

Genome-Wide Analysis of Basic Helix-Loop-Helix (bHLH) TFs in Sea Buckthorn (*Hippophae rhamnoides*)

Songfeng Diao (✉ diaosongfeng@outlook.com)

Chinese Academy of Forestry

Hong Liu

Chinese Academy of Medical Sciences & Peking Union Medical College Hospital of Skin Diseases and Institute of Dermatology

Zhongrui Lv

Chinese Academy of Forestry

Caiyun He

Chinese Academy of Forestry

Aiguo Duan

Chinese Academy of Forestry

Guoyun Zhang

Chinese Academy of Forestry

Jianguo Zhang

Chinese Academy of Forestry

Research article

Keywords: Hippophae rhamnoides, Genome-wide analysis, Basic helix-loop-helix TFs, Expression analysis, Fruit Development

Posted Date: August 21st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-53464/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

The basic helix-loop-helix (bHLH) transcription factor gene family is one of the largest gene families and extensively involved in plant growth, organ development, and stress responses. However, limited studies are available on the gene family in sea buckthorn.

Results

In this study, we focused on 144 HrbHLH genes, exploring their DNA and protein sequences and physicochemical properties. According to their protein sequence similarities, we classified the genes into 15 groups with specific motif structures. In order to explore their expressions, we performed gene expression profiling using RNA-Seq and identified 108 HrbHLH genes that expressed in five sea buckthorn tissue, including root nodule, root, leaf, stem and fruit. Furthermore, we found 11 increased expressed HrbHLH genes during sea buckthorn fruit development. We validated the expression pattern of HrbHLH genes using reverse transcription quantitative real-time PCR.

Conclusions

This study lays the foundation for future studies on gene cloning, transgenes, and biological mechanisms. We performed a genome-wide, systematic analysis of bHLH proteins in sea buckthorn. This comprehensive analysis provides a useful resource that enables further investigation of the physiological roles and molecular functions of the HrbHLH TFs.

1. Background

The basic helix-loop-helix (bHLH) TFs (TFs) family is widely existed in eukaryotes and plays an important role in plant growth and development [1]. However, limited studies are available in plants, especially in sea buckthorn. Since the first bHLH protein structure was analyzed [2], more than 100 bHLH proteins have been identified in various species. The bHLH TFs family is named for its highly conserved bHLH domain, which consists of about 60 amino acids, including the basic region distributed at the N-terminus and the helix-loop-helix (HLH) region distributed at the C-terminus of the polypeptide chain [3]. The basic region contains about 15 amino acids and plays an important role in recognizing and binding to the target DNA. The HLH region, ranging from 40 to 50 amino acids in length, consists of two amphiphilic alpha helices and a loop structure with uncertain length and sequence. The two amphiphilic alpha helices can form homodimer or heterodimeric to interact with other bHLH proteins. The bHLH TFs can identify the element called E-box (5'-CANNTG-3'), with the most common one is G-box (5'-CACGTG-3'). Studies have shown that the nucleotides in the flank of the external of core element also impact specific binding. Previous studies indicated that the known bHLH proteins in animals can be divided into six

groups (A–F) [1]. However, many of the identified bHLH proteins belong to group B in plants, the most members of which are characterized by binding to G-box [4].

The bHLH TFs have various functions in plant growth and development. The bHLH genes *SPATULA* and *ALCATRAZ* impact the development of carpel and fruit [5, 6], and MYC2, MYC3, while the MYC4 affects the flowering of *Arabidopsis thaliana* [7]. The bHLH gene *BIGPETALp* in *A. thaliana* plays an important role in controlling petal size [8], and the bHLH gene *BIM1* is associated with embryonic patterning [9]. *MdbHLH3* gene promotes anthocyanin accumulation and fruit coloration in apple [10]. *SIPRE2* affects plant morphology and fruit pigment accumulation in tomato [11]. The bHLH genes also regulate plant response to various abiotic stresses, such as drought [12–14], salinity [15–17], cold [18–20], aluminum and iron deficiency [21–23]. In addition, the bHLH genes play an important role in regulating multiple signal transduction pathways and impacting biosynthesis [24, 25].

Sea buckthorn (*Hippophae rhamnoides*) is a hardy, fast growing, deciduous, spiny shrub, naturally distributed in Asia and Europe [26–28]. Sea buckthorn fruit and leaves have high levels of vitamins C, E and A, organic acids, amino acids, fatty acids, carotenoids and flavonoids [29–31]. In this study, we focus on 144 bHLH genes in sea buckthorn, in order to investigate their structures and functions, especially differential gene expression patterns in different tissue and different fruit development stages.

2. Methods

2.1 Identification and Characterization of the Sea Buckthorn bHLH Gene Family

We downloaded the genome-wide protein sequences of sea buckthorn from *H. rhamnoides* Information Archive database. The Hidden Markov Model (HMM) profile of the bHLH domain (PF00010) was downloaded from the Pfam database [35], and was used as a query to scan the proteome file via HMMER software (version 3.1) with a default E-value [36]. The protein sequences for bHLH genes shown in those HMMER results were obtained from the proteome file using TBTools (<https://www.tbtools.com/>). Redundant sequences were removed with online ElimDupes software, and a few sequences with obvious errors were removed manually.

The GFF file, containing location data and annotation information for sea buckthorn genes, was downloaded from the *H. rhamnoides* Information Archive database. MapInspect software was used to map the sea buckthorn bHLH genes on different chromosomes, and annotation data in the GFF file were exhibited by the online tool WEGO2.0 [37]. The lengths, masses, isoelectric point (PI)-values, and charge at pH7.0 for these bHLH protein sequences were determined with DNASTAR software, and length distributions and functional annotations were analyzed with Blast2GO software [38].

2.2 Multiple Sequence Alignment and Phylogenetic Analysis

A total of 144 predicted HrbHLH proteins, with amino acids spanning the bHLH core domain, were subjected to a multiple sequence alignment using ClustalX 2.0 with the default parameters and the further multiple sequence alignment was performed using ClustalW 2.0 [39]. The phylogenetic tree representing HrbHLH proteins was generated using MEGA 7.0 software and the Maximum likelihood method, with the following settings: mode, “p-distance”; gap setting, “Complete Deletion”; and bootstrap test replicate, ‘1,000’ [40].

2.3 Gene Structure and Conserved Motifs Analyses of HrbHLH Genes

Information about the gene structure (intron-exon) of each putative bHLH gene was obtained from the GFF file, downloaded from the GDR database. The schematic structures of these genes were drawn with the online Gene Structure Display Server (GSDS2.0) [41]. Local MEME software was used to identify conserved motifs in the protein sequences according to the following parameters: -protein, -oc m12, -mod zoops, -nmotifs 12, -minw 6, and -maxw 70 [42]. The results from these analyses of gene structure and conserved motifs were arranged according to the order shown on the phylogenetic tree.

2.4 3D Protein Homology Modeling and Protein Properties

First of all, BLASTP search with the default parameters was performed in the Protein Data Bank (PDB) with all bHLH proteins for identification of the best template having similar sequence and known three-dimensional structure. Using ‘intensive’ mode in Protein Homology/Analogy Recognition Engine (Phyre2) [43], the data was analyzed for prediction of protein structure of sea buckthorn bHLHs. The theoretical isoelectric point (PI) and protein statistics were analyzed using ExPASy and Sequence Manipulation Suite, respectively.

2.5 Expression Analysis of HrbHLH Genes in Sea Buckthorn

To evaluate the sea buckthorn bHLH gene expression patterns, we downloaded the Illumina RNA-seq data of sea buckthorn different tissues from *H. rhamnoides* Information Archive database, including root nodule, root, leaf, stem and fruit tissues. First, all these raw data were spitted and converted to fastq format file by the NCBI SRA Toolkit’s fastq-dump command. The quality of fastq files were evaluated with FASTQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Also, low-quality reads were trimmed with local perl script. After final quality check, RNA-seq sequence datasets were aligned to the whole sea buckthorn genome by using TopHat2 [44]. Normalization of the gene expression values were carried out by the fragments per kilobase of exon model per million mapped reads (FPKM) algorithms. The absolute ratio of \log_2 foldchange ≥ 1 and FDR-value ≤ 0.001 were used for the identification of differentially expressed bHLH genes. The heatmaps of hierarchical clustering were visualized with OmicShare tools (<http://www.omicshare.com/tools>).

2.6 Plant Materials and Growth Conditions

The *H. rhamnoides* L. subsp. *Mongolica Rousi* was used in this study. All these sea buckthorns were planted in the desert Forest Experimental Center in Inner Mongolia, China. The records of planting,

preservation and identification of all plant materials are completed and archived by this center, and the management of all experimental materials in this center is supervised and managed by China National Laboratory Accreditation Committee (CNACL). Healthy fresh sea buckthorn fruits were harvested at 30, 37, 42, 49, 56 and 63 days post-anthesis, and quickly placed in the liquid nitrogen, and stored at -80°C until next use.

2.7 RNA Extraction and Real-Time Quantitative qRT-PCR Analysis

Plant materials were harvested, frozen in liquid nitrogen, and then ground under RNase-free conditions. The RNA was extracted with TRizol reagent (Invitrogen), following the manufacturer's instructions, and then treated with DNase I at 37°C for 30 min. The RNA ($1\ \mu\text{g}$) was then reverse-transcribed using a PrimeScript First-strand cDNA Synthesis Kit (Takara, Dalian, China) according to the manufacturer's instructions. A $10\text{-}\mu\text{L}$ aliquot of cDNA was diluted to $100\ \mu\text{L}$ with water, and $2\ \mu\text{L}$ of that diluted cDNA was used for the analyses. For real-time quantitative qRT-PCR, gene specific primers for our selected HrbHLHs were designed and synthesized by Sangon Biotech (Shanghai, China) Company (product size 110–130 bp; T_m $59\text{--}61^{\circ}\text{C}$; details are shown in **Supplementary Table S1**), and Hr18S was used as an internal control. All reactions were performed on an Icyler iQ5 system (Bio-Rad), using the SYBR Green Supermix Kit (Bio-Rad) according to the manufacturer's instruction. Expression levels of these genes were calculated as $2^{\Delta\text{CT}}$ values. Besides, relative expression levels for each gene along the time series were also calculated as $2^{\Delta\Delta\text{CT}}$ values. At least three biological replicates were used for the fluorescence-quantitative PCR reactions, with each biological repeat having at least three technical replicates. Each biological repeat contains at least 6 plantlets for mixing.

3. Results

3.1 Identification, Chromosomal Locations, and Functional Annotation of Sea Buckthorn bHLH Genes

For genome-wide identification of HrbHLH genes, we used the HMM file as a query to search the sea buckthorn proteome. A total of 180 putative HrbHLH protein sequences were obtained. After the existence of the conserved bHLH domain was confirmed by SMART and CD-Search and redundant sequences were removed, we finally identified 144 bHLH proteins in the sea buckthorn genome. Based on their chromosomal locations, these sea buckthorn bHLH genes were named from HrbHLH1 to HrbHLH144 (Fig. 1). Sequence analysis revealed that these HrbHLH proteins vary widely in length and have an average of 326.5 aa. The most length of HrbHLH proteins was HrbHLH76 (758 aa), and the least length of HrbHLH proteins was HrbHLH129 (92 aa). Their predicted molecular weights ranged from 10.23 kDa to 81.93 kDa, with average of 36.47 kDa. Their predicted pI values ranged from 4.50 to 10.61. Gene IDs, genomic positions, and annotations information were also summarized for these HrbHLH proteins (**Supplementary Table S1**).

According to the genomic position of putative HrbHLH genes, we showed that the HrbHLH genes were found across all the 12 chromosomes, ranging from 1 to 21 per chromosome (Fig. 1). However, the other 4 bHLHs (HrbHLH1-HrbHLH4) were localized to unassembled genomic contigs, and cannot be mapped to any particular chromosome according to what we currently know about this genome (Fig. 1).

Chromosome 3 and 12 has the most HrbHLHs (21 total), followed by chr 11 (17 genes) and chr 10 (16 genes). The sea buckthorn genome database provided GO and KEGG annotation information about these HrbHLH proteins. Among these HrbHLH, three HrbHLH genes (*HrbHLH7*, *HrbHLH133*, *HrbHLH39*) were related with organ development. In addition, two HrbHLH genes (*HrbHLH120* and *HrbHLH107*) were enriched into plant hormone signal transduction pathway.

3.2 Phylogenetic Analysis and Prediction of Conserved Motifs

The exact number of subgroup classifications for plant bHLH proteins is unknown, but is thought to be 15–32. To evaluate the evolutionary relationships of our 144 proteins, we conducted a phylogenetic analysis based on full-length protein sequences. Applying the ML method, we assigned the proteins to 15 main groups (G1-G15) (Fig. 2). The average size of the groups has approximately 9 members, ranging from 4 to 19. G15 group have the least members, including HrbHLH65, HrbHLH59, HrbHLH11, and HrbHLH45.

We identified 8 putative conserved protein motifs in HrbHLH family proteins using MEME online software (Motifs 1–8, Fig. 3). All HrbHLH family proteins contained Motif 1 and Motif 2, while HrbHLH in group G6 and G7 didn't have Motif 1. Also, all HrbHLHs in group G6 and G7 contained Motif 8. In each group, the components of the conserved motifs for most of the proteins were similar. For example, Motifs 1, 2, and 6 were identified in all 11 members of group G2, and Motifs 1, 2, and 3 were identified in all 19 members of group G1, and Motifs 4 were involved in 11 members of group G2. The evolutionary relationships among these HrbHLH proteins were also determined by analyzing their conserved motifs. These composition patterns tended to be consistent with the results from our phylogenetic tree, being nearly identical among genes within the same group, but varying greatly between groups.

3.3 Analysis of Gene Structures

The pattern of exon and intron positions can also provide important evidence to support phylogenetic relationships in a gene family. Gene structure analysis of the 144 HrbHLHs was performed to gain some information about exon-intron organization (**Supplementary Figure S1**). Among these 144 sea buckthorn genes, a total of 7 bHLH genes (*HrbHLH26*, *HrbHLH48*, *HrbHLH52*, *HrbHLH58*, *HrbHLH63*, *HrbHLH67*, *HrbHLH109*) without intron were found, which accounts for 5% of total HrbHLH genes. All these intronless genes were clustered into the group G3. According to exon-intron organization of bHLH genes in sea buckthorn, phylogenetically related proteins exhibited a closely related gene structure, in terms of intron number or exon length.

3.4 3D Homology Modeling of HrbHLH Proteins

To construct 3D homology model for bHLH proteins, BLASTP search was performed against the PDB. 15 HrbHLH proteins with high homology rate were selected. To predict homology modeling in Phyre2, which utilizes the alignment of hidden Markov models via HMM-HMM search, the detection rate method was used. The intensive mode was selected in Phyre2 to increase accuracy of alignment. In addition, it integrates a new abinitio folding simulation termed as poing to model areas of proteins without any significant homology for known structures. The 3D protein modelling of selected 15 bHLH proteins are predicted at > 90% confidence and the percentage residue varied from 80 to 100 (Fig. 4). The secondary structures were predominantly constituted of α helices and having rare incurrence of β sheets. Thus, all suggested protein structures are assessed to be highly reliable which offers a preliminary basis for understanding the molecular function of HrbHLH proteins.

3.5 Expression Patterns of HrbHLH Genes in Different Tissues

To further understand the function of sea buckthorn bHLH proteins, the expression patterns of sea buckthorn bHLH genes among 21 samples, including 5 tissues, were analyzed, according to the normalized RPKM data from RNA-seq. **Supplementary Table S2** shows the expression profiles of all bHLH genes in the 21 sea buckthorn samples. Among the 144 HrbHLH genes, 122 HrbHLH mRNAs had an RPKM value greater than 1 in at least one of the 21 samples, while the remaining 22 HrbHLH genes were expressed at very low levels in all 21 samples. Among these HrbHLH genes, *HrbHLH47*, *HrbHLH74*, *HrbHLH90*, *HrbHLH131* were the highest expression level in root nodule, root, leaf, stem and fruit tissues, respectively.

In particular, *HrbHLH9*, *HrbHLH42*, *HrbHLH69*, *HrbHLH74*, *HrbHLH75*, *HrbHLH85*, *HrbHLH90* and *HrbHLH131* were constitutively produced at a relatively high level in all 21 samples, suggesting that these four bHLH genes perform a variety of functions in different tissues (Fig. 5). Furthermore, four HrbHLH genes showed preferential tissue-specific expression, including one gene in root nodule (*HrbHLH35*), two genes in fruits (*HrbHLH43* and *HrbHLH58*), one gene in stem (*HrbHLH63*) (Fig. 6). The specific accumulation of these bHLH genes in a particular tissue suggests that they may play conserved regulatory roles in discrete cells, organs, or conditions.

3.6 Expression Profiling for HrbHLH Genes during Sea Buckthorn Fruit Development Process

To analyze the expression pattern of HrbHLH genes among different development stages, we used the Fragments Per Kilobase per Million (FPKM) normalized data from RNA-seq. Additional file showed the expression profiles of HrbHLH genes in sea buckthorn fruit different development stages. Among 144 HrbHLHs genes, 131 expressed at least in one fruit development stage, while the rest HrbHLHs genes not expressed in sea buckthorn fruit. Based on gene expression data, we identified 11 HrbHLH genes (*HrbHLH27*, *HrbHLH54*, *HrbHLH70*, *HrbHLH79*, *HrbHLH91*, *HrbHLH92*, *HrbHLH100*, *HrbHLH104*,

HrbHLH111, *HrbHLH127*, *HrbHLH132*) which showed a gradually decreased expression with fruit development and ripening, suggesting that they may function in fruit ripening. Among them, *HrbHLH54*, *HrbHLH91* and *HrbHLH92* exhibited a high expression in sea buckthorn ripe fruit, and the rest bHLH gene have low expression level in sea buckthorn fruit (Fig. 7). Furthermore, we identified 54 *HrbHLH* genes which showed a gradually increased expression with fruit development and ripening. Compared with down-regulated bHLH genes, these up-regulated bHLH genes have lower expression level. And, the expression level of 16 *HrbHLH* genes were confirmed with qRT-PCR (Fig. 8). Therefore, further characterization of the 11 *HrbHLHs* is highly important and will provide a new insight to understand the molecular mechanism of fruit development and ripening.

4. Discussion

Sea buckthorn is an important economic species of the *Elaeagnus* family [29]. The fruit of sea buckthorn is rich in nutrients, which can provide essential amino acids and have antioxidant activity [30]. With the finish of the sea buckthorn genome sequencing, we can identify and characterize transcription factor gene family at the whole genome level. Studying the transcription factor family of sea buckthorn is great significance for breeding excellent cultivars and improving fruit quality. However, no such detailed studies have been done with the bHLH family in sea buckthorn.

The second largest TF family across eukaryotic kingdoms is the bHLH superfamily [3]. Based on the highly conserved bHLH domain, the bHLH superfamily is categorized into two main groups: the basic region and the HLH region. The basic region is a DNA-binding interface located next to the N-terminus of the protein domain, whereas the HLH region serves as the dimerization domain and consists of two amphipathic α -helices linked by a loop [32]. Apart from the high frequency of bHLH conserved domains, some other motifs also occur within the bHLH superfamily. Based on the highly conserved domains and their evolutionary relationships, this superfamily is mainly divided into 20–25 subfamilies in agronomically important plants. A wide range of extensive studies have been reported on the bHLH superfamily-related genes in plant species after genome sequences have been produced, including cucumber [5], *Arabidopsis* [7], apple [10], tomato [24] and *Carthamus tinctorius* [32]. Additionally, functional and structural characterizations of most plant bHLH proteins have been described in detail. These findings demonstrated that bHLH proteins are involved in various biochemical and physiological networks across the plant kingdom. Studies on bHLH revealed its partial involvement in iron uptake, response against salt and drought stress, tanshinone biosynthesis, and petal growth development. The most prominent studies on the bHLH family demonstrated its association with anthocyanin biosynthesis and regulation in flowering plants and fruits [33].

Previous studies have suggested that the bHLH TFs were usually divided into 25 subfamilies in many plants. In this study, we identified 144 *HrbHLH* genes in sea buckthorn genome, which was less than bHLH genes in tomato and apple genome. Based on phylogenetic analysis, predicted conserved protein motifs and intron-exon organizations, all identified *HrbHLHs* were divided into 15 groups, which were different with other species [1, 3]. These results indicate that sea buckthorn bHLHs have different

classification model with other plants. Furthermore, our intron-exon gene structure and conserved protein motifs results strongly supported our classifications of the HrbHLHs.

To further understand the function of sea buckthorn bHLH proteins, the expression patterns of sea buckthorn bHLH genes among 21 samples, including 5 tissues and 3 different developmental stages, were analyzed and normalized RPKM data from RNA-seq. **Supplementary Table S1** shows the expression profiles of all bHLH genes in the 21 sea buckthorn tissues. Among the 144 HrbHLH genes, 122 HrbHLH mRNAs had an RPKM value greater than 1 in at least one of the 21 samples, while the remaining 22 HrbHLH genes were expressed at very low levels in all 21 samples. Among these 122 HrbHLH genes, the expression level of 46 HrbHLH were more than 1 in all five tissues. In particular, HrbHLH131, HrbHLH9, HrbHLH42, HrbHLH69 and HrbHLH85 were constitutively produced at a relatively high level in all 21 samples, suggesting that these five bHLH genes perform a variety of functions in different tissues [13]. Furthermore, four bHLH genes showed preferential tissue-specific expression, including one gene (HrbHLH35) in leaf, two genes (HrbHLH43 and HrbHLH58) in fruit, one gene (HrbHLH63) in stem. The specific accumulation of these bHLH genes in a particular tissue suggests that they may play conserved regulatory roles in discrete cells, organs, or conditions [20, 34].

Recently, more and more researches reveal that bHLH gene play important role in fruit development process, including tomato [11], peach [19], apple [20], almond [33] and Chinese jujube [34]. In this study, 91 HrbHLH mRNAs had an RPKM value greater than 1 in three fruit development stages. HrbHLH47 have a high expression level across whole fruit development process. *HrbHLH47* is the homolog of AtPRE5, a key factor regulating cell elongation and plant development. The homologous comparison and protein interaction prediction also indicated that it might be involved in fruit enlargement [10–11]. The expression *HrbHLH90* was significantly higher at the early stage of fruit development. This result also indicated that it might be involved in fruit enlargement [26, 33]. In addition, the expression of *HrbHLH91* and *HrbHLH92* were significantly higher at the ripen stage of fruit development.

5. Conclusions

Our study provides the first all-inclusive and systematic genome-wide analysis of the sea buckthorn bHLH superfamily. As a result, a set of 144 genes was characterized and classified into 15 subfamilies. Protein motifs, compositions, and their amino acid ratios were also thoroughly investigated. Gene expression analysis results provided preliminary information for future investigation of the sea buckthorn bHLH gene family during flavonoid metabolism. The current investigation lays the highlights for functional characterization and confirmation of the sea buckthorn bHLH gene superfamily and improves our understanding of the bHLH gene superfamily in higher plants.

Abbreviations

helix-loop-helix (bHLH)

Hidden Markov Model (HMM)

Protein Data Bank (PDB)

fragments per kilobase of exon model per million (FPKM)

China National Laboratory Accreditation Committee (CNACL)

Declarations

Ethics approval and consent to participate

Not applicable.

Consent to publish

Manuscript is approved by all authors for publication.

Availability of data and materials

The data used in this experiment are all downloaded from available public databases and analyzed by software. Usually these data do not need to be repeatedly provided to third parties because they can be easily obtained. Any reader who needs all the software analysis results can get it by contacting the corresponding author.

Competing interests

The authors declare no conflict of interest.

Funding

This research was supported financially by the special fund from Key Laboratory of Tree Breeding and Cultivation, National Forestry and Grassland Administration (ZDRIF2019). J.G.Z. hold the funding, and play the role of writing review and editing in the whole progress.

Authors' Contributions

Methodology, S.F.D. and H.L.; data curation, Z.R.L and A.G.D.; writing-original draft preparation, S.F.D. and H.L.; writing-review and editing, C.Y.H. and J.G.Z.; supervision, G.Y.Z. and J.G.Z.; project administration, G.Y.Z., C.Y.H. and J.G.Z.; funding acquisition, G.Y.Z. All authors have read and approved the manuscript.

Acknowledgements

Not applicable.

References

1. Carretero-Paulet L, Galstyan A, Roig-Villanova I, Martinez-Garcia JF, Bilbao-Castro JR, Robertson DL. Genome-wide classification and evolutionary analysis of the bHLH family of TFs in *Arabidopsis*, poplar, rice, moss, and algae. *Plant Physiol.* 2010;153:1398–412.
2. Ludwig SR, Wessler Sf. Maize R Gene Family: TissueSpecific Helix-loop-helix proteins. *Cell.* **1990**, *62*, 649–851.
3. Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC. The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. *Mol Biol Evol.* 2003;20:735–47.
4. 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.
5. Xiao X, Lv J, Xie J, Feng Z, Ma N, Li J, Yu J, Calderón-Urrea A. Transcriptome analysis reveals the different response to toxic stress in rootstock grafted and non-grafted cucumber seedlings. *Int J Mol Sci.* 2020;21:774.
6. Groszmann M, Paicu T, Alvarez JP, Swain SM, Smyth DR. SPATULA and ALCATRAZ, are partially redundant, functionally diverging bHLH genes required for *Arabidopsis* gynoecium and fruit development. *Plant J.* 2011;68:816–29.
7. Wang H, Li Y, Pan J, Lou D, Hu Y, Yu D. The bHLH TFs MYC2, MYC3, and MYC4 are required for Jasmonate-Mediated inhibition of flowering in *Arabidopsis*. *Mol Plant.* 2017;10:1461–4.
8. Szecsi J, Joly C, Bordji K, Varaud E, Cock JM, Dumas C, Bendahmane M. BIGPETALp, a bHLH transcription factor is involved in the control of *Arabidopsis* petal size. *EMBO J.* 2006;25:3912–20.
9. Chandler JW, Cole M, Flier A, Werr W. BIM1, a bHLH protein involved in brassinosteroid signalling, controls *Arabidopsis* embryonic patterning via interaction with DORNROSCHEN and DORNROSCHEN-LIKE. *Plant Mol Biol.* 2009;69:57–68.
10. Xie XB, Li S, Zhang RF, Zhao J, Chen YC, Zhao Q, Yao YX, You CX, Zhang XS, Hao YJ. The bHLH transcription factor MdbHLH3 promotes anthocyanin accumulation and fruit colouration in response to low temperature in apples. *Plant Cell Environ.* 2012;35:1884–97.
11. Zhu Z, Chen G, Guo X, Yin W, Yu X, Hu J, Hu Z. Overexpression of SIPRE2, an atypical bHLH transcription factor, affects plant morphology and fruit pigment accumulation in tomato. *Sci Rep.* 2017;7:5786.
12. Li Z, Liu C, Zhang Y, Wang B, Ran Q, Zhang J. The bHLH family member ZmPTF1 regulates drought tolerance in maize by promoting root development and abscisic acid synthesis. *J Exp Bot.* 2019;70:5471–86.

13. Yao PF, Li CL, Zhao XR, Li MF, Zhao HX, Guo JY, Cai Y, Chen H, Wu Q. Overexpression of a tartary buckwheat gene, FtbHLH3, enhances drought/oxidative stress tolerance in transgenic *Arabidopsis*. *Front Plant Sci.* 2017;8:625.
14. Waseem M, Rong X, Li Z. Dissecting the role of a basic helix-loop-helix transcription factor, SlbHLH22, under salt and drought stresses in transgenic *Solanum lycopersicum* L. *Front Plant Sci.* 2019;10:734.
15. Zhao K, Li S, Yao W, Zhou B, Li R, Jiang T. Characterization of the basic helix-loop-helix gene family and its tissue-differential expression in response to salt stress in poplar. *PeerJ.* **2018**, 6, e4502.
16. Chen HC, Cheng WH, Hong CY, Chang YS, Chang MC. The transcription factor OsbHLH035 mediates seed germination and enables seedling recovery from salt stress through ABA-dependent and ABA-independent pathways, respectively. *Rice.* 2018;11:50.
17. Samira R, Li B, Kliebenstein D, Li C, Davis E, Gillikin JW, Long TA. The bHLH transcription factor ILR3 modulates multiple stress responses in *Arabidopsis*. *Plant Mol Biol.* 2018;97:297–309.
18. Yao P, Sun Z, Li C, Zhao X, Li M, Deng R, Huang Y, Zhao H, Chen H, Wu Q. Overexpression of *Fagopyrum tataricum* FtbHLH2 enhances tolerance to cold stress in transgenic *Arabidopsis*. *Plant Physiol Biochem.* 2018;125:85–94.
19. Jin C, Huang XS, Li KQ, Yin H, Li LT, Yao ZH, Zhang SL. Overexpression of a bHLH1 transcription factor of *Pyrus ussuriensis* confers enhanced cold tolerance and increases expression of stress-responsive genes. *Front Plant Sci.* 2016;7:441.
20. Xu H, Wang N, Wang Y, Jiang S, Fang H, Zhang J, Su M, Zuo W, Xu L, Zhang Z, et al. Overexpression of the transcription factor MdbHLH33 increases cold tolerance of transgenic apple callus. *Plant Cell Tissue Organ Cult.* 2018;134:131–40.
21. Li X, Zhang H, Ai Q, Liang G, Yu D. Two bHLH TFs, bHLH34 and bHLH104, regulate iron homeostasis in *Arabidopsis thaliana*. *Plant Physiol.* 2016;170:2478–93.
22. Colangelo EP, Guerinot ML. The essential basic helix-loop-helix protein FIT1 is required for the iron deficiency response. *Plant Cell.* 2004;16:3400–12.
23. Zhang J, Liu B, Li M, Feng D, Jin H, Wang P, Liu J, Xiong F, Wang J, Wang HB. The bHLH transcription factor bHLH104 interacts with IAA-LEUCINE RESISTANT3 and modulates iron homeostasis in *Arabidopsis*. *Plant Cell.* 2015;27:787–805.
24. Gao Y, Liu J, Chen Y, Tang H, Wang Y, He Y, Ou Y, Sun X, Wang S, Yao Y. Tomato SIAN11 regulates flavonoid biosynthesis and seed dormancy by interaction with bHLH proteins but not with MYB proteins. *Hortic Res.* 2018;5:27.
25. Yang T, Yao S, Hao L, Zhao Y, Lu W, Xiao K. Wheat bHLH-type transcription factor gene TabHLH1 is crucial in mediating osmotic stresses tolerance through modulating largely the ABA-associated pathway. *Plant Cell Rep.* 2016;35:2309–23.
26. Zhang G, Diao S, Zhang T, Chen D, He C, Zhang J. Identification and characterization of circular RNAs during the sea buckthorn fruit development. *RNA Biol.* 2019;16:354–61.

27. Zhang G, Chen D, Zhang T, Duan A, Zhang J, He C. Transcriptomic and functional analyses unveil the role of long non-coding RNAs in anthocyanin biosynthesis during sea buckthorn fruit ripening. *DNA Res.* 2018;25:465–76.
28. Zhang G, Duan A, Zhang J, He C. Genome-wide analysis of long non-coding RNAs at the mature stage of sea buckthorn (*Hippophae rhamnoides* Linn) fruit. *Gene.* 2017;596:130–6.
29. TIITINEN KM, HAKALA MA, KALLIO HP. Quality components of sea buckthorn (*Hippophae1 rhamnoides*) varieties. *J Agric Food Chem.* 2005;53:1692–9.
30. TIITINEN KM, YANG B, HARALDSSON GG, JONSDOTTIR S, KALLIO HP. Fast analysis of sugars, fruit acids, and Vitamin C in Sea Buckthorn (*Hippophae1 rhamnoides* L.) varieties. *J Agric Food Chem.* 2006;54:2508–13.
31. Pop RM, Weesepeol Y, Socaciu C, Pintea A, Vincken JP, Gruppen H. Carotenoid composition of berries and leaves from six Romanian sea buckthorn (*Hippophae rhamnoides* L.) varieties. *Food Chem.* 2014;147:1–9.
32. Hong Y, Ahmad N, Tian Y, Liu J, Wang L, Wang G, Liu X, Dong Y, Wang F, Liu W, Li X, Zhao X, Yao N, Li H. Genome-wide identification, expression analysis, and subcellular localization of *Carthamus tinctorius* bHLH TFs. *Int J Mol Sci.* 2019;20:3044.
33. Sánchez-Pérez R, Pavan S, Mazzeo R, Moldovan C, Cigliano RA, Del Cueto J, ... López-Marqués RL. Mutation of a bHLH transcription factor allowed almond domestication. *Science.* 2019;364:1095–8.
34. Li H, Gao W, Xue C, Zhang Y, Liu Z, Zhang Y, Meng X, Liu M, Zhao J. Genome-wide analysis of the bHLH gene family in Chinese jujube (*Ziziphus jujuba* Mill.) and wild jujube. *BMC Genomics.* 2019;20:568.
35. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 2016;44:D279–85.
36. Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* 2013;41:e121.
37. Ye J, Zhang Y, Cui H, Liu J, Wu Y, Cheng Y, Xu H, Huang X, Li S, Zhou A, et al. WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update. *Nucleic Acids Res.* 2018;46:W71–5.
38. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics.* 2005;21:3674–6.
39. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007;23:2947–8.
40. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33:1870–4.
41. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. GSDD 2.0: an upgraded gene feature visualization server. *Bioinformatics.* 2015;31:1296–7.

42. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009;37:W202–8.
43. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc.* 2015;10:845–58.
44. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology.* 2013, 14, R36.

Tables

Table 1

Differences in fruit traits between *Vitis vinifera* / *Schisandra chinensis* and self-rooted *Vitis vinifera* Hongruibao

Plant	Length of fruit / mm	Width of fruit / mm	Weight of fruit / g
<i>Vitis vinifera</i> / <i>Schisandra chinensis</i>	23.06 ± 1.21	23.17 ± 1.21	6.86 ± 0.82
<i>Vitis vinifera</i> 'Hongruibao'	30.07 ± 1.14**	23.53 ± 0.86	10.46 ± 1.15**

Note: The double asterisks showed highly significant differences according to Student's t-test.

Table 2

Summary of Illumina transcriptome sequencing data for Vv and Vs

Type	Vv	Vs	Sum
Raw Reads	138 096 216	153 048 950	291 145 166
Valid Reads	136 975 222	151 704 222	288 679 444
Valid Reads Ratio (%)	99.19	99.11	/
Mapped Reads	103 135 533	100 706 990	203 842 532
Q20 (%)	98.98	99.12	/
GC Content (%)	48.67	50.17	/
Member of Transcript	/	/	30 609
GO Annotate	/	/	21 109
KEGG Annotate	/	/	4 151

Table 3
Summary of metabolome results

Mode	Comparison	All	Up	Down
Pos	Vv/Vs	11 757	734	454
Pos	Sc/Vv	11 757	2 232	2 099
Pos	Sc/Vs	11 757	2 299	1 911
Neg	Vv/Vs	7 868	353	260
Neg	Sc/Vv	7 868	1423	1360
Neg	Sc/Vs	7 868	1522	1255

Figures

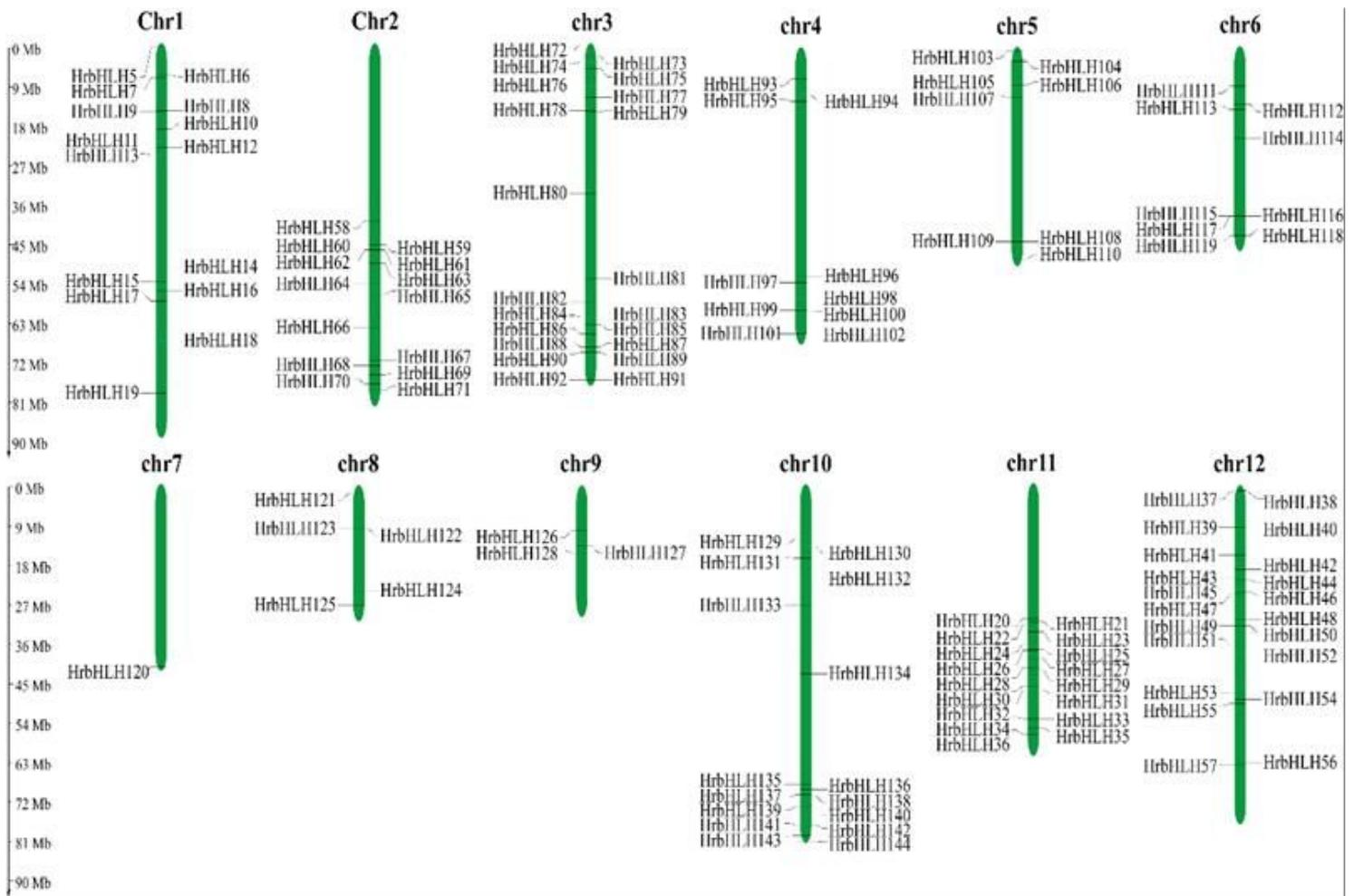


Figure 1

Distribution of 144 HrbHLH genes onto 12 sea buckthorn chromosomes

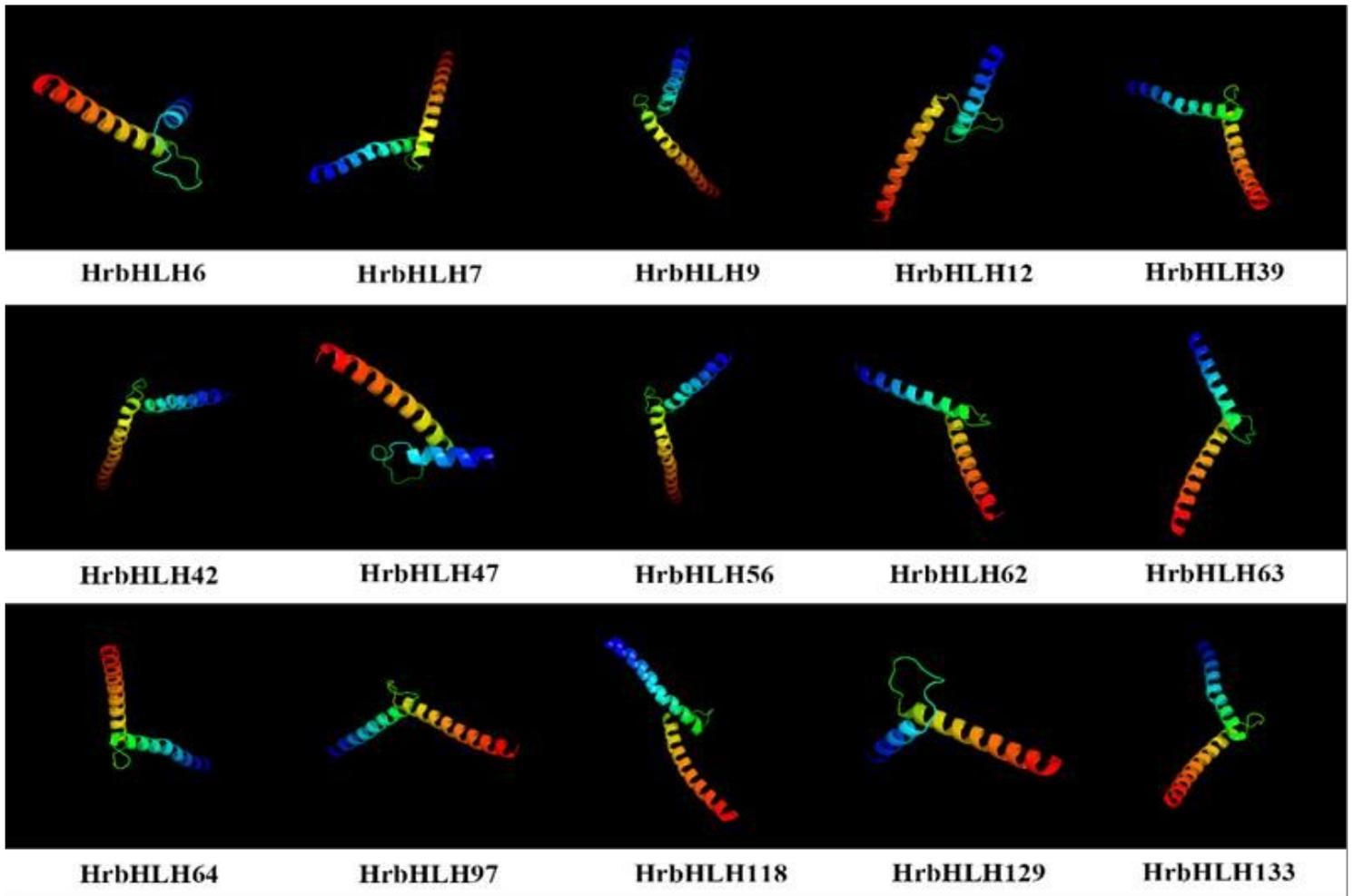


Figure 4

Predicted 3D structures of bHLH proteins in sea buckthorn.

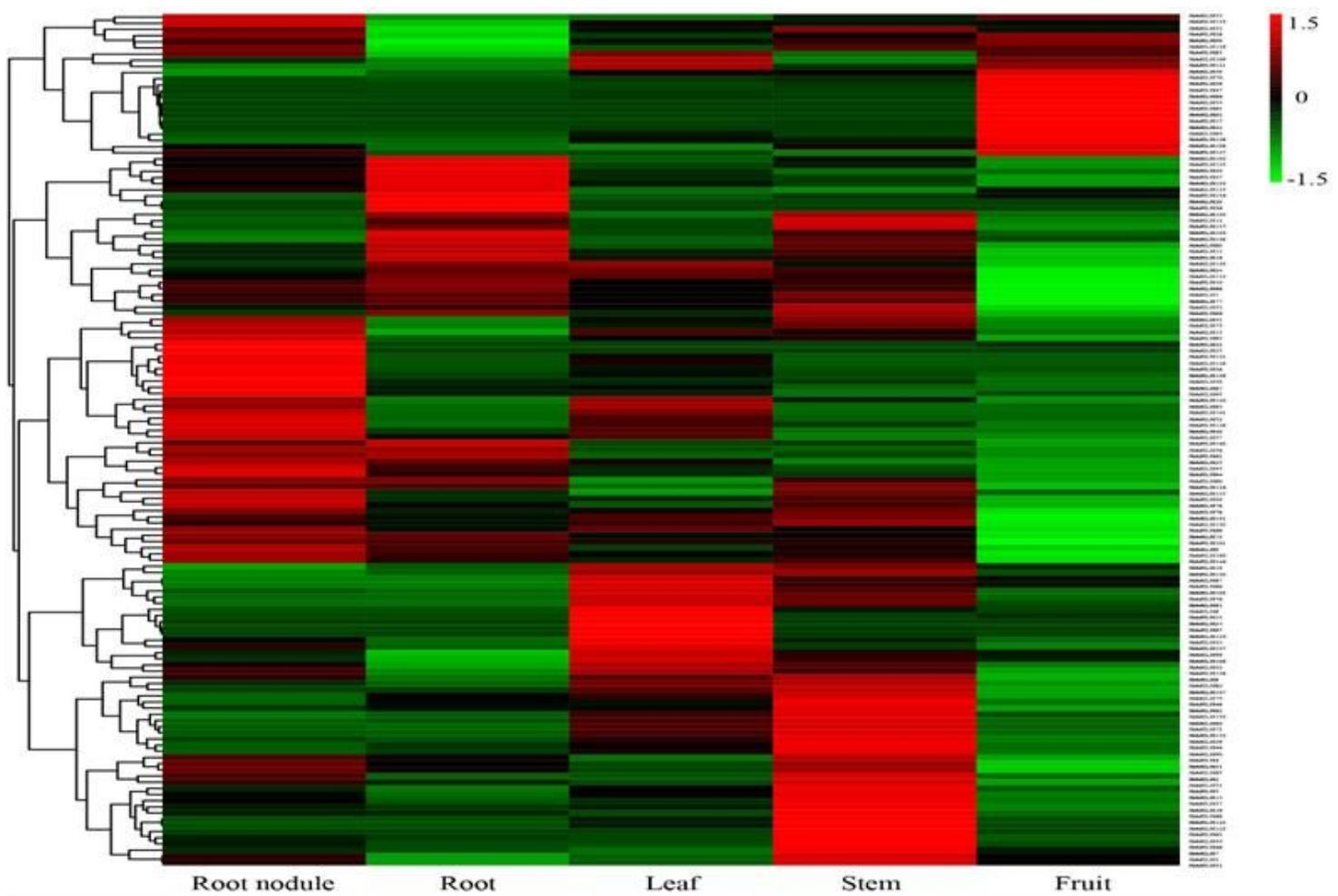


Figure 5

Expression heatmap of HrbHLH genes in different tissues of sea buckthorn.

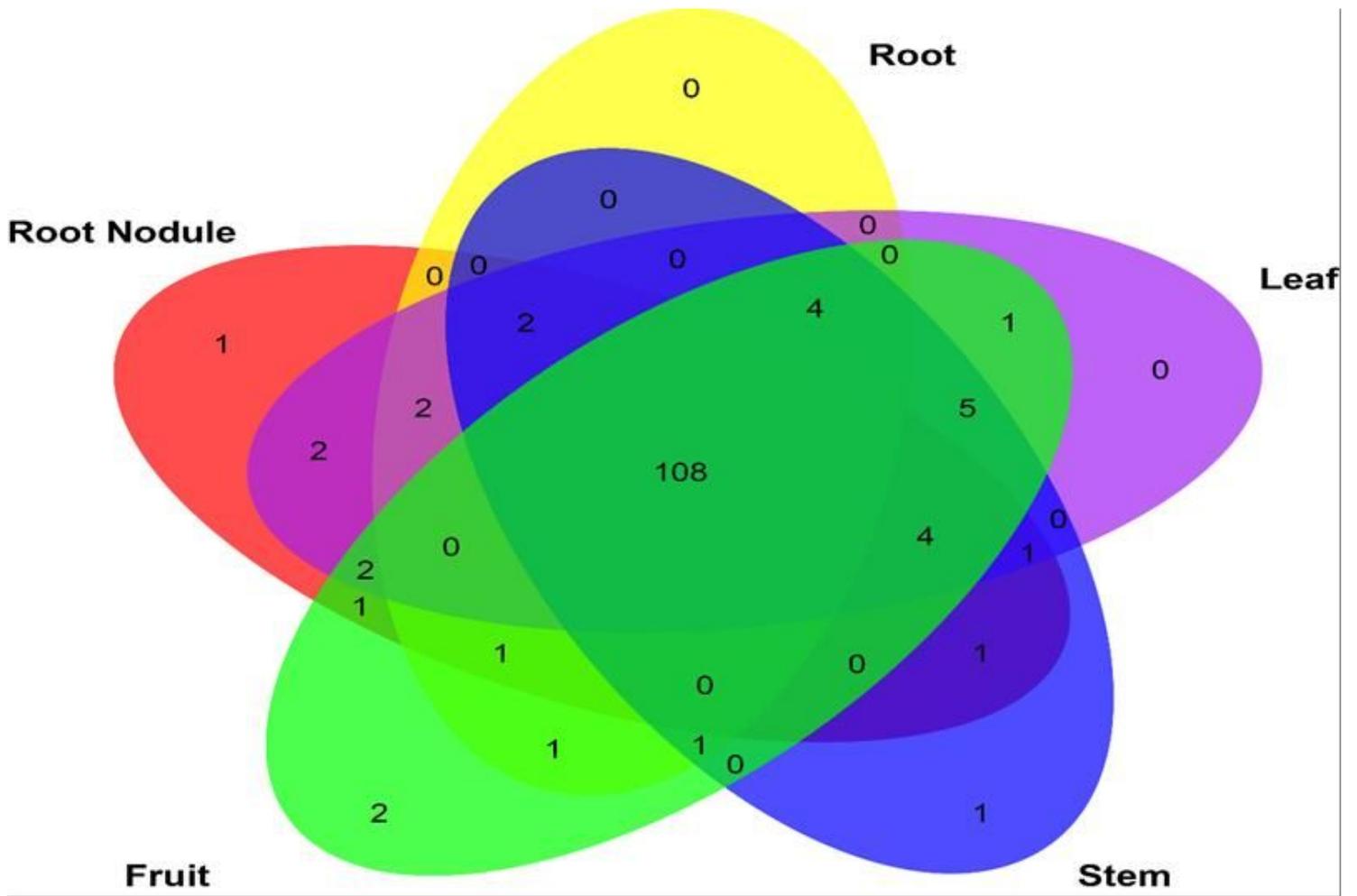


Figure 6

Venn of expressed HrbHLH gene in different tissues

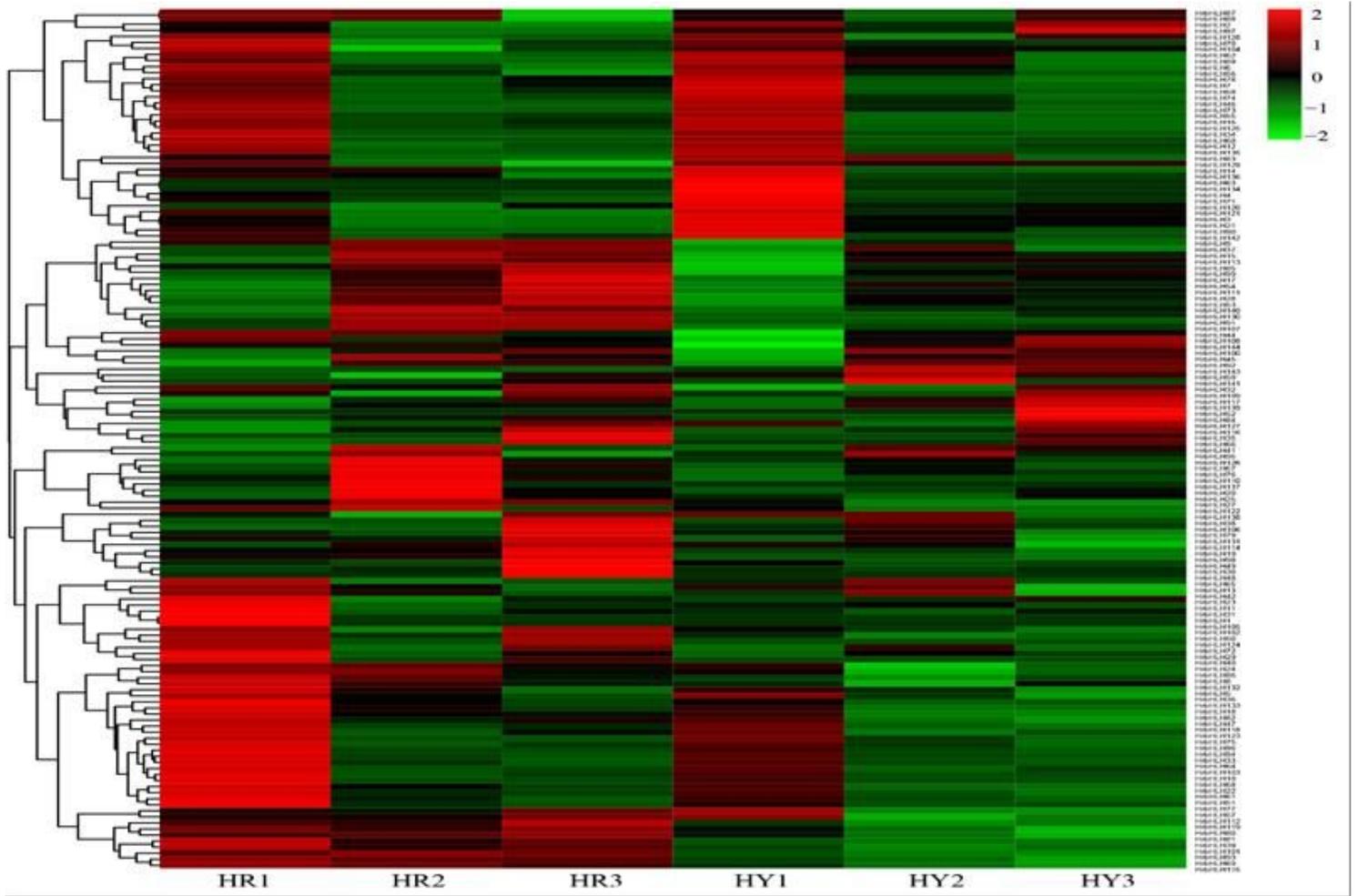


Figure 7

Expression heatmap of HrbHLH genes in different development stages of sea buckthorn fruit.

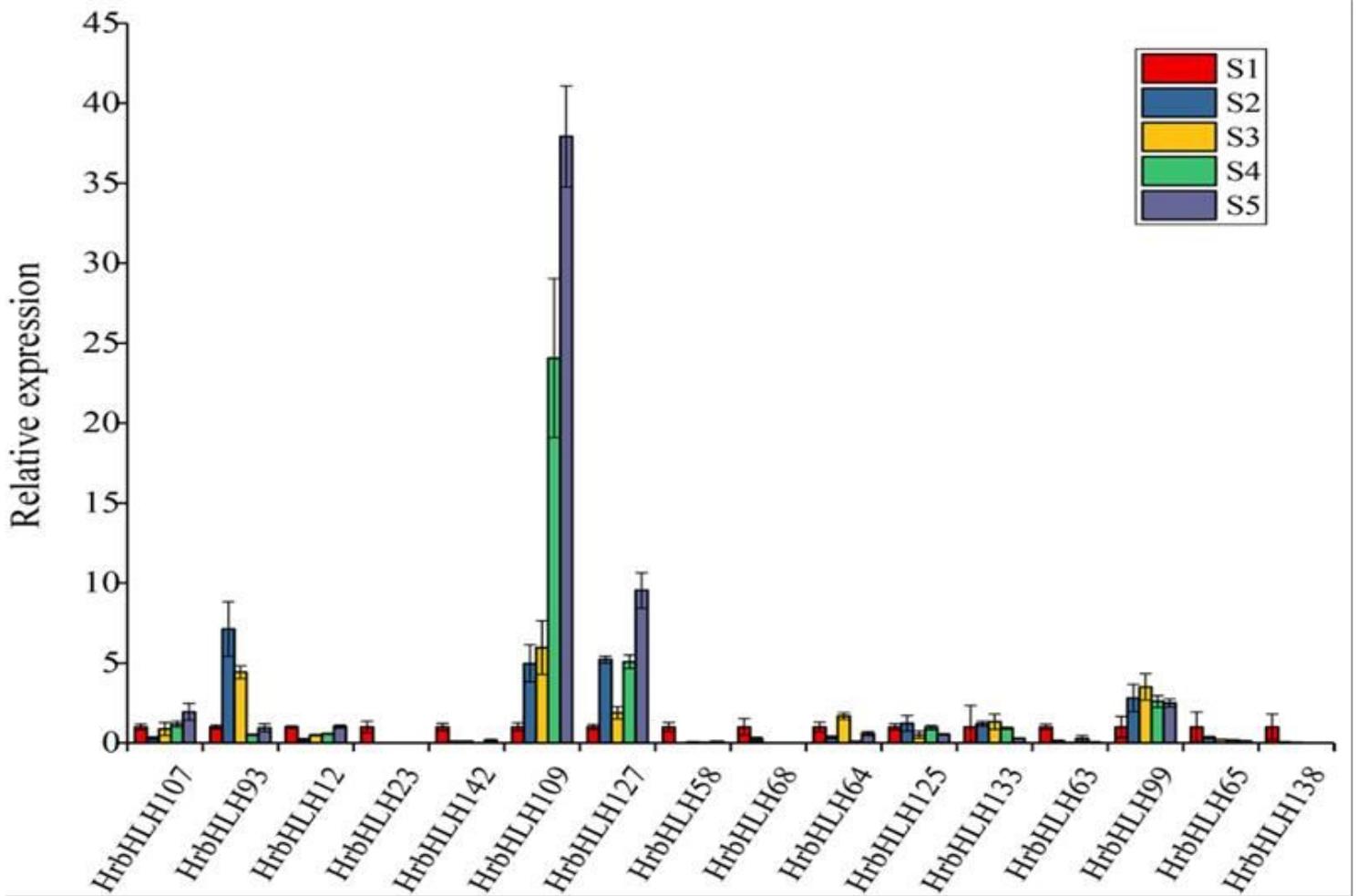


Figure 8

Expression patterns of the HrbHLH candidate genes in different development stages of sea buckthorn fruit by qRT-PCR. S1-S5: five development stages of sea buckthorn fruit.

Supplementary Files

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

- [SupplementaryTableS2.xlsx](#)
- [SupplementaryTableS1.xlsx](#)