN

*Harbin University of Commerce, School of Pharmacy (Research Center of Pharmaceutical Engineering Technology), No. 138, Tongda Street, Daoli District, Harbin, Heilongjiang Province, China; wei2013zheng@163.com (*corresponding author); xm_2019Dong@163.com; yuxuefei83@163.com; Qiuying88Zhang@163.com; 67404306@qq.com*

Abstract

In plants, the basic helix-loop-helix (*bHLH*) transcription factors (TFs) play pivotal roles in many biological processes including growth, stress response, and secondary metabolite synthesis. To date, many *bHLH* genes have been identified and characterized in diverse plant species. However, little is known regarding the *bHLH* genes in *Dimocarpus longan* Lour. (*D. longan*). Based on RNA-seq data, we identified 42 putative *bHLH* genes from *D. longan* and determined their putative functions using the NCBI Conserved Domain Search Tool and Pfam databases. The physicochemical properties, phylogenetic relationships, conserved motifs, gene ontology (GO) annotations, protein-protein interactions, and tissue-specific expression patterns of these *bHLH* genes were systematically explored. In total, ten motifs were found in DlbHLH proteins using MEME, among which two were highly conserved. Phylogenetic tree analysis found that DlbHLH proteins can be divided into nine groups, with group 2 being the largest. GO annotation results showed that the *DlHLH* genes were involved in various molecular functions. RNA-seq and qRT-PCR results revealed important differences in the expression patterns of 17 of the *DlbHLH* genes. In particular, *DlbHLH-9*, *DlbHLH-19, DlbHLH-25, DlbHLH-26,* and *DlbHLH-35* were found to show significantly different expression patterns in root and leaf tissues. The results of this study will further enrich our knowledge regarding *bHLH* transcription factor genes and lay a foundation for enhancing the production of active secondary metabolites by genetic engineering in *D. longan*.

Keywords: bHLH transcription factor; bioinformatics analysis; *Dimocarpus longan* Lour.; expression patterns

Introduction

Transcription factors (TFs) are an important group of DNA-binding proteins that recognize and bind to specific DNA sequences to control transcription from DNA to mRNA at specific times and places. TFs are usually characterized by the possession of four functional regions, including a nuclear localization signal, a DNA binding domain, a transcription regulation domain, and an oligomerization site (Yang *et al*., 2012; Yamasaki *et al*., 2013; Guo and Wang, 2017). In plants, *bHLH* TFs are the second largest family after *MYB* TFs (Sun *et al*., 2018; Yu *et al*., 2019). These TFs contain the highly conserved bHLH domain, which includes both a basic region and a HLH region. The basic region is usually located at the N-terminus of the bHLH domain, and permits binding to E-box sequences (5'-CANNTG-3') in target gene promoters (Heim *et al*., 2003). The HLH region is usually located at the C-terminus of the bHLH domain, and is approximately 50 amino acids long. It contains two alpha helixes separated by a loop, and forms homodimeric or heterodimeric complexes with other bHLH proteins, thereby regulating their activity (Massari and Murre, 2000).

To date, the diversity of the bHLH family has been explored in many species including *Arabidopsis thaliana*, *Brachypodium distachyon*, peanut, Chinese cabbage, rice, *Salvia miltiorrhiza*, and tomato (Toledo-Ortiz *et al*., 2003; Li *et al*., 2006; Wang *et al*., 2015a; Zhang *et al*., 2015; Wu *et al*., 2016b; Chao *et al*., 2017; Niu *et al*., 2017). Moreover, it has also been shown that bHLH TFs play an important role in active secondary metabolism. The Lc protein, the first

Received: 27 Apr 2019. Received in revised form: 20 Jun 2019. Accepted: 24 Jun 2019. Published online: 01 Jul 2019.

reported bHLH TF, is encoded by the maize *R* gene, and has been shown to be involved in regulating the expression of two structural genes related to the maize anthocyanin metabolic pathway (Ludwig *et al*., 1989). In snapdragon, the expression of the *Delila* gene, which contains a helix-loop-helix domain, was also found to have close relationships with anthocyanin accumulation (Goodrich *et al*., 1992). Moreover, bHLH TFs participate in the regulation of terpenoid biosynthesis. *AtMYC2*, an *Arabidopsis thaliana* bHLH TF, can interact with the target promoter regions of sesquiterpene synthases genes, thereby activating their transcription and increasing sesquiterpene accumulation (Hong *et al*., 2012). A diterpenoid phytoalexin factor (DPF) belonging to the bHLH family has been shown to positively regulate the transcript level of rice diterpenoid phytoalexin (DP) genes, and is thereby linked to DP accumulation (Yamamura *et al*., 2015).

The functions of members of the bHLH TF family involved in secondary metabolite synthesis have been widely studied in many plant species. However, to date no studies have examined the bHLH TF family in *D. longan*. *D. longan* is a common fruit tree in China that is valuable for human consumption and medicine. However, the active secondary metabolites that are the medically active components of *D. longan* are accumulated in only small amounts in root and leaf tissues. Thus, identification of *bHLH* TF genes related to the accumulation of secondary metabolites may facilitate the development of *D. longan* plant resources in southern China. Recently, the genome sequencing of *D. longan* was performed, the results of which showed its genetic diversity. The genome sequence data analysis not only revealed the unique characteristics of *D. longan*, but also highlighted the genes that are possibly involved in the accumulation of secondary metabolites in *D. longan* (Lin *et al*., 2017). Genetic engineering and targeted breeding programs could then be used to enhance the production of these active secondary metabolites in *D. longan*.

In the present study, 42 bHLHs were identified and their physicochemical properties, motif compositions, phylogenetic relationships, gene ontology (GO) annotations, and protein-protein interactions were examined. In addition, we used RNA-seq and quantitative real time PCR (qRT-PCR) to study the expression patterns of these *bHLHs* in root and leaf tissues. The results of this study will lay the foundation for further studies of the biosynthesis of secondary metabolites in *D. longan* and further highlight the importance of *bHLH* TFs in plants.

Materials and Methods

Plant material

D. longan plant tissue was obtained from plants cultivated in a greenhouse with the humidity of 50% and temperature of 25 °C. Leaves from the upper peripheral branches and roots from 10 plants were collected from individuals after 2 months growth. All samples were immediately put in liquid nitrogen for later RNA isolation.

Identification of bHLH genes in D. longan

We obtained sequences from the RNA-seq data deposited to the non-redundant (NR) NCBI database (NCBI accession number: SRP155595) to identify 42 putative *DlbHLH* genes based on NR annotation. Moreover, we examined all sequences using the NCBI Conserved Domain Search Tool (https:// www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and the Pfam database (http://pfam.xfam.org/) to further confirm the identity of all putative DlbHLH TFs. All 42 confirmed DlbHLH TFs were retained for further analyses.

Bioinformatics analysis of the DlbHLH genes in D. longan

ExPasy software was used to investigate the molecular weight, protein sequence length, instability index, aliphatic index, isoelectric point, and grand average of hydropathicity (GRAVY) for all DlbHLH proteins. The Self-Optimized Prediction method With Alignment (SOPMA) tool was used to predict the proportions of extended strands, alpha helices, beta turns, and random coils in all proteins. The conserved motifs present in the DlbHLH proteins were identified using Multiple Expectation Maximization for Motif Elicitation (MEME). A phylogenetic tree was generated by MEGA 7.0 using the neighbor-joining method with 1,000 bootstrap iterations (Tamura *et al*., 2011). The functional regulatory network of *D. longan* bHLH proteins was studied using the Protein-Protein Interaction Networks (STRING) online tool. Finally, Blast2GO PRO was applied to analyze the functional classification of bHLH proteins and to acquire detailed annotations (Conesa and Gotz, 2008). The online softwares used in this study were shown in Table 1.

Table 1. Online software for bioinformatics analysis

Name	Function	Web address
NCBI CD search	Conservative domain prediction	http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi
ExPasy	Physical and chemical properties prediction	http://web.expasy.org/protparam/
MEME	Conserved motifs prediction	http://meme-suite.org/tools/meme
STRING	Protein-protein interactions prediction	http://string-db.org/
SOPMA	Secondary structure prediction	https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html

RNA isolation and quantitative real time PCR analysis

Total RNA was extracted from root and leaf of *D. longan* using the cetyltrime thylammonium bromide (CTAB) method (Jaakola *et al*., 2001). First strand cDNAs were synthesized using the TransScript® One-Step gDNA Removal and cDNA Synthesis SuperMix kit (TransGen Biotech, Beijing, China). The TransStart® Top Green qPCR SuperMix (TransGen Biotech, Beijing, China) was used for all qRT-PCR reactions according to the manufacturer's instructions. The reaction system were 20 μL, containing 10 μL 2×TransStart® Top Green qPCR SuperMix, l μL cDNA template, 7 μL ddH_2O, 1 μL forward primer and 1 μL reverse primer. Primers used for qRT-PCR are shown in Table 2. The *D. longan tubulin* gene was used as reference. The qRT-PCR reactions conditions were as follows: 95℃ for 1 min, followed by 95℃ for 5 s, 60℃ for 30s, and 72℃ for 30 s. All experiments were performed in triplicate. Relative gene expression was computed using the relative quantification ($2^{-\Delta\Delta CT}$) method (Schmittgen and Livak, 2008).

Results

Identification of D. longan bHLH genes and their physicochemical properties

Using NR annotations, the NCBI Conserved Domain Search Tool, and the Pfam database to analyze *D. longan* RNA-seq data (NCBI accession number: SRP155595), we identified 42 genes as putative *DlbHLH* TF genes. We designated the 42 *DlbHLH* genes *DlbHLH-1* to *DlbHLH-42* according to the order of these genes in the original RNA-seq experiment. We then assessed the physicochemical properties of these TFs; our analyses included determinations of their open reading frame (ORF) length, theoretical isoelectric point, aliphatic index, molecular weight, instability index (II), and grand average of hydropathicity (GRAVY), as well as the number of alpha helices, extended strands, beta turns, and random coils. The detailed results were shown in Table 3.

Analysis of conserved motifs in DlbHLH proteins

MEME was used to investigate the motif compositions of the DlbHLH proteins identified here. In total, we identified 10 conserved motifs, and the positions where these motifs were found in DlbHLH proteins are shown in Fig. 1. All 42 DlbHLH proteins were found to contain Motif 1, and Motif 2 was also present in 40 of 42 DlbHLHs.

Phylogenetic analysis of DlbHLH proteins

A phylogenetic tree was used to investigate the evolutionary relationships among the 42 bHLH proteins from *D. longan*. Based on the classifications of rice and *Arabidopsis thaliana* (Toledo-Ortiz *et al*., 2003), the DlbHLH proteins could be divided into nine groups, with DlbHLH-3, DlbHLH-14, DlbHLH-25, and DlbHLH-40 unable to be classified (Fig. 2). Group 2, which contains 13 DlbHLH proteins, was the largest among these groups, whereas Groups 3, 4, 5, and 8 were the smallest-each had 2 DlbHLH proteins. The other 4 groups each contained 4-5 DlbHLH proteins.

Gene Ontology annotation of DlbHLH genes

GO annotation of all 42 *DlbHLH* genes as biological process, molecular function, or cellular component genes was performed using Blast2GO v5.2.5 with graph level 2 (Fig. 3). In total, 22 *DlbHLH* genes were identified as metabolic process genes, accounting for 37% of all *DlbHLH* genes identified. 22 *DlbHLH* genes were also identified as cellular process genes (37% of the total), followed by biological regulation (13%) and regulation of biological process (13%). Within the molecular function category, 26 and 9 *DlbHLH* genes were predicted to be associated with binding (74%) and transcription regulator activity (26%), respectively. Cellular component prediction showed that genes associated with five different terms were defined, including cell parts (27%), organelles (27%), cells (27%), protein-containing complexes (10%), and organelle parts (10%), respectively. Moreover, those *DlbHLH* genes that were predicted to have multiple classifications are listed in Table 4.

Table 2. Primer sequences for qRT-PCR

Gene	Primer sequence	Gene	Primer sequence
DlbHLH-3	QF: CGATGCCTTGTCCATCTTGC QR: GCGTCTACTGCGGCTTCTAA	*DlbHLH-24*	QF:CTTACTGCGGTGGAGTGGAG QR:AACTCCTTCATCTGTGCGGG
DlbHLH-5	QF:TCCAGAGCTCTCATGGTGGA QR:CTGCAAGGGTGGAGGTTAGG	*DlbHLH-25*	QF:TCTGTCCGGTTCTCGTGTTG QR:CCGGGTCAGAGAACGGAAAA
DlbHLH-6	QF: GTGGCGATGAACCAACTTCG QR: ACTTGACTCAGGCGCTTGTT	*DlbHLH-26*	QF: TGTTGAACCCCAGTGGATGG QR:GCTGCTTGGTCACTGAAAGC
DlbHLH-9	QF: TGTTGAGAGAAACCGGCGAA QR: TAGAGGCTTGGTCACCCCTT	*DlbHLH-35*	QF: GATGCATGAGGGAGCGAAGA QR:CCCGGGACAGACGAAAAAGA
DlbHLH-14	QF: GCTGTCTCGAGCATGGAACT QR: CCAACTGGACTCATGACGGT	*DlbHLH-36*	QF:ATCAACGGAGCCATGCAAGA QR:TACCTCCGCCAGCAAAGAAG
DlbHLH-15	QF: ATTACAGCAGTTGCAGCCCT QR: GGCTGTGATGGAAAGGGTGT	*DlbHLH-38*	QF:TAGCGAAACCTCTCTTGGCG QR:CCGGTCTAATGGCCTGACTC
DlbHLH-16	QF: GACCCATGCCAACCTCAAGA QR:CGGTTTGATCAACGGTGGTG	*DlbHLH-40*	QF:ACCTGCCACCAACTCTAAGC QR:ACCCAAATGACGTGGGAAGT
DlbHLH-18	QF: GAACCCTACAAGGCACCGAA QR:TCCCGTCTCACCCGAAACTA	*DlbHLH-42*	QF:GCTCTCCACCATGGTCACTT QR:CGGGTGTGAGAACTCAGCTT
DlbHLH-19	QF: CACGGGTGATGAGCTGTTCT QR:GTCCAGTGAAGCGTGATCCA	*Tubulin*	QF: CTCATGTATGCCAAGCGTGC QR:CTCTGCAGACTCAGCACCAA

Table 3. Physicochemical properties of the *bHLH* genes in *D. longan*

Gene	Gene ID	ORF (aa)	PI	Aliphatic index	MW (Da)	Instability index (II)	GRAVY	Alpha helix	Extended strand	Beta turn	Random coil
DlbHLH-1	c20189.graph_c0	531	5.19	78.64	56786.08	46.23	-0.434	27.12%	8.66%	3.01%	61.21%
DlbHLH-2	c26370.graph_c0	620	6.09	79.73	68652.31	44.77	-0.425	36.45%	12.26%	2.42%	48.87%
DlbHLH-3	c37707.graph_c0	244	5.17	91.60	27683.59	56.97	-0.442	53.28%	9.02%	1.64%	36.07%
DlbHLH-4	c32269.graph_c0	269	7.13	67.17	29650.99	57.31	-0.723	30.48%	11.15%	3.35%	55.02%
DlbHLH-5	c30836.graph_c0	481	5.37	76.63	53078.37	39.77	-0.484	23.49%	12.06%	3.74%	60.71%
DlbHLH-6	c31725.graph_c0	426	5.92	59.04	46163.33	55.26	-0.702	21.83%	5.87%	3.76%	68.54%
DlbHLH-7	c10261.graph_c0	91	7.93	102.86	10327.74	91.43	-0.505	64.84%	1.10%	0.00%	34.07%
DlbHLH-8	c26732.graph_c0	314	5.38	81.62	35354.95	43.56	-0.481	37.26%	14.01%	2.23%	46.50%
DlbHLH-9	c27950.graph_c0	305	8.90	88.56	34432.25	51.92	-0.432	42.62%	9.51%	2.30%	45.57%
DlbHLH-10	c47015.graph_c0	76	9.52	62.89	8628.79	60.79	-1.018	43.42%	11.84%	0.00%	44.74%
DlbHLH-11	c12611.graph_c0	347	5.96	80.69	37985.72	64.46	-0.550	40.06%	8.36%	2.88%	48.70%
DlbHLH-12	c30269.graph_c0	364	8.37	58.65	38754.60	59.71	-0.599	30.22%	6.04%	2.47%	61.26%
DlbHLH-13	c20620.graph_c0	208	8.63	77.74	23814.08	56.30	-0.725	50.48%	12.50%	5.29%	31.73%
DlbHLH-14	c29442.graph_c0	362	6.00	76.96	40549.33	50.42	-0.635	33.43%	15.47%	3.59%	47.51%
DlbHLH-15	c37749.graph_c0	435	5.26	63.61	48033.64	53.45	-0.620	34.94%	5.98%	2.76%	56.32%
DlbHLH-16	c10452.graph_c0	335	5.72	70.69	37341.05	63.75	-0.526	33.13%	11.34%	1.79%	53.73%
DlbHLH-17	c40711.graph_c0	200	6.45	78.95	22884.83	32.78	-0.661	46.50%	14.00%	2.00%	37.50%
DlbHLH-18	c21544.graph_c0	247	6.54	64.01	27398.56	71.73	-0.802	36.84%	3.24%	2.43%	57.49%
DlbHLH-19	c34174.graph_c0	239	9.26	87.24	27042.97	50.92	-0.595	40.17%	11.30%	2.09%	46.44%
DlbHLH-20	c13201.graph_c0	238	7.72	92.18	26915.67	46.19	-0.464	50.42%	7.56%	0.84%	41.18%
DlbHLH-21	c10742.graph_c0	458	5.75	65.13	47664.84	45.64	-0.465	34.93%	5.46%	3.06%	56.55%
DlbHLH-22	c31814.graph_c0	415	6.14	52.43	44204.76	47.32	-0.749	15.42%	6.99%	2.65%	74.94%
DlbHLH-23	c10942.graph_c0	276	6.01	63.33	29657.81	56.89	-0.686	28.62%	8.70%	3.99%	58.70%
DlbHLH-24	c31651.graph_c0	237	7.58	86.67	26638.30	45.96	-0.486	40.93%	12.66%	3.38%	43.04%
DlbHLH-25	c17501.graph_c0	227	8.83	73.00	26522.09	55.40	-0.854	61.23%	7.49%	0.88%	30.40%
DlbHLH-26	c13202.graph_c0	341	4.84	71.79	38555.29	56.63	-0.582	40.76%	7.62%	2.35%	49.27%
DlbHLH-27	c40111.graph_c0	84	6.07	111.31	9252.52	51.42	-0.299	63.10%	0.00%	2.38%	34.52%
DlbHLH-28	c14185.graph_c0	320	6.67	58.56	35520.25	54.68	-0.993	35.62%	4.38%	1.88%	58.13%
DlbHLH-29	c31147.graph_c0	546	6.10	60.75	59805.46	52.76	-0.698	21.79%	2.93%	2.75%	72.53%
DlbHLH-30	c23863.graph_c0	229	7.14	70.22	26077.50	47.95	-0.631	40.17%	8.73%	5.68%	45.41%
DlbHLH-31	c11341.graph_c0	357	8.49	59.52	39482.12	54.91	-0.965	35.01%	5.32%	1.96%	57.70%
DlbHLH-32	c12978.graph_c0	93	6.57	98.60	10462.73	89.75	-0.504	64.52%	1.08%	4.30%	30.11%
DlbHLH-33	c25784.graph_c0	262	8.27	88.97	28904.70	51.24	-0.490	46.95%	12.60%	2.67%	37.79%
DlbHLH-34	c19493.graph_c0	222	5.42	78.60	25118.29	44.12	-0.496	45.95%	13.51%	2.25%	38.29%
DlbHLH-35	c33268.graph_c0	384	5.69	81.48	43692.38	65.59	-0.626	45.31%	8.33%	2.60%	43.75%
DlbHLH-36	c11569.graph_c0	230	8.77	69.13	25526.75	48.17	-0.643	46.09%	7.83%	0.87%	45.22%
DlbHLH-37	c29549.graph_c0	247	8.89	64.74	28079.85	48.26	-0.858	34.82%	11.34%	4.86%	48.99%
DlbHLH-38	c32080.graph_c0	441	8.95	51.09	48989.84	50.12	-0.936	24.49%	7.26%	2.95%	65.31%
DlbHLH-39	c20949.graph_c0	91	7.94	97.58	10284.59	74.50	-0.596	70.33%	1.10%	2.20%	26.37%
DlbHLH-40	c28975.graph_c0	355	5.89	78.31	39983.16	53.17	-0.532	38.59%	8.17%	1.69%	51.55%
DlbHLH-41	c16315.graph_c0	505	5.92	76.40	55680.71	44.75	-0.467	40.00%	11.68%	2.97%	45.35%
DlbHLH-42	c26560.graph_c0	279	7.20	72.37	31038.69	63.04	-0.628	39.43%	8.96%	2.51%	49.10%

Table 4. Function-based classification of the *DlbHLH* genes

Gene Name	Biological Process	Molecular function	Cellular Component
DlbHLH-1		√	√
DlbHLH-2		√	
DlbHLH-3		√	
DlbHLH-4	√	√	
DlbHLH-5		√	
DlbHLH-6	√	√	
DlbHLH-7	√	√	
DlbHLH-8		√	
DlbHLH-9		√	
DlbHLH-10	√	√	√
DlbHLH-11	√	√	
DlbHLH-12		√	
DlbHLH-13	√	√	√
DlbHLH-14		√	
DlbHLH-15	√	√	
DlbHLH-16		√	
DlbHLH-17	√	√	√
DlbHLH-18		√	
DlbHLH-19	√	√	√
DlbHLH-20	√	√	
DlbHLH-21		√	
DlbHLH-22	√	√	
DlbHLH-23	√	√	
DlbHLH-24	√	√	
DlbHLH-25		√	
DlbHLH-26		√	
DlbHLH-27	√	√	
DlbHLH-28		√	
DlbHLH-29	√	√	
DlbHLH-30	√	√	√
DlbHLH-31		√	
DlbHLH-32	√	√	√
DlbHLH-33	√	√	
DlbHLH-34	√	√	√
DlbHLH-35		√	
DlbHLH-36	√	√	
DlbHLH-37	√	√	√
DlbHLH-38		√	√
DlbHLH-39	√	√	
DlbHLH-40		√	
DlbHLH-41	√	√	√
DlbHLH-42		√	

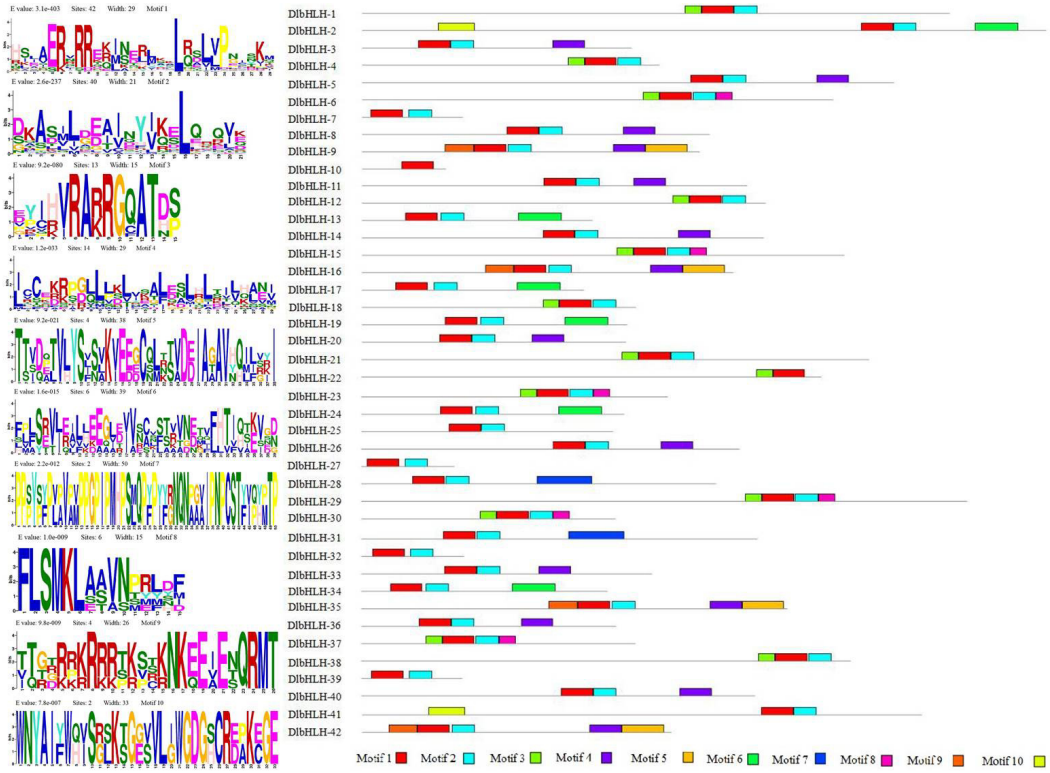

Fig. 1. Motif compositions and distributions of *D. longan* bHLH TF proteins; E-values represent expected values. Sites refer to the number of identified DlbHLH proteins containing the motif. Width refers to the length of the motif sequence. The height of each letter is proportional to amino acid frequency. Each colored boxes represent different motifs and black lines indicate the non-conserved regions

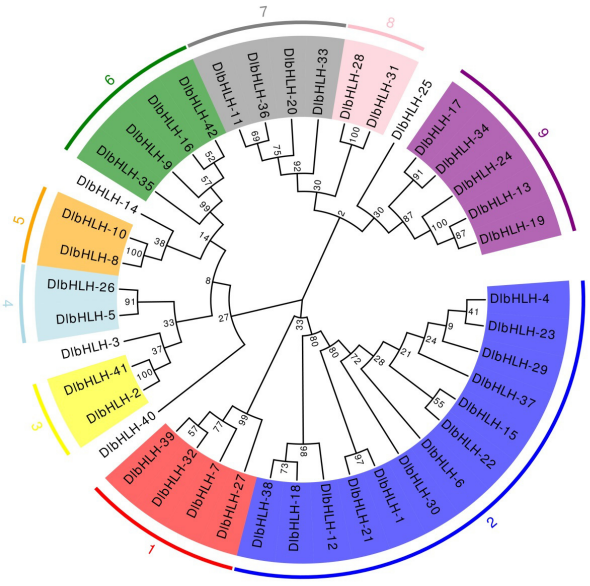

Fig. 2. Phylogenetic tree of bHLH proteins identified in *D. longan*; the tree was generated with MEGA 7.0 using the neighbor-joining method with 1000 bootstrap replicates. Numbers indicate bootstrap values; colours indicate different DlbHLH proteins groups

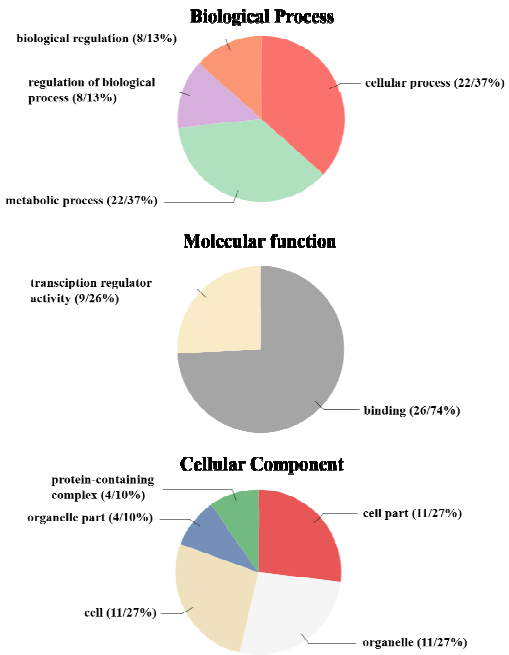

Fig. 3. Gene Ontology (GO) annotation results for *bHLH* genes identified in *D. longan*. Blast2GO was used to predict the classification of *DlbHLH* genes in three broad categories (biological processes, molecular functions, and cellular components); more specific breakdowns of each class are shown here. Different colours indicate the different functions

Analysis of protein-protein interactions

In this study, Protein-Protein Interaction Networks (STRING) software was used to predict DlbHLH protein interactions. The results are shown in Fig. 4. As genomic data for *D. longan* were not available in STRING, bHLHs proteins from *Arabidopsis thaliana* with high homology to those from *D. longan* were selected as representatives for protein interaction studies, since-to some extent-these are likely to reflect the relationships among *D. longan* bHLH proteins.

As shown in Fig. 4, most DlbHLH proteins were predicted to interact with more than one bHLH protein. We found 20 DlbHLH proteins (DlbHLH-2, DlbHLH-3, DlbHLH-6, DlbHLH-9, DlbHLH-12, DlbHLH-14, DlbHLH-15, DlbHLH-16, DlbHLH-20, DlbHLH-22, DlbHLH-23, DlbHLH-25, DlbHLH-27, DlbHLH-32, DlbHLH-33, DlbHLH-35, DlbHLH-37, DlbHLH-39, DlbHLH-41, and DlbHLH-42) that could interact with three or more bHLH proteins. However, five DlbHLH proteins (DlbHLH-1, DlbHLH-18, DlbHLH-28, DlbHLH-30, and DlbHLH-31) were not predicted to interact with any bHLH proteins.

Among the DlbHLH proteins that were predicted to interact with others, DlbHLH-3 (homologous to AT5G57150) and DlbHLH-25 protein (homologous to BHLH92) were predicted to be co-expressed-connected with a black line. Proteins linked with purple lines indicated interactions that were experimentally validated. In *Arabidopsis thaliana*, it has been experimentally determined that AT1G68810 and AT5G51780 could interact with AT2G31220 and AT3G61950, respectively. Therefore, since homologous proteins often have similar biological functions, we speculated that DlbHLH-11/DlbHLH-36 and DlbHLH-19/DlbHLH-24 could interact with DlbHLH-14 and DlbHLH-35, respectively.

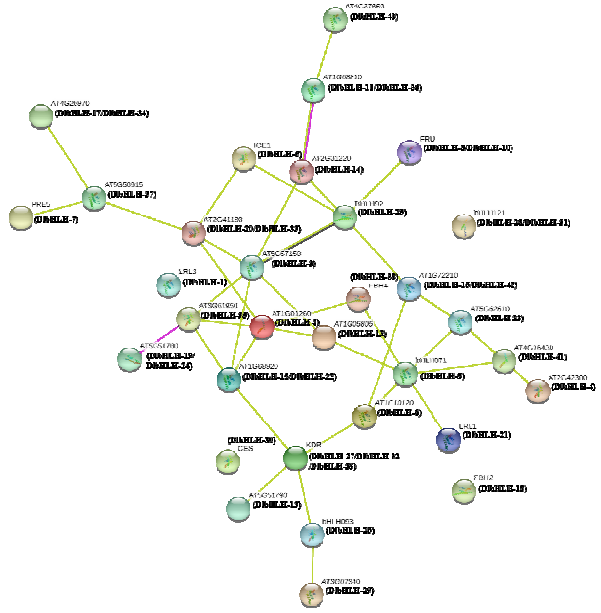

Fig. 4. Functional regulatory network of *D. longan* bHLH proteins; purple lines represent experimentally validated interactions, yellow lines show relationships derived from previous studies and the black lines represent protein co-expression

DlbHLH gene expression patterns in root and leaf tissues

The comparative expression of *D. longan bHLH* genes were analyzed in root and leaf tissues. Using previously published RNA-seq data, the expression patterns of the identified 42 *DlbHLH* genes were shown in a heat map in Fig. 5. As shown in Fig. 5, only 17 *DlbHLH* genes (40.48% of the total), including *DlbHLH-3, DlbHLH-5, DlbHLH-6, DlbHLH-9, DlbHLH-14, DlbHLH-15, DlbHLH-16, DlbHLH-18, DlbHLH-19, DlbHLH-24, DlbHLH-25, DlbHLH-26, DlbHLH-35, DlbHLH-36, DlbHLH-38, DlbHLH-40,* and *DlbHLH-42*, were shown to have obviously different expressions between root and leaf (Fig.5).

Next, in order to further validate the RNA-seq-derived patterns of gene expression, we performed qRT-PCR analysis on the 17 *DlbHLH* genes expressed in both root and leaf. These results, shown in Fig. 6, revealed that all 17 tested *DlbHLH* genes were expressed, with different levels of expression in root and leaf. The expression levels of 12 genes (including *DlbHLH-3, DlbHLH-6, DlbHLH-14, DlbHLH-16, DlbHLH-18, DlbHLH-19, DlbHLH-24, DlbHLH-25, DlbHLH-26, DlbHLH-38, DlbHLH-40*, and *DlbHLH-42*) were higher in roots than that in leaves, whereas the expression patterns of the other five genes (*DlbHLH-5, DlbHLH-9, DlbHLH-15, DlbHLH-35,* and *DlbHLH-36*) were the opposite. Significant differences in gene expression between root and leaf tissues were found for *DlbHLH-9*, *DlbHLH-19, DlbHLH-25, DlbHLH-26,* and *DlbHLH-35*. The expression levels of *DlbHLH-9 and DlbHLH-35* in leaf were 43- and 80-fold higher than that in root, respectively. In contrast, the expression levels of *DlbHLH-19, DlbHLH-25,* and *DlbHLH-26* in root were 29-, 33-, and 27-fold higher than in leaf, respectively. Moreover, the patterns of expression of the 17 *DlbHLH* genes revealed by qRT-PCR were in accordance to the patterns previously found in the RNA-seq data.

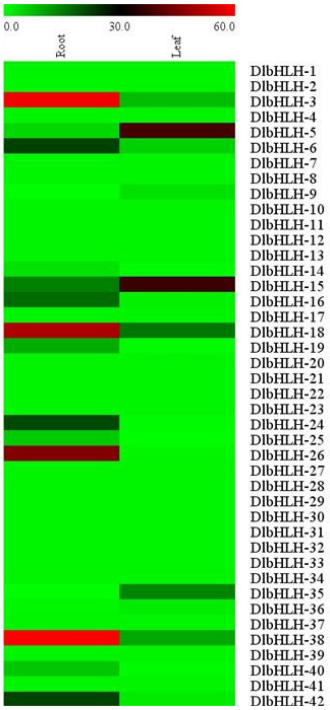

Fig. 5. Heatmap showing the expression profiles of the 42 *DlbHLH* genes

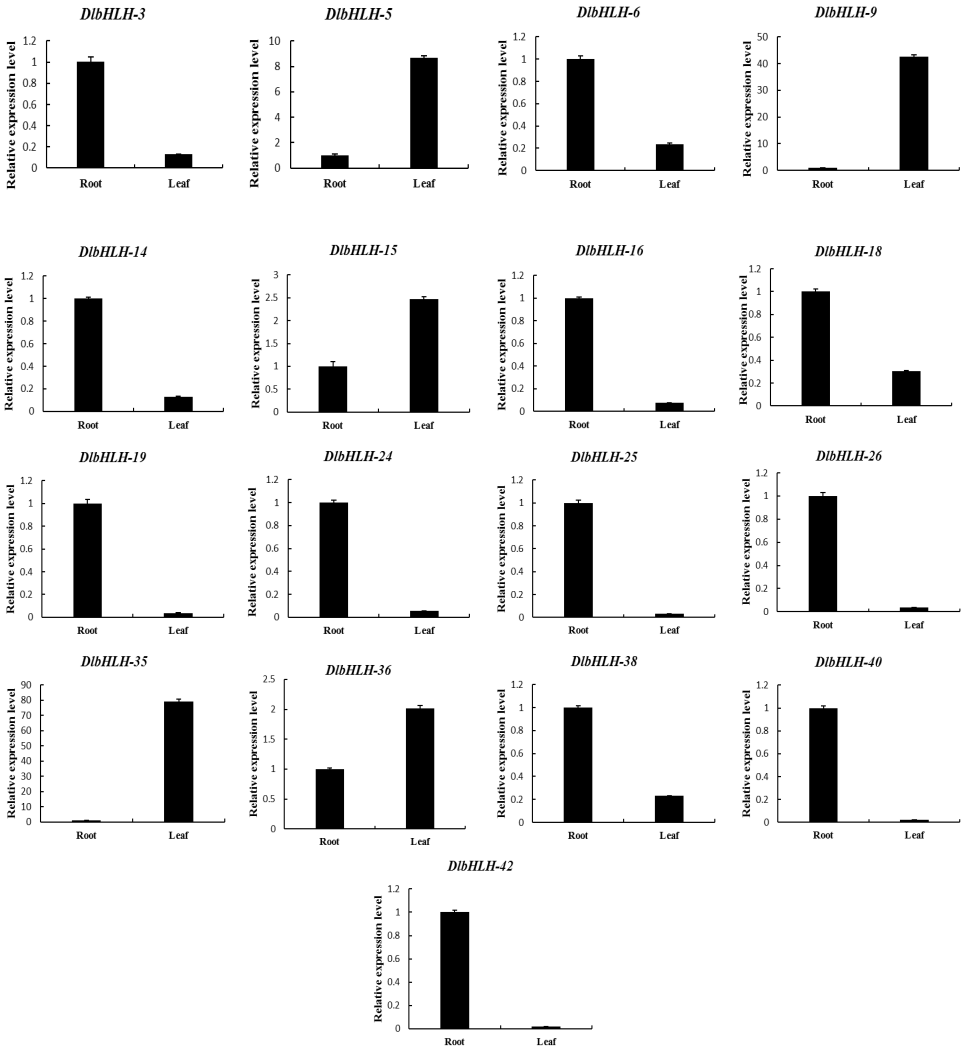

Fig. 6. Expression profiles of the 17 selected *DlbHLH* genes in root and leaf; error bars indicate standard error of three independent replicates

Discussion

Transcription factors are important regulatory genes involved in diverse biological processes, including plant growth, development, stress response and secondary metabolite synthesis. To date, only a few transcription factor families-such as the WRKY TF family-have been systematically studied in *D. longan* (Jue *et al*., 2018). No studies of the bHLH TFs have yet been performed in *D. longan*, although bHLH TFs have been identified and studied in many other plant species, including Arabidopsis, Brachypodium distachyon, and rice (Toledo-Ortiz *et al*., 2003; Li *et al*., 2006; Niu *et al*., 2017). Studies of bHLH TFs have demonstrated that they are closely related to diverse biological functions, especially those involved in secondary metabolite synthesis (Heim *et al*., 2003). *D. longan*, which is consumed both for food and medicine, has important commercial and medicinal values. Moreover, the root and leaf tissues of *D. longan* have been shown to possess many bioactive chemicals, including flavonoids, phytosterols, glycosides, and organic acids (Xue *et al*., 2015). Moreover, enhancing the production of active secondary metabolites present in *D. longan* roots and leaves by gene engineering can significantly expand the scope of its application and increase its value as a crop.

In this study, *D. longan* RNA-seq data was used to identify and characterize the 42 putative *DlbHLH* genes. The number of *bHLH* genes varies among higher plants, lower plants, and fungi. Higher plants such as *Brassica napus*, *Glycine max*, and *Panicum virgatum* contain numerous *bHLH* genes, while only one *bHLH* gene was identified in the lower plants and fungi such as *Bathycoccus*, *Ostreococcus tauri*, *Ostreococcus lucimarinus*, and *Helicosporidium* (Zhang *et al*., 2018). Based on these facts, we speculate that the *bHLH* gene family had undergone expansion during evolution, and that this expansion had likely resulted in the emergence of novel biological functions.

The length of *bHLH* genes in *D. longan* varied from 228 bp to 1,860 bp. In *Panax ginseng*, the length of *bHLH* genes ranged from 283 bp to 2,857 bp (Chu *et al*., 2018). By contrast, the longest *bHLH* gene in *D. longan* was significantly shorter (i.e. by about 1,000bp) than that in *Panax ginseng*, suggesting that *bHLH* gene lengths vary significantly among different species. The span of theoretical isoelectric points in *D. longan* bHLH proteins was large, ranging from 4.84 to 9.52, suggesting that different DlbHLH proteins might be functional in diverse microenvironments. The theoretical isoelectric points of *Panax ginseng* bHLH proteins were close to those found in *D. longan*, and varied from 4.81 to 10.16 (Chu *et al*., 2018).

We also evaluated the stability of the DlbHLH proteins. A protein whose instability index is larger than 40 is likely unstable, while those with values under 40 are likely stable (Guruprasad *et al*., 1990). The instability indexes of the DlbHLH-5 and DlbHLH-17 proteins were both predicted to be under 40 (39.77 and 32.78, respectively), while the instability index values of all other DlbHLH proteins were above 40. Thus *D. longan* contains both stable and unstable bHLH proteins, but the unstable bHLH proteins predominate. In addition, we found that all 42 DlbHLH proteins identified in *D. longan* had negative GRAVY scores. Since proteins with negative GRAVY scores are predicted to be soluble (Kyte and Doolittle, 1982), this means that all 42 DlbHLH proteins are likely soluble. This conclusion is consistent with the general requirement that transcription factors should be soluble.

MEME was used to predict the conserved motifs in the 42 DlbHLH proteins identified in *D. longan*. In total, 10 conserved motifs were found, with motifs 1 and 2 present in many proteins. Because of their ubiquity, we speculate that motifs 1 and 2 are likely related to the core functions of bHLH proteins. A neighbor-joining phylogenetic tree was created, in which DlbHLH proteins with bootstrap values above 50 clustered together (Toledo-Ortiz *et al*., 2003). In general, the bHLH TFs of plants clustering in the same group participate in similar biological processes (Pires and Dolan, 2010). In *D. longan*, 42 DlbHLH proteins were divided into 9 groups, and were likely to be involved in 9 biological processes. This result suggested the possible biological processes that the 42 DlbHLH proteins are involved in, and each of these putative functions requires further examination in future work. Predicting protein-protein interactions is useful for investigating the physiological functions of proteins (Zhang *et al*., 2018), and can be especially valuable for those that, like bHLH family proteins, interact with each other. In this study, 37 DlbHLH proteins were predicted to interact with each other, which suggested that they may not function alone but require the presence of other DlbHLH proteins.

Next, we systematically explored the expression profiles of *DlbHLH* genes in *D. longan* root and leaf tissues. According to our RNA-seq dataset, 17 of 42 *DlbHLH* genes had different expression levels in root and leaf. To further confirm these results, qRT-PCR was performed to investigate the expression profiles of *DlbHLH* genes that showed differential expression level in the RNA-seq data. In conclusion, our qRT-PCR results were in accordance with the RNA-seq data.

The genetic engineering of transcription factors has proven to be an effective strategy to enhance the accumulation of secondary metabolites and to increase the yield of crops and medicinal plants (Gantet and Memelink, 2002). In the hairy roots of *Salvia miltiorrhiza*, overexpression of the *SmbHLH10* gene has been shown to enhance the accumulation of tanshinones (Xing *et al*., 2018a), and overexpression of the *SmbHLH148* gene induced tanshinone and phenolic acid productions (Xing *et al*., 2018b). Therefore, we speculate that the five *DlbHLH* genes that showed significantly different expression patterns in root and leaf (i.e. *DlbHLH-9*, *DlbHLH-19, DlbHLH-25, DlbHLH-26,* and *DlbHLH-35*) deserves further study in their potential to enhance the production of valuable secondary metabolites in *D. longan*.

Conclusions

In this study, 42 *DlbHLH* genes were identified in *D. longan* using transcriptomic data, the NCBI Conserved Domain Search Tool, and the Pfam database. The physicochemical properties, phylogenetic relationships, conserved motifs, GO annotations, and protein-protein interactions of these genes were then examined using bioinformatics tools. Moreover, RNA-seq data and qRT-PCR results indicated that 17 of 42 *DlbHLH* genes expressed differently in root and leaf. Among these *DlbHLH* genes, *DlbHLH-9*, *DlbHLH-19, DlbHLH-25, DlbHLH-26,* and *DlbHLH-35* exhibited significant tissue-specific expression, which is deserving of further investigation in the future. The results of this study will enrich our knowledge of the *bHLH* TF family in *D. longan* and lay a foundation for enhancing the production of active secondary metabolites by genetic engineering in *D. longan*.

Acknowledgements

This work is financially supported by University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province (UNPYSCT-2017210), Scientific Research Projects at Harbin University of Commerce (17XN007), Doctoral Science Foundation of Harbin University of Commerce (14LG06); Graduate Student Innovation Research Project at Harbin University of Commerce (YJSCX2018-545HSD), Heilongjiang Natural Science Fund (H2017001) and Heilongjiang Postdoctoral Foundation (LBH-Z16095).

Conflicts of interest

The authors declare that there are no conflicts of interest related to this article.

References

Chao G, Sun J, Wang C, Dong Y, Xiao S, Wang X, Jiao Z (2017). Genome-wide analysis of basic/helix-loop-helix gene family in peanut and assessment of its roles in pod development. PLoS One 12(7):e0181843.

Chu Y, Xiao SM, Su H, Liao BS, Zhang JJ, Xu J, Chen SL (2018). Genome-

wide characterization and analysis of bHLH transcription factors in *Panax ginseng*. Acta Pharmaceutica Sinica B 8(4):666-677.

Conesa A, Gotz S (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. International Journal of Plant Genomics 2008:1-12.

Gantet P, Memelink J (2002). Transcription factors: tools to engineer the production of pharmacologically active plant metabolites. Trends in Pharmacological Sciences 23(12):563-569.

Goodrich J, Carpenter R, Coen ES (1992). A common gene regulates pigmentation pattern in diverse plant species. Cell 68(5):955-964.

Guo XJ, Wang JR (2017). Global identification, structural analysis and expression characterization of bHLH transcription factors in wheat. BMC Plant Biology 17(1):90.

Guruprasad K, Reddy BVB, Pandit MW (1990). Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence. Protein Engineering, Design and Selection 4(2):155-161.

Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC (2003). The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. Molecular Biology and Evolution 20(5):735-747.

Hong GJ, Xue XY, Mao YB, Wang LJ, Chen XY (2012). *Arabidopsis* MYC2 interacts with DELLA proteins in regulating sesquiterpene synthase gene expression. Plant Cell 24(6):2635-2648.

Jaakola L, Pirttila AM, Halonen M, Hohtola A (2001). Isolation of high quality RNA from bilberry (*Vaccinium myrtillus L.*) fruit. Molecular Biotechnology 19(2):201-203.

Jue DW, Sang XL, Liu LQ, Shu B, Wang YC, Liu CM, ... Shi SY (2018). Identification of WRKY gene family from *Dimocarpus longan* and its expression analysis during flower induction and abiotic stress responses. International Journal of Molecular Sciences 19(8):2169.

Kyte J, Doolittle RF (1982). A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology 157(1):105-132.

Li X, Duan X, Jiang H, Sun Y, Tang Y, Yuan Z, ... Yin J (2006). Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and *Arabidopsis*. Plant Physiology 141(4):1167-1184.

Lin YL, Min JM, Lai RL, Wu ZY, Chen YK, Yu LL, ... Lai ZX (2017). Genome-wide sequencing of longan (*Dimocarpus longan* Lour.) provides insights into molecular basis of its polyphenol-rich characteristics. Gigascience 6(5):1-14.

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 86(18):7092-7096.

Massari ME, Murre C (2000). Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. Molecular and Cellular Biology 20(2):429-440.

Niu X, Guan Y, Chen S, Li H (2017). Genome-wide analysis of basic helix-loop-helix (bHLH) transcription factors in *Brachypodium distachyon*. BMC Genomics 18(1):619.

Pires N, Dolan L (2010). Origin and diversification of basic-helix-loop-helix proteins in plants. Molecular Biology and Evolution 27(4):862-874.

Schmittgen TD, Livak KJ (2008). Analyzing real-time PCR data by the comparative CT method. Nature Protocols 3(6):1101-1108.

Sun X, Wang Y, Sui N (2018). Transcriptional regulation of bHLH during plant response to stress. Biochemical and Biophysical Research Communications 503(2):397-401.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28(10):2731-2739.

Toledo-Ortiz G, Huq E, Quail PH (2003). The Arabidopsis Basic/Helix-Loop-Helix transcription factor family. The Plant Cell 15(8):1749-1770.

Wang J, Hu Z, Zhao T, Yang Y, Chen T, Yang M, ... Zhang B (2015a). Genome-wide analysis of bHLH transcription factor and involvement in the infection by yellow leaf curl virus in tomato (*Solanum lycopersicum*). BMC Genomics 16(1):39.

Wu P, Song XM, Wang Z, Duan WK, Hu R, Wang WL, ... Hou XL (2016b). Genome-wide analysis of the BES1 transcription factor family in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). Plant Growth Regulation 80(3):291-301.

Xing BC, Yang DF, Yu HZ, Zhang BX, Yan KJ, Zhang XM, ... Liang ZS (2018a). Overexpression of *SmbHLH10* enhances tanshinones biosynthesis in *Salvia miltiorrhiza* hairy roots. Plant Science 276:229-238.

Xing BC, Liang LJ, Liu L, Hou ZN, Yang DF, Yan KJ, ... Liang ZS (2018b). Overexpression of *SmbHLH148* induced biosynthesis of tanshinones as well as phenolic acids in *Salvia miltiorrhiza* hairy roots. Plant Cell Reports 37(12):1681-1692.

Xue Y, Wang W, Liu Y (2015). Two new flavonol glycosides from *Dimocarpus longan* leaves. Natural Product Research 29(2):163-168.

Yamamura C, Mizutani E, Okada K, Nakagawa H, Fukushima S, Tanaka A, ... Mori M (2015). Diterpenoid phytoalexin factor, a bHLH transcription factor, plays a central role in the biosynthesis of diterpenoid phytoalexins in rice. Plant Journal 84(6):1100-1113.

Yamasaki K, Kigawa T, Seki M, Shinozaki K, Yokoyama S (2013). DNA-binding domains of plant-specific transcription factors: structure, function, and evolution. Trends in Plant Science 18(5):267-276.

Yang CQ, Fang X, Wu XM, Mao YB, Wang LJ, Chen XY (2012). Transcriptional regulation of plant secondary metabolism. Journal of Integrative Plant Biology 54(10):703-712.

Yu J, Ai G, Shen DY, Chai CY, Jia YL, Liu WJ, ... Dou DL (2019). Bioinformatical analysis and prediction of *Nicotiana benthamiana* bHLH transcription factors in *Phytophthora parasitica* resistance. Genomics 111(3):473-482.

Zhang TT, Lv W, Zhang HS, Ma L, Li PH, Ge L, Li G (2018). Genome-wide analysis of the basic Helix-Loop-Helix (bHLH) transcription factor family in maize. BMC Plant Biology 18(1):235.

Zhang X, Luo H, Xu Z, Zhu Y, Ji A, Song J, Chen S (2015). Genome-wide characterization and analysis of bHLH transcription factors related to tanshinone biosynthesis in *Salvia miltiorrhiza*. Scientific Reports 5:11244.