

Genome-wide identification of SrbHLH transcription factors highlight its potential role in rebaudioside A (RA) biosynthesis in *Stevia rebaudiana*

Yuping Li

South China Botanical Garden, Chinese Academy of Sciences

Yuan Qiu

South China Botanical Garden, Chinese Academy of Sciences

Xin Xu

Southwest Forestry University

Ming Luo (✉ luoming@scbg.ac.cn)

South China Botanical Garden, Chinese Academy of Sciences

Research Article

Keywords: Terpenoids, bHLH transcription factor, Rebaudioside A, *Stevia rebaudiana*

Posted Date: May 11th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-2859524/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Plant Biology on July 6th, 2023. See the published version at <https://doi.org/10.1186/s12870-023-04353-1>.

Abstract

Stevia rebaudiana Bertonii is a valuable medicinal plant and an important source of natural sweetener, steviol glycosides (SGs), with rebaudioside A (RA) being one of the main components of SGs. bHLH genes play a crucial role in plant development and secondary metabolism. In this study, a total of 159 *SrbHLH* genes were identified from the *S. rebaudiana* genome, and each gene was named based on its chromosome location. The *SrbHLH* proteins were then clustered into 18 subfamilies through phylogenetic analysis. The classification of the *SrbHLH* family was further supported by the analysis of conserved motifs and gene structure. Chromosomal location and gene duplication events of *SrbHLH* genes were also studied. Moreover, based on the RNA-Seq data of different tissues of *S. rebaudiana*, 28 *SrbHLH*s were found to be co-expressed with structural genes involved in RA biosynthesis. The expression pattern of candidate *SrbHLH* genes was confirmed by qPCR. Finally, dual luciferase reporter assays (DLAs) and subcellular localization analysis verified *SrbHLH22*, *SrbHLH111*, *SrbHLH126*, *SrbHLH142*, and *SrbHLH152* as key regulators of RA biosynthesis. This study provides new insights into the function of *SrbHLH*s in regulating SGs biosynthesis and lays the foundation for future applications of *SrbHLH* genes in molecular breeding of *S. rebaudiana*.

Introduction

Transcription factors (TFs) play a crucial role in regulating various biological processes, including plant development, flowering, metabolism, and abiotic stress responsiveness (Lu et al., 2022, Qiao et al., 2015). Basic helix-loop-helix (bHLH) is one of the largest families of transcription factors, and these transcription factors are widely distributed in both plants and animals (Zuo et al., 2023). The conserved bHLH domain consists of a basic DNA-binding region and two amphipathic α -helices separated by a loop of variable length (HLH) (Pires and Dolan, 2010). Based on conserved domains and phylogenetic relationships, bHLH can be divided into 15 to 25 subfamilies (Buck and Atchley, 2003).

In recent years, with the development of whole genome sequencing, numerous *bHLH* gene families have been identified and analyzed in various plant species, such as *Arabidopsis* (Toledo-Ortiz et al., 2003), rice (Li et al., 2006), *Erigeron breviscapus* (Gao et al., 2022), *Aquilaria sinensis* (Sun et al., 2022), and Tartary buckwheat (Sun et al., 2020). Many *bHLH* genes have been reported to be involved in regulating terpenoid biosynthesis in plants. For instance, *Catharanthus roseus* CrMYC1 or *Artemisia annua* AaMYC2-like regulates the expression of structural genes by binding to the G-box of their promoter regions, thereby affecting the accumulation of indole alkaloids or artemisinins (Majid et al., 2019, Van Moerkercke et al., 2015). Meanwhile, *Medicago truncatula* TSAR1 and TSAR2 regulate the biosynthesis of triterpene saponins by binding to the N-box of MtHMGR1 promoter (Mertens et al., 2016). Additionally, some key transcription factors can regulate the expression of multiple structural genes in a metabolic pathway and affect the accumulation of compounds. For example, the bHLH transcription factors BI (Bitter leaf) and Bt (Bitter fruit) in cucumber (*Cucumis sativus*) can bind to the promoters of oxosqualene cyclase, fatty acyltransferase, and cytochrome P450 genes for bitter C synthesis simultaneously, regulating the accumulation of bitter taste in cucumber leaves and fruits (Shang et al., 2014).

S. rebaudiana is a medicinal herb with a sweet taste and is an important source of steviol glycosides (SGs), which are up to three hundred times sweeter than sucrose (Agulló et al., 2022) and have potential applications in the control of diseases such as diabetes, obesity, or hypertension (Olas, 2022). The biosynthesis of SGs involves backbone formation and glycosylation (Zhang et al., 2021), with kaurene synthase (*ent*-KS), kaurene oxidase (*ent*-KO), and kaurenoic acid 13-hydroxylase (*ent*-KAH) catalyzing formation of the backbone, and UDP glycosyltransferases (UGTs) being responsible for glycosylation at the C13-hydroxyl and/or C19-carboxylic acid positions of the skeleton (Wang et al., 2021, Zhang et al., 2021). Among these SGs, stevioside (ST) and rebaudioside A (RA) are the main components, with RA being more valuable due to its high sweetness and good taste (Raspe et al., 2022). To date, four UGTs, UGT74G1, UGT85C2, UGT91D2, and UGT76G1, involved in the RA biosynthesis have been identified. However, the molecular regulatory mechanisms of SG synthesis are poorly understood (Zhang et al., 2021).

In this study, we aimed to identify and characterize the bHLH gene family in the *S. rebaudiana* genome, construct a phylogenetic tree to confirm the relationships between *SrbHLH* and *AtbHLH* genes, identify segmentally duplicated *SrbHLH* genes, and analyze the contribution of duplication events to the expansion of *SrbHLH*s. We conducted co-expression analysis between *SrbHLH* genes and structural genes involved in RA synthesis, identified the relative expression levels of selected *SrbHLH* genes using qPCR, and confirmed key regulators for RA biosynthesis using transient dual luciferase reporter assays (DLA) and subcellular localization analysis. This study provides important clues for future research on the bHLH family in *S. rebaudiana*, and sheds light on the molecular regulatory mechanisms underlying SG biosynthesis.

Methods

Plant growth and MeJA treatment

S. rebaudiana seeds ("Zhongke No. 1" (S1) and "Zhongke No. 4" (S4)) used in this study were identified and planted at South China Botanical Garden, Chinese Academy of Sciences, Guangzhou City, Guangdong Province. All the study protocol for this species comply with relevant institutional, national, and international guidelines and legislation. Both variety S1 and S4 were used for tissue expression analysis, and S4 was used for MeJA treatment.

After removing the villi from the seed surface, the seeds were soaked in ddH₂O for 2 hours. Floating seeds were removed, and the remaining plump seeds were sterilized using a 1% NaClO solution for 10 minutes. Furthermore, the sterilized seeds were washed 5 times with sterilized ddH₂O and planted on MS plates and moved to a culture room with a temperature of 25 °C and a photoperiod cycle of 16 hours of light and 8 hours of dark.

For the MeJA treatment experiment, 1-week-old seedlings were transferred to MS plates with 100 μM MeJA. After 6 hours, 12 hours, and 24 hours, leaves from 30 seedlings were harvested and pooled as one

sample, then immediately frozen with liquid nitrogen and stored at -80°C. All experiments were performed with three biological replicates.

Identification of bHLH genes in the *S. rebaudiana* genome

The *S. rebaudiana* genome was sequenced and assembled in our laboratory. The Hidden Markov model (HMM) file of bHLH domain (PF00010) was downloaded from Pfam database (<http://pfam.xfam.org/>), the SrbHLH gene family members were identified by searching the *S. rebaudiana* genome using the HMM PF00010; To further verify the existence of bHLH conserved domain, the PFAM and SMART (<http://smart.emblheidelberg.de/>) programs were used. Finally, the amino acid lengths, molecular weights, and isoelectric points of candidate SrbHLHs were analyzed via the tools on the ExPASy website (<https://www.expasy.org/>), and subcellular localizations of SrbHLH proteins were predicted on the WoLF PSORT website (<https://wolfsort.hgc.jp/>).

Gene structure, genome distribution, and phylogenetic analysis

The generic feature format (GFF) file of the *S. rebaudiana* genome was used to obtain the gene structure and chromosomal location information of each SrbHLH. The conserved motif of each SrbHLH was predicted using the MEME web tool (Bailey et al., 2015), and the gene structures were displayed using TBtools (Chen et al., 2020).

To construct the phylogenetic tree, protein sequences of *Arabidopsis* were downloaded from the Phytozome database (<https://phytozome-next.jgi.doe.gov/>). The Neighbor-Joining (NJ) method was used in MEGA 7.0, and 1000 bootstrap replicates were performed using the SrbHLH and AtbHLH proteins.

RNA-Seq co-expression analysis and candidate selection

Using RNA-Seq data from different tissues of two *S. rebaudiana* varieties, we identified enzyme-coding structural genes involved in RA biosynthesis, such as *ent-KS*, *ent-KO*, *ent-KAH*, *UGT74G1*, *UGT85C2*, *UGT91D2*, and *UGT76G1*. These structural genes were used as bait to screen bHLH transcription factors that potentially regulate RA biosynthesis. Co-expression relationships between structural genes and *SrbHLHs* were estimated using Pearson correlation analysis. We filtered a set of co-expressed *SrbHLHs* genes using a Pearson coefficient threshold of 0.95 and $p < 0.05$. The OmicShare tools, a free online platform for data analysis (<https://www.omicshare.com/tools>), were applied to construct the structural-SrbHLH gene network.

RNA extraction and qPCR analysis

Total RNA extraction was conducted using a HiPure Total RNA Mini Kit (Code No. R4151-03). Vazyme HiScript Reverse Transcriptase (Cat No. R101-01) was used for cDNA synthesis using the manufacturer's instructions. qPCR was performed in 384-well plates using Vazyme HiScript Q RT SuperMix (Code No. R122-01) with a LightCycler 480 (Roche, Switzerland). The determination of the expression levels of each tested gene relative to internal reference gene (Sractin) was conducted using the $2^{-\Delta\Delta CT}$ method (Livak

and Schmittgen, 2002). All primers used in this study are listed in Table S1. Three biological replicates and three technical replicates were conducted.

Transient dual luciferase reporter assays

The *UGT76G1* promoter sequence was cloned into the pGreen II 0800-LUC reporter vector, and the full length coding sequence of *SrbHLHs* was cloned into the pCambia1300-UBQ-GFP effector vector. The constructed effector and reporter plasmids were transformed into *Agrobacterium* GV3101 and injected into the leaves of tobacco (*Nicotiana benthamiana* Domin) alongside an agrobacterium-mediated transient expression system (Hellens et al., 2005). Empty pCambia1300-UBQ-GFP vector was used as a negative control. Firefly and Renilla luciferase activity was measured using a dual luciferase reporter assay system (Promega, Madison, WI, USA). Results were evaluated as the ratio of LUC to REN activity across three independent biological replicates. Primers for vector construction are listed in supplemental Table S1.

Subcellular localization assays in tobacco leaves

The subcellular localizations of *SrbHLHs* were verified through transient expression assays in *N. benthamiana* leaves. The full-length CDSs of candidate *SrbHLHs* were cloned into pCambia1300-UBQ-GFP in-frame with a GFP reporter. The primers are listed in Table S1. The fusion constructs were co-transfected into *N. benthamiana* leaves alongside a nuclear localization marker (Histone3-mCherry). After incubation for 3 days, the localization of the target proteins was observed using a confocal fluorescence microscope (Leica, SP8).

Results

Identification of the *SrbHLH* gene family in *S. rebaudiana*

After using NCBI Batch CD-Search tools to eliminate redundant proteins by, we obtained 159 bHLHs from the *S. rebaudiana* genome. To distinguish these genes, we named them as *SrbHLH1* to *SrbHLH159*, based on their location on the chromosomes of *S. rebaudiana* (Table 1). Next, we examined various physicochemical properties of the *SrbHLHs*, including amino acid lengths, isoelectric point (PI), protein molecular weight (MW), and subcellular localization. Among the 159 *SrbHLH* proteins, their amino acid sequences varied from 93 aa (*SrbHLH44*) to 661 aa (*SrbHLH35*). The MWs of *SrbHLH* proteins ranged from 10.33 kDa to 70.78 kDa, while the PIs of the *SrbHLH* proteins ranged from 4.38 (*SrbHLH128*) to 10.15 (*SrbHLH159*). Additionally, based on subcellular localization predictions, most of the *SrbHLH* proteins were expected to be located in the nucleus, cytoplasm, or chloroplast, with only *SrbHLH34* likely to be found in the cytoskeleton (Table 1).

Phylogenetic analysis and classification of *SrbHLH* genes

To evaluate the classification and evolutionary relationships among *SrbHLHs* in plants, we utilized bHLH amino acid sequences from *S. rebaudiana* and *Arabidopsis* to construct a Neighbor-Joining (NJ) tree

using MEGA 7.0 software. The phylogenetic analysis results revealed that 159 *SrbHLH*s were categorized into 18 subfamilies (Fig. 1), consistent with previous research (Gao et al., 2022, Li et al., 2006). Subfamily C contained 40 *SrbHLH*s, while subfamily P and subfamily V contained 18 and 14 *SrbHLH*s, respectively. Furthermore, subfamily E, subfamily S, and subfamily R included 13 *SrbHLH*s, whereas subfamily I, subfamily G, subfamily U, and subfamily T, each had only one assigned *SrbHLH*. Notably, subfamilies H, W, and F did not have any *SrbHLH*s assigned to them.

Gene structure and motif analysis of *SrbHLH* genes in *S. rebaudiana*

To investigate the characteristic regions of *SrbHLH* proteins, we utilized MEME to identify conserved motifs among the 159 *SrbHLH* proteins (Fig. 2). Ten distinct motifs were identified and designated as motif 1 through 10. Figure 2A illustrates that *SrbHLH* proteins contain different numbers of conserved motifs ranging from 1 to 5. Almost all *SrbHLH* family members contain two highly conserved motifs (motif 1 and 2). Additionally, we observed that *SrbHLH* proteins belonging to the same subfamily typically share similar composition and relative position. For instance, *SrbHLH* proteins within subfamily P contain motif 1, motif 2, motif 4, motif 7, and motif 8, whereas the members of subfamily V share common motifs 1, 2, and 5 (Fig. 2A). The fact that members of the same subfamily have similar gene structure and motif composition and tend to gather together in the phylogenetic trees, supports the accuracy and authenticity of the subfamily classification.

We also analyzed the UTR, exons, and introns structure of identified *SrbHLH* genes to further explore their structural composition (Fig. 2B). By comparing the number and position of exons and introns, we observed that the 159 *SrbHLH* genes possess varying numbers, ranging from 1 to 11, with only one *SrbHLH* gene being intronless. Some subfamilies contained genes with a specific number of exons, such as subfamily A where 3 *SrbHLH* genes in this clade contained 2 exons. Generally, genes clustered into the same group in the phylogenetic tree have a similar number of exons, such as *SrbHLH*47, *SrbHLH*48, *SrbHLH*49, and *SrbHLH*125 which each contain four exons, with similar relative positions to the UTR.

Chromosome distribution, gene duplication, and syntenic analysis of *SrbHLH* genes in *S. rebaudiana* genome

TBtools was utilized to investigate the regions of interest for *SrbHLH* genes on the chromosome. The analysis involved anchoring 159 *SrbHLH* genes to their corresponding chromosomes. The results demonstrated that the distribution of *SrbHLH* genes on the *S. rebaudiana* genome was irregular, and the genes were randomly located on all 11 chromosomes (Fig. 3). Chromosome 1 had the most *SrbHLH*s, with 40 genes, followed by chromosome 2, which had 33 genes. Conversely, chromosome 9 had the least *SrbHLH*s, with only 2 genes.

Genome duplication events are considered primary drivers of genome evolution and gene family expansion, with tandem and segmental duplication being the primary patterns. In this study, we observed no tandem repeat events in the *SrbHLH* family, but we did identify 263 pairs of segmental duplicates that were unevenly distributed on different chromosomes in the *S. rebaudiana* genome (Fig. 4). Among these

pairs of collinear relationships, some *SrbHLH*s were paired with more than one gene. For instance, *SrbHLH149* had a collinear relationship with *SrbHLH11*, *SrbHLH97*, *SrbHLH115* and *SrbHLH122*, while *SrbHLH152* has collinear relationship with *SrbHLH91*, *SrbHLH126*, *SrbHLH141*, and *SrbHLH142*, respectively (Fig. 4, Supplementary table S2). These findings suggest that gene duplication events may be the primary cause of the expansion of the *SrbHLH* family.

Correlation analysis between *SrbHLH*s and structural genes involved in RA biosynthesis

Gene expression patterns provides valuable insights into gene function. As the content of stevioside varies in different tissues of *S. rebaudiana* and the leaves are the main site of SG accumulation (Olas, 2022), we analyzed the expression of *SrbHLH* genes in different tissues. A heatmap was generated using Fragments Per Kilobase per Million (FPKM) values from RNA-seq data from of two *S. rebaudiana* varieties (with higher RA content in S1 than S4) to show their tissue-specific expression patterns (Fig. 5A). Most *SrbHLH* genes exhibited tissue-specific expression patterns, such as *SrbHLH25*, *SrbHLH47*, *SrbHLH48*, and *SrbHLH49*, which were highly expressed in flowers, *SrbHLH111*, *SrbHLH154*, and *SrbHLH155*, which were highly expressed in leaves, *SrbHLH57*, *SrbHLH72*, and *SrbHLH76*, which were highly expressed in roots, and *SrbHLH69* and *SrbHLH70*, which were highly expressed in stems. Furthermore, some *SrbHLH* genes showed differential expression between the two varieties. For example, the expression level of *SrbHLH15*, *SrbHLH16*, *SrbHLH18*, and *SrbHLH19* in the roots of S1 was higher than that of S4, the expression level of *SrbHLH57*, *SrbHLH58*, and *SrbHLH88* in the flowers of S4 was higher than that of S1, and the expression level of *SrbHLH124* in the leaves of S1 was higher than that of S4 (Fig. 5A). These findings provide a basis for identifying bHLH transcription factors that regulate RA biosynthesis.

The biosynthesis pathway of RA in *S. rebaudiana* has been elucidated, with specificity mainly occurring in the later cyclization and modification pathways. Seven structural genes, including *ent-KS*, *ent-KO*, *ent-KAH*, *UGT74G1*, *UGT85C2*, *UGT91D2*, and *UGT76G1*, are involved in the formation of RA in *S. rebaudiana* (Xu et al., 2021, Zhang et al., 2021). Using RNA-Seq data from different tissues of *S. rebaudiana*, co-expression analysis was conducted to identify key *SrbHLH* genes regulating RA biosynthesis. Notably, we found that the expression levels of 11 *SrbHLH* genes were significantly negatively correlated with the seven structural genes, while the expression levels of 17 *SrbHLH* genes were significantly positively correlated with the seven structural genes (Fig. 5B). Therefore, we hypothesized that these 28 *SrbHLH* genes may be involved in the regulation of RA biosynthesis.

SrbHLH gene expression patterns in various tissues.

To validate the RNA-Seq results, we conducted qPCR assays to assess the gene expression of *SrbHLH*s in different tissues of *S. rebaudiana* varieties S1 and S4. We focused on 12 *SrbHLH* genes with the highest correlation with the critical RA biosynthesis genes *UGT74G1* and *UGT76G1*. As depicted in Fig. 6, the qPCR results were consistent with the RNA-Seq data. Notably, *SrbHLH22*, *SrbHLH111*, and *SrbHLH152*

exhibited higher expression in leaves than other tissues, whereas *SrbHLH126*, *SrbHLH142*, and *SrbHLH148* showed higher expression levels in all tissues. Additionally, only *SrbHLH134* demonstrated significantly higher expression in the roots of variety S4 than that of other tissues. Overall, these results suggest that *SrbHLH22*, *SrbHLH111*, and *SrbHLH152* may have crucial roles in RA biosynthesis and accumulation.

Expression of *SrbHLH* genes in response to MeJA treatment

MeJA has been shown to regulate the biosynthesis of multiple secondary metabolites (Chen et al., 2022), such as flavonoid and soyasaponin biosynthesis in *G. uralensis* (Li et al., 2020, Tamura et al., 2018). Therefore, we examined the expression patterns of 12 *SrbHLH* genes selected in our study under MeJA treatment. As depicted in Fig. 7, all *SrbHLH* genes except for *SrbHLH38* and *SrbHLH54*, were responsive to MeJA treatment, with four genes down-regulated and ten up-regulated. These results indicate that most of the *SrbHLH* genes are MeJA-responsive and may play a role in JA-regulated secondary metabolic processes.

Functional verification of key *SrbHLH* genes regulating rebaudioside A accumulation

It is well established that *UGT74G1* and *UGT76G1* are two major genes involved in RA biosynthesis. *UGT74G1* catalyzes the conversion of steviolmonoside to rubusoside, while stevioside is converted to RA under catalysis by *UGT76G1* (Petit et al., 2019, Zhang et al., 2021). In this study, we cloned and analyzed the transcriptional activation or repression of 14 selected *SrbHLH* genes on *UGT76G1* using a dual luciferase assay (DLA). Results showed that 5 *SrbHLH*s (*SrbHLH22*, *SrbHLH111*, *SrbHLH126*, *SrbHLH142*, *SrbHLH152*) transcriptionally activated the expression of *UGT76G1*, while *SrbHLH6*, 12, 60, 72, 83, 85, 134, 147, and 148 were implicated in regulating the expression of other structural genes involved in RA biosynthesis (Fig. 8).

Subcellular localization of key bHLHs involved in regulating rebaudioside A accumulation

Subcellular localization is a crucial factor in understanding gene function. Our prediction analysis revealed that *SrbHLH*s exhibit a variety of subcellular localization patterns (Table 1). To further investigate this, we experimentally determined the subcellular localization of 5 *SrbHLH*s that regulate RA biosynthesis, as confirmed by DLA. Full-length cDNAs of these *SrbHLH*s were fused to Green Fluorescent Protein (GFP) under the ubiquitin promoter and transiently expressed in *N. benthamiana* leaves. Our findings showed that *SrbHLH111* and *SrbHLH142* localized in both the nucleus and cytosol, while *SrbHLH22*, *SrbHLH126*, and *SrbHLH152* were mainly located in the nucleus (Fig. 9). These nuclear localization results were consistent with the bioinformatics prediction and the expected role of *SrbHLH*s as transcription factors.

Discussion

The bHLH transcription factors (TFs) play crucial roles in plants, responses to stress, and metabolic processes. Current genome-wide characterization of *bHLH* genes has been reported in numerous plants. However, little is known about this gene family in *S. rebaudiana*, prompting us to systematically examine bHLH proteins in *S. rebaudiana*, especially looking into their function in regulating RA biosynthesis. Here, we identified 159 *SrbHLH* genes by investigating the *S. rebaudiana* genome, and based on classification methods from a previous study (Toledo-Ortiz et al., 2003), these 159 *SrbHLH* proteins were clustered into 18 groups. Generally, genes in the same branch had similar functions. There were many *SrbHLH* members that were proximal to the bHLH proteins in *Arabidopsis*, which were predicted to be homologous genes. For example, *AtbHLH29* regulates iron acquisition processes in *Arabidopsis*, and *SrbHLH114* is homologous gene of *AtbHLH29*. Thus, we predicted that *SrbHLH114* proteins in *S. rebaudiana* also participate in the regulation of iron acquisition. These findings will contribute to the prediction of functions of *SrbHLH* genes in *S. rebaudiana*.

The gene expression level is critically important for regulating plant biological progress. In general, genes with the same expression patterns are more likely to be involved in production of similar bioactive compounds (Lau and Sattely, 2015). In this study, based on RNA-Seq data, we discovered that 159 *SrbHLH* genes presented different expression patterns, and numerous tissue-specific *SrbHLH* genes were identified (Fig. 5A). These findings will contribute to the screening of candidate genes regulating RA synthesis. Interestingly, co-expression analysis using RNA-Seq data from different tissue samples demonstrated that 11 *SrbHLH* genes were significantly negatively correlated with the seven structural genes involved in RA biosynthesis, whereas 17 *SrbHLH* genes were significantly positively correlated with structural genes (Fig. 5B). Therefore, in the future, these co-expressed *SrbHLHs* with structural genes may be candidate transcription factors regulating RA synthesis.

Several studies have shown that MeJA treatment induces the accumulation of secondary metabolites in various plant species. For example, terpenoids were induced in *Catharanthus roseus* cells (Zhang et al., 2018) and soyasaponins in *G. uralensis* (Li et al., 2020). MeJA treatment also affected the expression of key genes in the biosynthesis of SGs in *S. rebaudiana* (Tavakoli et al., 2019). In this study, we identified 14 *SrbHLH* genes that were co-expressed with structural genes and responded to MeJA treatment. Among them, 4 *SrbHLHs* were down-regulated and 10 were up-regulated (Fig. 7), suggesting their important roles in JA-mediated biological processes, including RA biosynthesis. It is worth noting that bHLH proteins have been reported to regulate metabolic process by interacting with other transcription factors or co-factors (Pérez-Llorca et al., 2023). For example, the COI1/JAZs/MYC2 module has been identified as a core regulator of JA-mediated terpenoid accumulation (Patra et al., 2018). Therefore, investigating the regulatory function of *SrbHLHs* and their interaction with co-factors in RA biosynthesis in future studies could be of great interest.

Transcription factors typically bind to specific *cis*-elements in the promoter region of target genes to regulate the synthesis of compounds catalyzed by the target genes by regulating their expression. In the case of the bHLH gene family, binding sites usually include G-box, E-box, or N-box. For example, CrMYC1 or AaMYC2-like proteins have been shown to regulate the expression of structural genes by binding to the

G-box of their promoter regions, affecting the biosynthesis of indole alkaloids or artemisinin (Majid et al., 2019, Van Moerkercke et al., 2015). In the current study, we used DLA to examine the binding of candidate *SrbHLH*s to the promoter regions of structural genes involved in RA synthesis. The results showed that 5 *SrbHLH*s (*SrbHLH22*, *SrbHLH111*, *SrbHLH126*, *SrbHLH142*, *SrbHLH152*) transcriptionally activated the expression of *UGT76G1* (Fig. 8). Furthermore, the nuclear localization of these *SrbHLH* proteins results was consistent with their expected role as transcription factors (Fig. 9). Therefore, these 5 *SrbHLH*s appear to be the most promising candidates for regulating RA synthesis. In the future, transgenic plants will be constructed to evaluate the function of these 5 *SrbHLH*s in the RA biosynthesis pathway.

Conclusion

In this study, we identified 159 *bHLH* (*SrbHLH*) genes in the *S. rebaudiana* genome. We then constructed a phylogenetic tree to confirm the relationships between *SrbHLH* and *AtbHLH* proteins, and found that *SrbHLH*s in the same group had similar protein motifs and gene structures. In addition, 263 pairs of segmental duplicated *SrbHLH* genes were identified, indicating duplication events which contributed to the expansion of *SrbHLH*s. Based on RNA-Seq data, we found that expression patterns of *SrbHLH* differed across various tissues. Moreover, in order to identify candidate *SrbHLH*s which regulate RA biosynthesis, co-expression analysis between *SrbHLH* genes and structural genes involved in RA synthesis was conducted. The relative expression levels of the selected *SrbHLH* genes was identified by qPCR. Finally, *SrbHLH22*, *SrbHLH111*, *SrbHLH126*, *SrbHLH142*, and *SrbHLH152* were confirmed as regulators of RA biosynthesis through transient dual luciferase reporter assays (DLA) and subcellular localization analysis. These results may eventually contribute to further understanding of the functions of *SrbHLH*s in SGs biosynthesis regulation, and provide theoretical basis for the application of *SrbHLH* genes in molecular breeding of *S. rebaudiana*.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and/or analysed during this study are available in NGDC (<https://ngdc.cncb.ac.cn/>) with the accession no. PRJCA016649. Other data are available in the supplementary table.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by Key Laboratory of Molecular Analysis and Genetic Improvement of South China Agricultural Plants Foundation of South China Botanical Garden, Chinese Academy of Sciences (KF200204), the National Natural Science Foundation of China (32170362), the Guangdong Natural Science Funds for Distinguished Young Scholars (2022B1515020026), the Youth Innovation Promotion Association, Chinese Academy of Sciences (Y2021094), the South China Botanical Garden, the Chinese Academy of Sciences (QNXM-02), the Natural Science Foundation of Guangdong Province (2021A515110122), Basic and Applied Research Foundation of Guangzhou City (202201010756) and Innovation Training Programs for Undergraduates, CAS (KCJH-80107-2020-041).

Authors' contributions

YPL and ML designed the experiments. YPL, YQ, and XX performed and analyzed the data. YPL and ML wrote and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank editors and reviewers for the careful reading and valuable comments. We apologize to researchers whose studies are not cited due to space limitations.

References

1. Agulló V., García-Viguera C., Domínguez-Perles R., The use of alternative sweeteners (sucralose and stevia) in healthy soft-drink beverages, enhances the bioavailability of polyphenols relative to the classical caloric sucrose, *Food chemistry*, 2022, 370: 131051.
2. Buck M.J., Atchley W.R., Phylogenetic analysis of plant basic helix-loop-helix proteins, *Journal of molecular evolution*, 2003, 56: 742-750.
3. Chen C., Liu F., Zhang K., Niu X., Zhao H., Liu Q., Georgiev M.I., Xu X., Zhang X., Zhou M., MeJA-responsive bHLH transcription factor *LjbHLH7* regulates cyanogenic glucoside biosynthesis in *Lotus japonicus*, *Journal of experimental botany*, 2022, 73: 2650-2665.
4. Gao Q., Song W., Li X., Xiang C., Chen G., Xiang G., Liu X., Zhang G., Li X., Yang S., Zhai C., Zhao Y., Genome-wide identification of bHLH transcription factors: Discovery of a candidate regulator related to flavonoid biosynthesis in *Erigeron breviscapus*, *Frontiers in plant science*, 2022, 13: 977649.

5. Hellens R.P., Allan A.C., Friel E.N., Bolitho K., Grafton K., Templeton M.D., Karunairetnam S., Gleave A.P., Laing W.A., Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants, *Plant methods*, 2005, 1: 13.
6. Lau W., Sattely E.S., Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone, *Science*, 2015, 349: 1224-1228.
7. Li X., Duan X., Jiang H., Sun Y., Tang Y., Yuan Z., Guo J., Liang W., Chen L., Yin J., Ma H., Wang J., Zhang D., Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and Arabidopsis, *Plant physiology*, 2006, 141: 1167-1184.
8. Li Y., Liang X., Zhou X., Wu Z., Yuan L., Wang Y., Li Y., Selection of reference genes for qRT-PCR analysis in medicinal plant Glycyrrhiza under abiotic stresses and hormonal treatments, *Plants (Basel)*, 2020, 9: 1441.
9. Livak K.J., Schmittgen T.D., Analysis of relative gene expression data using real-time quantitative PCR, *Methods*, 2002, 25: 402-408.
10. Lu X., Zhang H., Hu J., Nie G., Khan I., Feng G., Zhang X., Wang X., Huang L., Genome-wide identification and characterization of bHLH family genes from orchardgrass and the functional characterization of DgbHLH46 and DgbHLH128 in drought and salt tolerance, *Functional & integrative genomics*, 2022, 22: 1331-1344.
11. Majid I., Kumar A., Abbas N., A basic helix loop helix transcription factor, AaMYC2-Like positively regulates artemisinin biosynthesis in *Artemisia annua* L, *Industrial Crops and Products*, 2019, 128: 115-125.
12. Mertens J., Pollier J., Vanden Bossche R., Lopez-Vidriero I., Franco-Zorrilla J.M., Goossens A., The bHLH transcription factors TSAR1 and TSAR2 regulate triterpene saponin biosynthesis in *Medicago truncatula*, *Plant physiology*, 2016, 170: 194-210.
13. Olas B., *Stevia rebaudiana* Bertoni and its secondary metabolites: Their effects on cardiovascular risk factors, *Nutrition*, 2022, 99-100: 111655.
14. Patra B., Pattanaik S., Schluttenhofer C., Yuan L., A network of jasmonate-responsive bHLH factors modulate monoterpenoid indole alkaloid biosynthesis in *Catharanthus roseus*, *The New phytologist*, 2018, 217: 1566-1581.
15. Pérez-Llorca M., Pollmann S., Müller M., Ethylene and jasmonates signaling network mediating secondary metabolites under abiotic stress, *International journal of molecular sciences*, 2023, 24 (6): 5990.
16. Petit E., Jacques A., Daydé J., Vallejo V., Berger M., UGT76G1 polymorphism in *Stevia rebaudiana*: New variants for steviol glycosides conjugation, *Plant physiology and biochemistry*, 2019, 135: 563-569.
17. Pires N., Dolan L., Origin and diversification of basic-helix-loop-helix proteins in plants, *Molecular biology and evolution*, 2010, 27: 862-874.
18. Qiao X., Li M., Li L., Yin H., Wu J., Zhang S., Genome-wide identification and comparative analysis of the heat shock transcription factor family in Chinese white pear (*Pyrus bretschneideri*) and five other

Rosaceae species, *BMC plant biology*, 2015, 15: 12.

19. Raspe D.T., da Silva C., Cláudio da Costa S., Compounds from *Stevia rebaudiana* Bertoni leaves: An overview of non-conventional extraction methods and challenges, *Food Bioscience*, 2022, 46: 101593.
20. Shang Y., Ma Y., Zhou Y., Zhang H., Duan L., Chen H., Zeng J., Zhou Q., Wang S., Gu W., Liu M., Ren J., Gu X., Zhang S., Wang Y., Yasukawa K., Bouwmeester H.J., Qi X., Zhang Z., Lucas W.J., Huang S., Plant science. Biosynthesis, regulation, and domestication of bitterness in cucumber, *Science*, 2014, 346 (6213): 1084-1088.
21. Sun P.W., Gao Z.H., Lv F.F., Yu C.C., Jin Y., Xu Y.H., Wei J.H., Genome-wide analysis of basic helix-loop-helix (bHLH) transcription factors in *Aquilaria sinensis*, *Scientific reports*, 2022, 12: 7194.
22. Sun W., Jin X., Ma Z., Chen H., Liu M., Basic helix-loop-helix (bHLH) gene family in Tartary buckwheat (*Fagopyrum tataricum*): Genome-wide identification, phylogeny, evolutionary expansion and expression analyses, *International journal of biological macromolecules*, 2020, 155: 1478-1490.
23. Tamura K., Yoshida K., Hiraoka Y., Sakaguchi D., Chikugo A., Mochida K., Kojoma M., Mitsuda N., Saito K., Muranaka T., Seki H., The basic helix-loop-helix transcription factor GubHLH3 positively regulates soyasaponin biosynthetic genes in *Glycyrrhiza uralensis*, *Plant Cell Physiol.*, 2018, 59: 778-791.
24. Tavakoli H., Tavakoli N., Moradi F., The effect of the elicitors on the steviol glycosides biosynthesis pathway in *Stevia rebaudiana*, *Functional plant biology : FPB*, 2019, 46: 787-795.
25. Toledo-Ortiz G., Huq E., Quail P.H., The Arabidopsis basic/helix-loop-helix transcription factor family, *The Plant cell*, 2003, 15: 1749-1770.
26. Van Moerkercke A., Steensma P., Schweizer F., Pollier J., Gariboldi I., Payne R., Vanden Bossche R., Miettinen K., Espoz J., Purnama P.C., Kellner F., Seppänen-Laakso T., O'Connor S.E., Rischer H., Memelink J., Goossens A., The bHLH transcription factor BIS1 controls the iridoid branch of the monoterpenoid indole alkaloid pathway in *Catharanthus roseus*, *Proceedings of the National Academy of Sciences of the United States of America*, 2015, 112: 8130-8135.
27. Wang Y., Sun X., Jia X., Zhu L., Yin H., Comparative transcriptomic of *Stevia rebaudiana* provides insight into rebaudioside D and rebaudioside M biosynthesis, *Plant physiology and biochemistry*, 2021, 167: 541-549.
28. Xu X., Yuan H., Yu X., Huang S., Sun Y., Zhang T., Liu Q., Tong H., Zhang Y., Wang Y., Liu C., Wu L., Hou M., Yang Y., The chromosome-level *Stevia* genome provides insights into steviol glycoside biosynthesis, *Horticulture research*, 2021, 8: 129.
29. Zhang S., Yang Y., Lyu C., Chen J., Li D., Liu Y., Zhang Z., Liu Y., Wu W., Identification of the key residues of the uridine diphosphate glycosyltransferase 91D2 and its effect on the accumulation of steviol glycosides in *Stevia rebaudiana*, *Journal of agricultural and food chemistry*, 2021, 69: 1852-1863.
30. Zhang X.N., Liu J., Liu Y., Wang Y., Abozeid A., Yu Z.G., Tang Z.H., Metabolomics analysis reveals that ethylene and methyl jasmonate regulate different branch pathways to promote the accumulation of

- terpenoid indole alkaloids in *Catharanthus roseus*, *Journal of natural products*, 2018, 81: 335-342.
31. Zuo Z.F., Lee H.Y., Kang H.G., Basic Helix-Loop-Helix Transcription factors: Regulators for plant growth development and abiotic stress responses, *International journal of molecular sciences*, 2023, 24 (2): 1419.

Table

Table 1 is available in the Supplementary Files section.

Figures

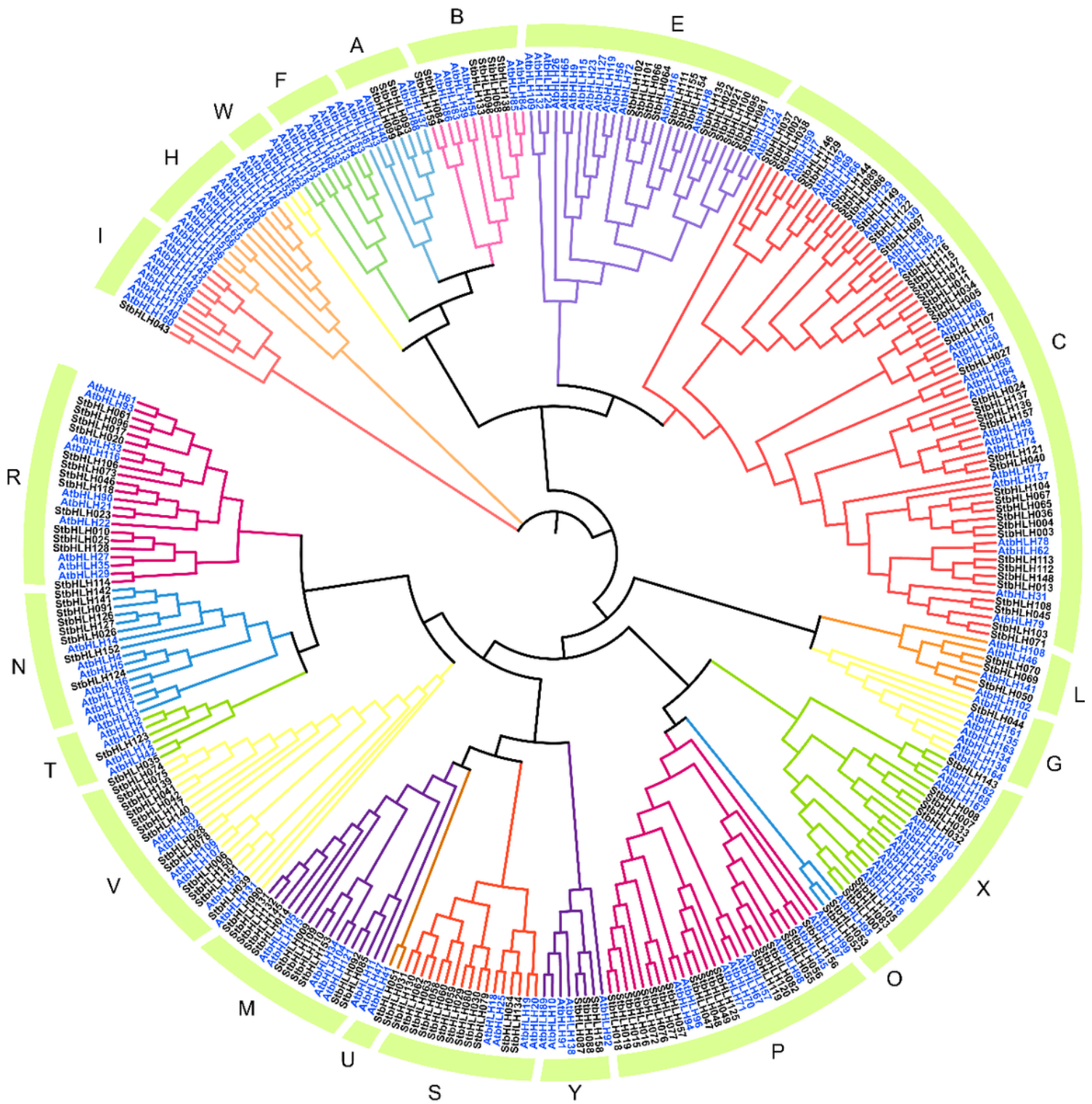


Figure 1

Phylogenetic tree of bHLH proteins in *S. rebaudiana* and *Arabidopsis thaliana*.

The phylogenetic tree was obtained using the NJ method in MEGA7. The tree shows 21 phylogenetic subfamilies, each subfamily represented by different colors. bHLH protein from *Arabidopsis* is labeled in blue.



Figure 2

The gene structure and motif distribution analysis of the Sr bHLH proteins in *S. rebaudiana*.

A. Phylogenetic trees of Sr bHLH proteins constructed by the NJ method; B. Ten conserved motif in the StbHLH protein; C. Exons and introns distribution of StbHLH genes, and green rectangles and gray lines show the exons and introns, respectively.

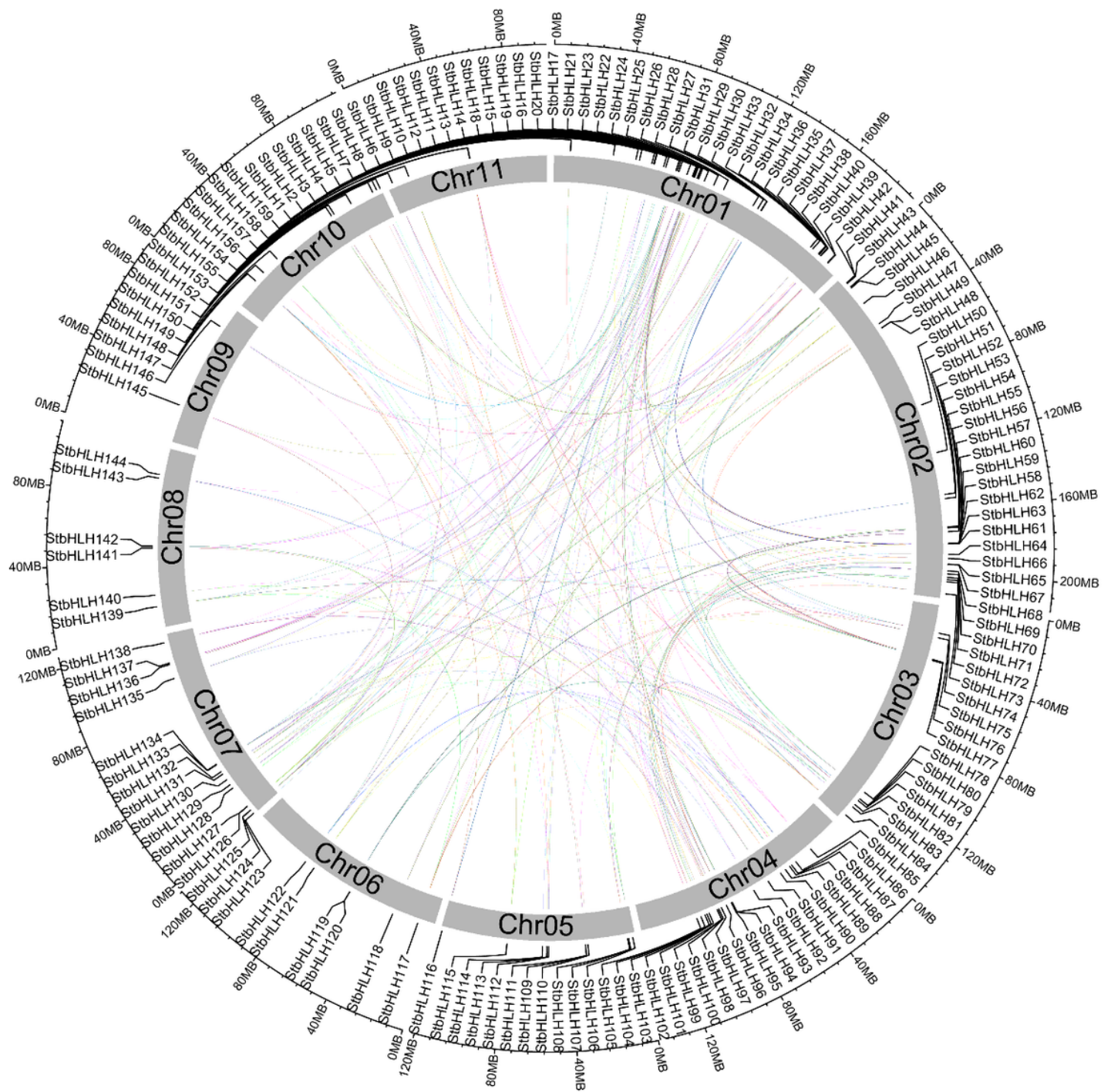


Figure 4

Collinearity analysis of the SrhHLHs in *S. rebaudiana* genome.

The duplicated SrhHLHs were mapped to different chromosomes using shinyCircos. The collinearity relationships among SrhHLHs are linked by colorful lines and the chromosome is presented by grey boxes.

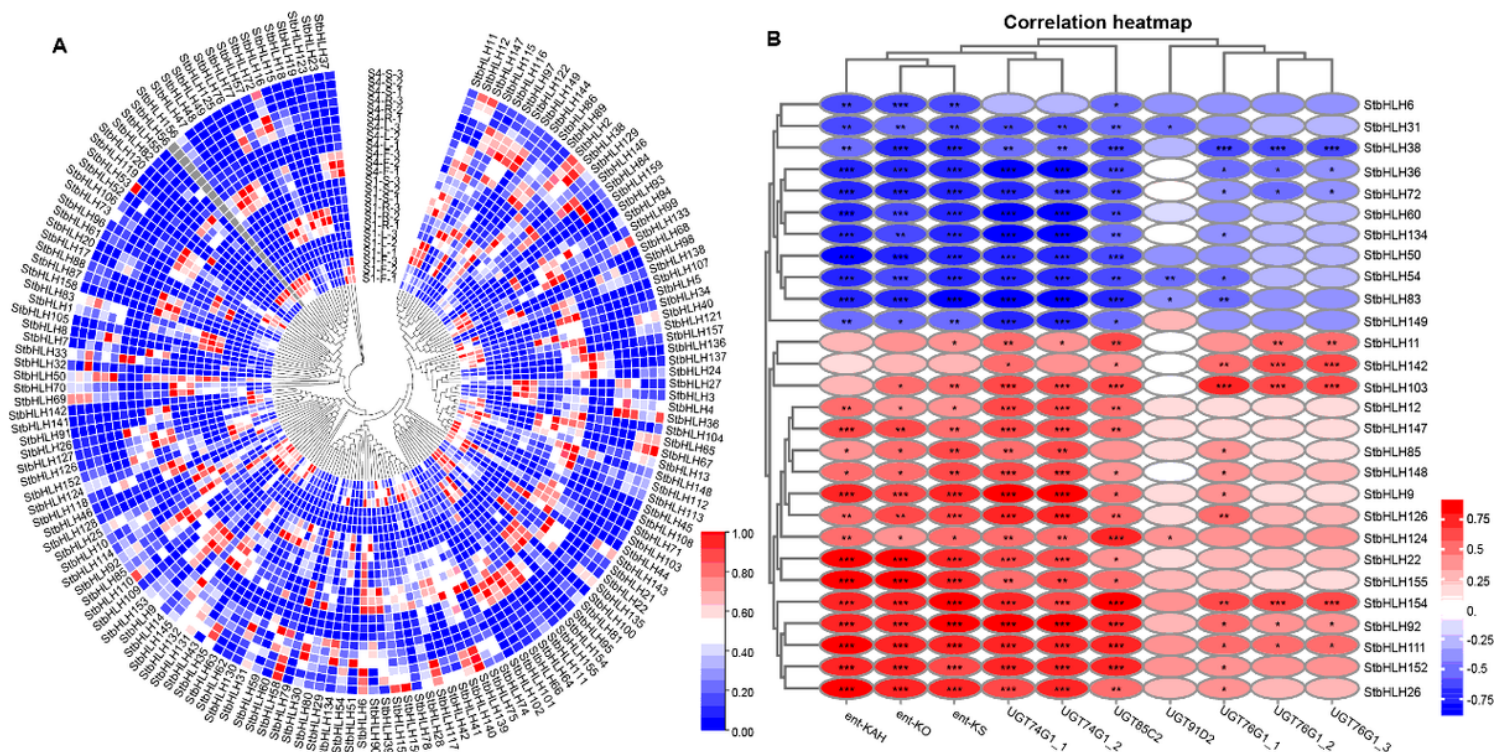


Figure 5

The expression patterns of *StbHLH* genes and the correlation between the expression between *StbHLH*s and structural genes involved in RA biosynthesis.

A. The expression patterns of *StbHLH* genes in different tissues of *S. rebaudiana*; B. the co-expression relationship between *StbHLH*s and structural genes involved in RA biosynthesis. Red circle: positively correlated; blue circle: negatively correlated. *, **, *** indicate a significant correlation at the 0.05, 0.01, or 0.001 level.

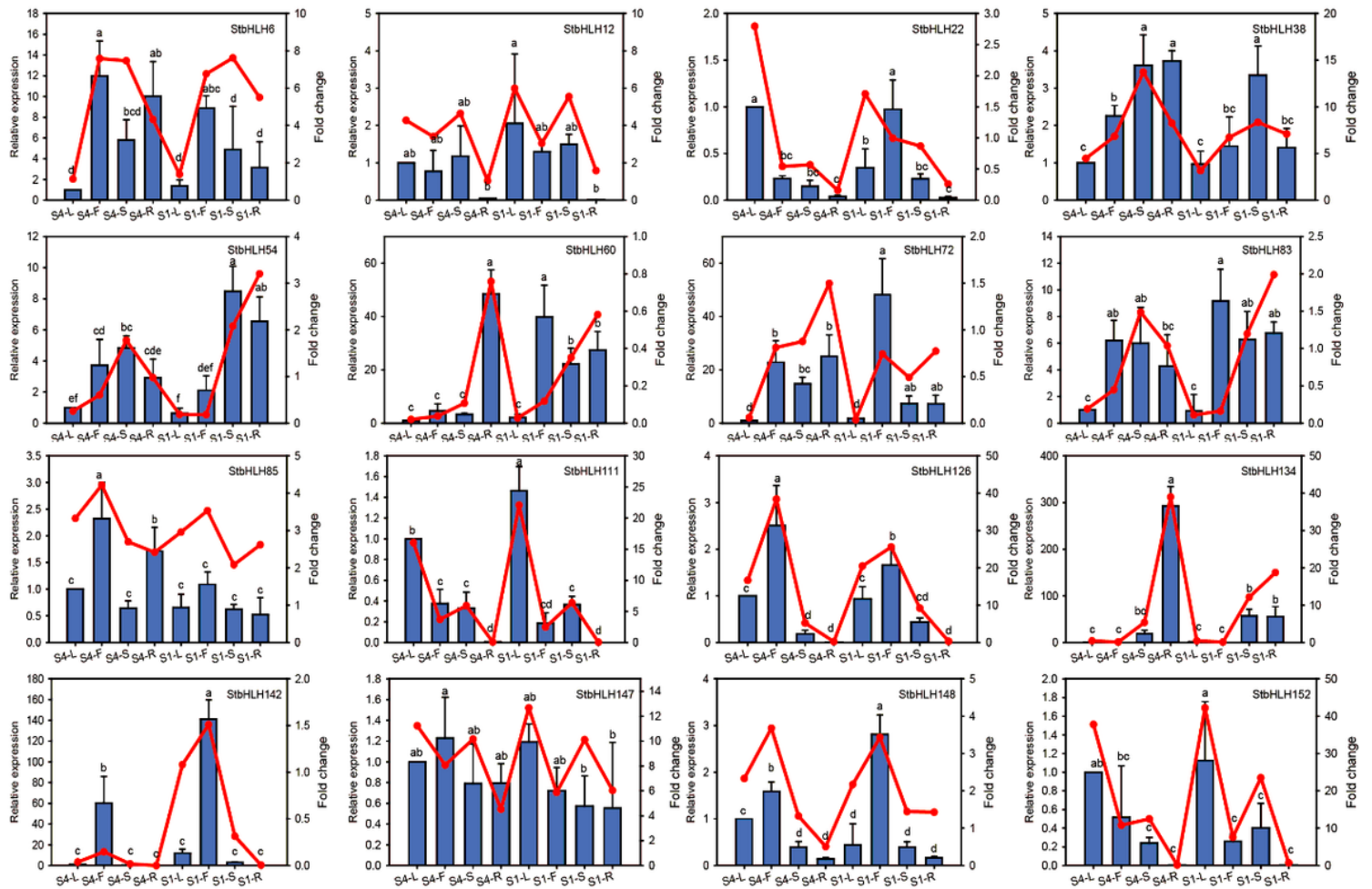


Figure 6

The expression patterns of candidate *StbHLH* genes in different tissues of *S. rebaudiana* measured by qPCR.

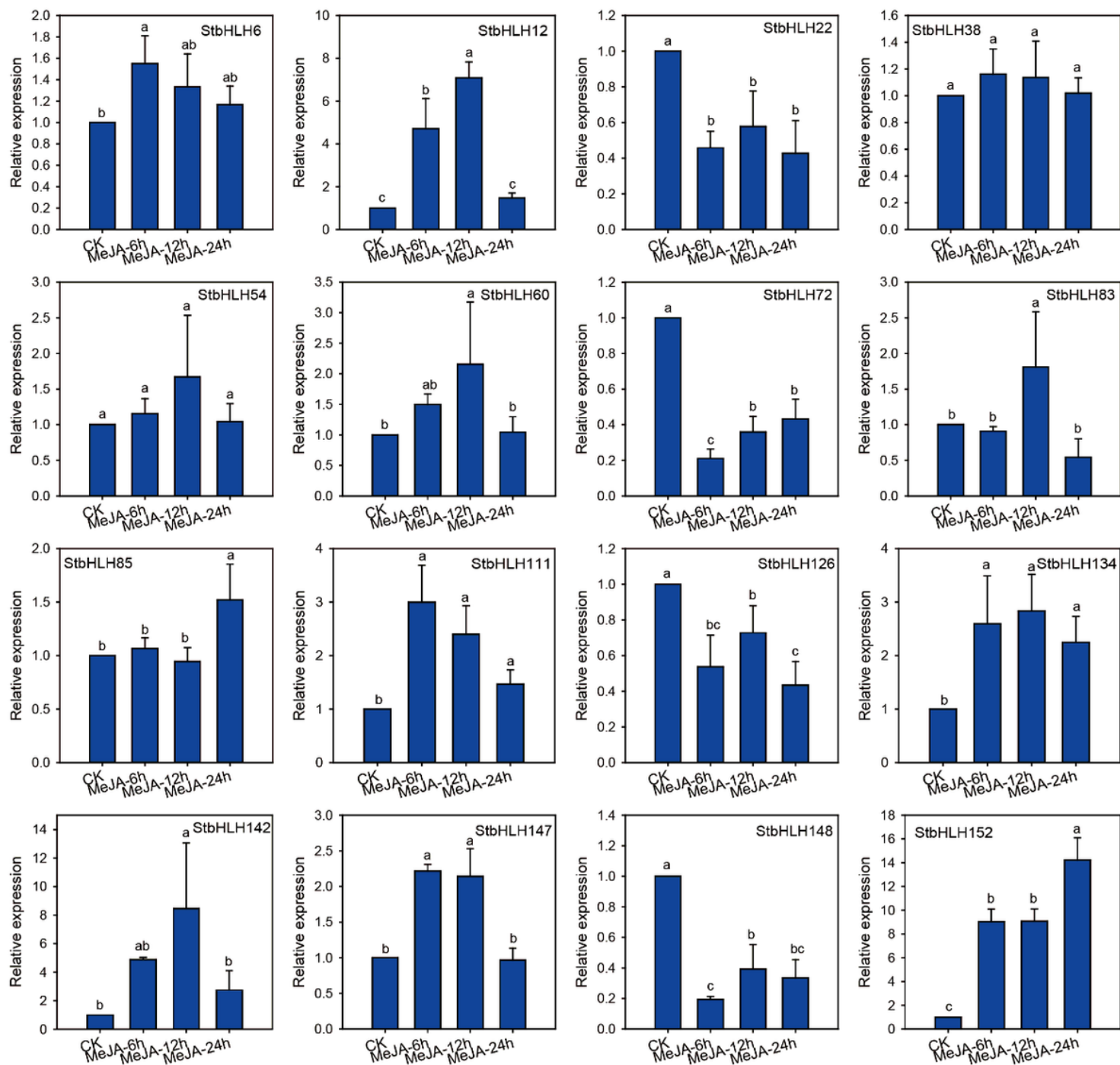


Figure 7

The expression patterns of candidate *StbHLH* genes under MeJA treatment measured by qPCR.

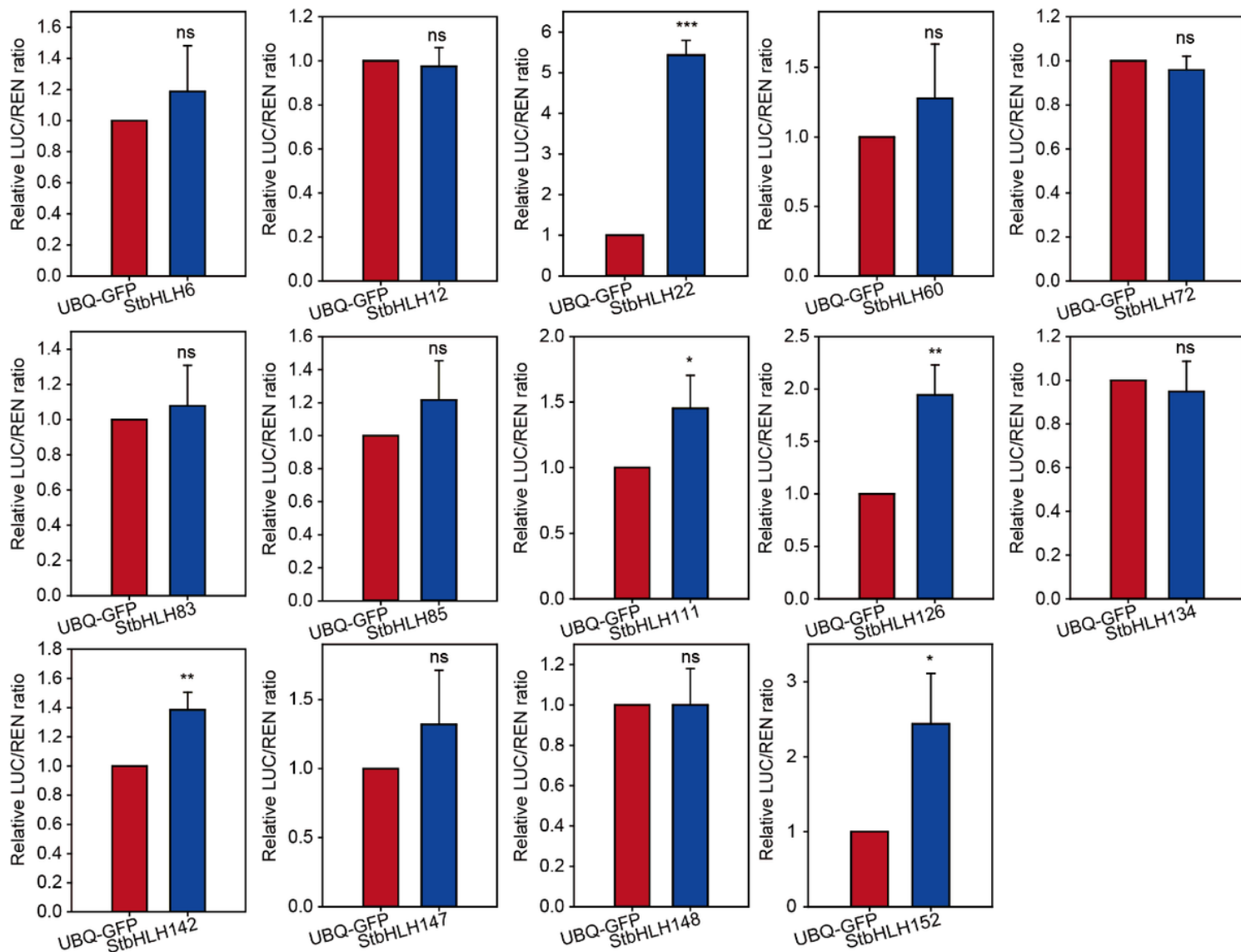


Figure 8

Transcriptional activation of UGT76G1 by candidate StbHLHs.

The values shown the means $\pm$ SD of dual LUC/REN ratio. *, **, *** indicate a significant at the 0.05, 0.01, or 0.001 level.

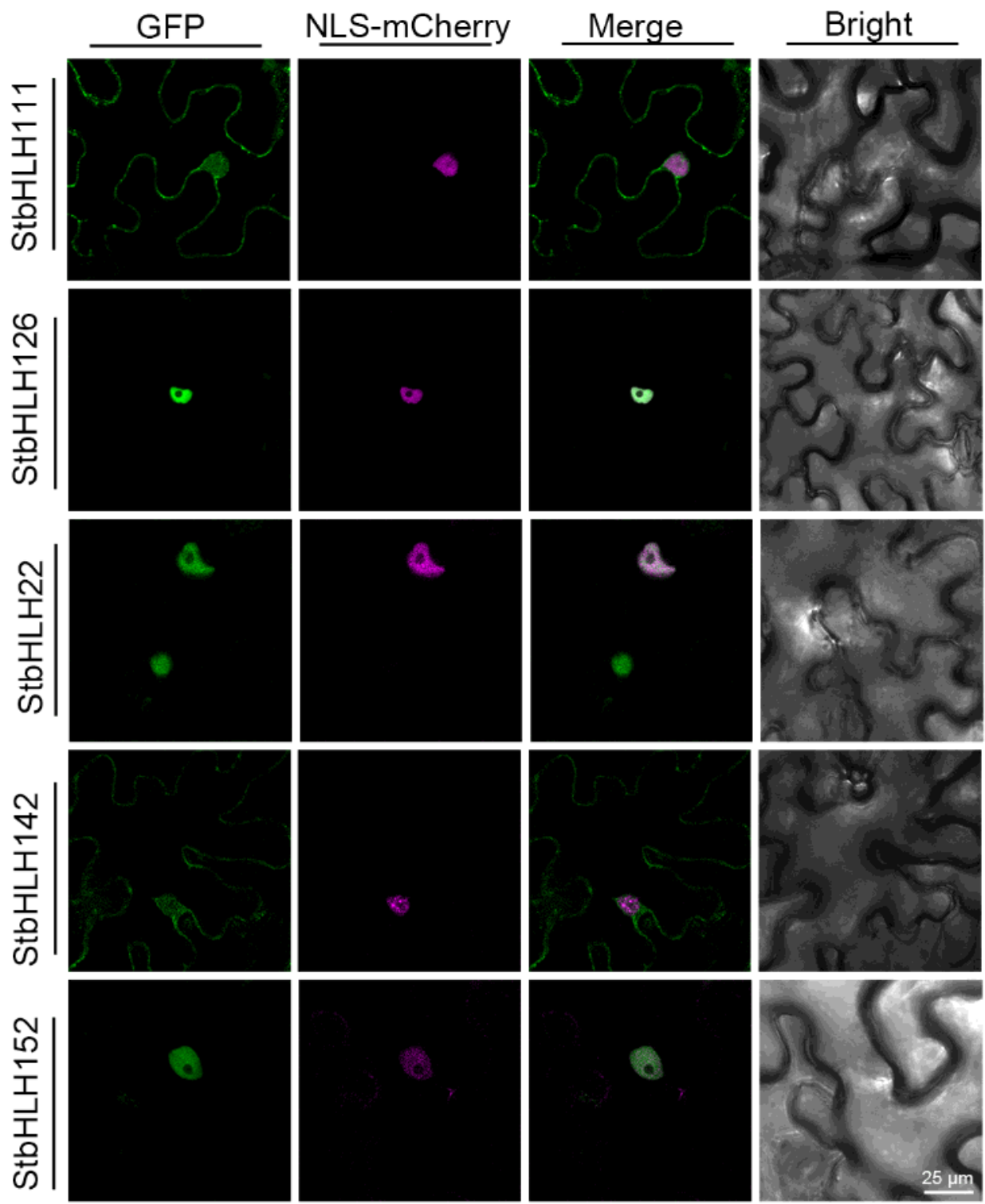


Figure 9

The subcellular localation of candidate StbHLH proteins.

Supplementary Files

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

- [TableS1.xlsx](#)
- [Table1b.xlsx](#)