BMC Plant Biology

RESEARCH

Open Access

Transcriptome-wide analysis of *AlbHLH* gene family in *Atractylodes lancea* (Thunb.) DC and functional exploration of AlbHLH3 in regulating terpenoid biosynthesis

Baiyi Zhang[1†], Peina Zhou[3†], Xiaoli Yang[1], Bingyan Gao[1], Yucheng Cao[1], Juan Deng[1,2], Xiao Huang[1,2], Kun Yu[1,2*] and Ling Gong[1,2*]

Abstract

Background *Atractylodes lancea* (Thunb.) DC., a medicinal species of Asteraceae with multi-pharmacological applications in rheumatic disorders and gastrointestinal syndromes due to its bioactive terpenoid-enriched essential oils. Although basic helix-loop-helix (bHLH) transcription factors are well known regulators of secondary metabolism biosynthesis pathways, their functional characterization in terpenoid regulation in *A. lancea* remains unexplored.

Results 176 *AlbHLHs* were systematically identified through multi-tissue transcriptomic profiling (leaf, stem, rhizome, rhizome buds) and categorized them into 21 evolutionarily conserved subfamilies via phylogenetic analysis with *Arabidopsis thaliana* bHLH proteins. Conserved motif architecture revealed high intra-subfamily conservation. Transcriptome analysis revealed that 36 *AlbHLHs* with significant differential expression across tissues, 19 of exhibited highly expressed in rhizomes and rhizome buds, which consistent with the accumulation trend of terpenoids determined by GC-MS analysis. Moreover, yeast one-hybrid (Y1H), electrophoretic mobility shift assay (EMSA) confirmed that AlbHLH3 could interact with the promoters of the key rate-limiting enzyme in terpenoid biosynthesis pathway genes 3-hydroxy-3-methylglutaryl-CoA reductase (*AlHMGR3*). Transient overexpression of *AlbHLH3* significantly upregulated expression of not only *AlHMGR3* but also downstream terpenoid pathway genes (*AlMVK*, *AlMVD*, *AlPMK*), suggesting its regulatory roles in terpenoid biosynthesis.

Conclusions This study contributes significantly to the understanding of the regulatory role of *AlbHLHs* in terpenoid biosynthesis and provides potential possibilities genetic modification strategies to enhance the medicinal value of *A. lancea*.

Keywords *Atractylodes lancea*, bHLH transcription factor, Terpenoids biosynthesis

†Baiyi Zhang and Peina Zhou contributed equally to this work.

*Correspondence:
Kun Yu
yukun_hbtcm@163.com
Ling Gong
gl1224@163.com

[1]College of Pharmacy, Hubei University of Chinese Medicine, Wuhan 430065, China
[2]Hubei Shizhen Laboratory, Wuhan 430065, China
[3]Nanjing Institute for Comprehensive Utilization of Wild Plants, Nanjing 211111, China

Background

Atractylodes lancea, known as "Cangzhu" in traditional Chinese medicine, is a perennial herb with a long and rich history of medicinal use. Its dried rhizomes have been prized for centuries for their remarkable therapeutic properties, dating back to the Shen-Nong-Ben-Cao-Jing, which documented its ability to eliminate dampness, fortify the spleen, dispel wind and cold, and even improve eyesight [1]. As a key component in numerous classic Chinese medicine formulas, such as Ping-Wei-San and Cangzhu Shaoyao Soup, *A. lancea* plays a pivotal role in treating rheumatism and digestive disorders [2, 3]. Sesquiterpenoids constitute a vital class of bioactive constituents in *A. lancea*, with key representatives including atractylon, *β*-eudesmol, and hinesol [4, 5]. Extensive modern pharmacological investigations have elucidated that *A. lancea* exhibits a spectrum of significant pharmacological activities, encompassing anti-inflammatory, anti-tumor, anti-gastric ulcer, hepatoprotective, anti-pulmonary injury, and hypoglycemic effects [6–9].

Given the substantial pharmacological evidence supporting the value of these terpenoids, significant attention has been focused on elucidating the terpenoids biosynthesis pathway in *A. lancea*. Sesquiterpenoids are mainly derived from the MVA (mevalonate) pathway, commences with the utilization of acetyl-CoA. The conversion of acetyl-CoA to IPP (isopentenyl diphosphate) and DMAPP (dimethylallyl diphosphate) transpires via a cascade of enzymatic reactions facilitated by AACT (Acetyl-CoA acetyltransferase), HMGS (3-hydroxy-3-methylglutaryl-CoA synthase), HMGR (HMG-CoA reductase), MVK (Mevalonate kinase), PMK (Phosphomevalonate kinase), MPDC (Mevalonate pyrophosphate decarboxylase) [10]. Then, FPP (Farnesyl pyrophosphate), the key precursor of sesquiterpenoids, is generated by FPPS (Farnesyl pyrophosphate synthase). TPS (Terpenoid synthase) and CYP450s (Cytochrome P450s) catalyzed the production of atractylon, *β*-eudesmol, and hinesol [11]. Among these enzymes, HMGR plays a critical role as a rate-limiting catalyst in the MVA pathway [12–15]. Beyond catalytic enzymes, upstream transcription factors (TFs)—including AP2/ERF, bHLH, MYB, WRKY, and bZIP families—play critical regulatory roles in sesquiterpenoid biosynthesis.

The bHLH family, one of the largest TF families in plants [16–18], is characterized by a conserved 50–60 amino acid domain, comprising an N-terminal basic region and the C-terminal helix-loop-helix (HLH) region [19]. The N-terminal basic region (10–15 amino acids) binds to the E-box motif (CANNTG), while the C-terminal HLH region (40–50 amino acids) enables the formation of homodimeric and heterodimeric protein complexes [20]. Generally, the bHLH family is divided into 15–25 subfamilies in plants [21].

Members of the bHLH gene family have been identified in an increasing number of plant species, such as *Areca catechu* [22], *Coptis chinensis* [23], *Dracaena cambodiana* [21] and other plants. TFs act as intermediaries in signal transduction pathways, modulating the expression of target genes. This regulation leads to the production of specific enzymes, which in turn are involved in the biosynthetic pathways of secondary metabolites, thereby indirectly contributing to their synthesis. For instance, overexpression of *LaMYC7* and *LaMYC4* has been shown to enhance the production of linalool and caryophyllene in *Lavandula angustifolia*, respectively [24, 25]. Similarly, *CpMYC2* and *CpbHLH13* of *Chimonanthus praecox* regulated the synthesis of linalool and *β*-caryophyllene [26], while *AabHLH112* overexpression increased *β*-caryophyllene levels in *Artemisia annua* [27]. In *Santalum album*, *SaMYC1* was involved in the santalol synthesis pathway by activating the promoters of both *SaSSy* and *SaCYP736A167* [28]. SIMYC2 of *Solanum lycopersicum* could improve the contents of terpene via directly regulating the expression of *SlJIG* [29]. In *Dendrobium officinale*, DobHLH4 served as a positive regulator in the biosynthesis of linalool [30]. Overexpression of *AaMYC3* from *A. annua* could boost the density of glandular-secreting trichome (GST) and the level of artemisinin significantly [31]. The accumulation of terpene trilactones could be increased by *GbMYC2* under JA induction in *Ginkgo biloba* [32]. As mentioned above, bHLH TFs are integral to the regulation of sesquiterpenoid biosynthesis. Consequently, elucidating the characteristics and functions of AlbHLHs is essential for enhancing sesquiterpenoid accumulation in *A. lancea*. However, the specific roles of bHLH TFs in modulating sesquiterpenoid biosynthesis in *A. lancea* remain incompletely understood. To address this knowledge gap, we conducted PacBio SMRT sequencing in *A. lancea* and identified 176 *AlbHLH* genes, which were divided into 21 subfamilies. Through integrated transcriptome and GC-MS analyses, we revealed that AlbHLH3 is likely involved in sesquiterpenoid biosynthesis in *A. lancea.* Molecular characterization using Y1H and EMSA demonstrated that AlbHLH3 could directly bind to the promoter region of *AlHMGR3*, a key gene in the sesquiterpenoid biosynthetic pathway. Functional validation demonstrated that *AlbHLH3* overexpression in *A. lancea* not only significantly upregulated *AlHMGR3* expression (1.7-fold, $p < 0.01$) but also coordinately activated downstream sesquiterpenoid biosynthetic genes. These findings suggest that AlbHLH3 plays a significant role in regulating sesquiterpenoid biosynthesis in *A. lancea*. The results of the present study offer valuable insights into the pivotal role of AlbHLHs in the biosynthesis of sesquiterpenoids in *A. lancea*. These findings lay a foundation for future investigations into the genetic regulatory

mechanisms underlying sesquiterpenoid biosynthesis in this species.

Materials and methods

Plant materials

Healthy 3-year-old *A. lancea* plants used in this experiment were collected from private lands near Zhanhe Reservoir in Yingshan County, Huanggang City, Hubei Province, China (29°36.32'N, 109°15.98'E) in October 2022. Prior to sampling, the permission was obtained from the landowners. They were identified as *A. lancea* by Professor Xiuqiao Zhang (Hubei University of Chinese Medicine). Six individual plants were selected based on stringent criteria: absence of pest infestation or pathogenic symptoms, structural integrity of aerial and subterranean tissues, fully expanded leaves without chlorosis or necrosis, and robust rhizomes free from rot or desiccation. From each plant, the stems (S), leaves (L), rhizomes (R), and rhizome buds (RB)were collected using sterilized instruments. Tissues were rinsed three times with sterile water, blotted dry with sterile filter paper, and immediately flash-frozen in liquid nitrogen. Samples were stored at −80 °C for subsequent analysis, with three biological replicates maintained per group.

Determination of terpenoid in *A. lancea*

To analyze the terpenoid compounds in S, L, R and RB, headspace solid-phase micro-extraction (HS-SPME) combined with gas chromatography-mass spectrometry (GC-MS) was employed. Briefly, the tissue samples were freeze-dried under vacuum to a constant weight and powered using liquid nitrogen. Approximately 0.3 g of each powdered sample transferred into a sealed headspace vial, then extracted at 50 ℃ for 30 min using a DVB/CAR/PDMS (SPME fiber assembly, divinylbenzene, Carboxen, Polydimethylsiloxane, Supelco, Sigma-Aldrich, Shanghai, China). The extracted compounds were desorbed for 5 min utilizing a GC-MS QP 2010 Ultra system, equipped with an Agilent 19091 S-433HP-5ms capillary column (30 m × 0.25 mm × 0.25 μm). High-purity helium (> 99.999%) served as the carrier gas, with splitless injection adopted and a column flow rate of 1 mL/min. The temperature program was set as follows: the initial temperature of 100 °C for 2 min, followed by an increased to 180 °C at a rate of 2℃/min for 6 min, a further raised to 270 °C at 30 °C/min, and a final hold for 5 min. Additional parameters included an inlet temperature of 280℃, an electron ionization source operating at 200℃, and a quadrupole temperature maintained at 150℃. Full MS SCAN mode was selected within an m/z range of 20 ~ 300.

PacBio SMRT sequencing and RNA-seq in *A. lancea*

The total RNA was extracted from S, L, R and RB using an RNA prep Pure Kit for Plants (Tiangen, Beijing, China) following the manufacturer's instructions. The quality, quantity and concentration and integrity number (RIN) of RNA were assessed using Nanodrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA) and Agilent 2100 (Agilent Technologies, Palo Alto, CA, USA). The RNA of samples should meet the following conditions: OD260/280 = 1.8–2.2; the concentration of RNA > 40 ng/μL; RIN ≥ 6.5. Qualified RNA from each sample was pooled for cDNA library construction using the SMARTer™ PCR cDNA Synthesis Kit. The library was then sequenced on the PacBio Sequel System (Pacific Biosciences, Menlo Park, CA, USA). Raw sequencing dates were filtered using SMRTLink v5.0 with default parameters (minLength = 200, minReadScore = 0.75). Circular consensus sequences (CCS) were obtained through self-correction of subreads from the same polymerase reads. These sequences were categorized into full-length non-concatemer (FLNC) and non-full-length sequences based on the presence of 5' and 3' primers and a poly-A tail. FLNC reads were clustered the using ICE to obtain Cluster consensus sequences, which were then polished using non-full-length sequences to generate high-quality sequences for subsequent analysis.

Additionally, total RNA from the same tissues was used constructed cDNA library and sequence on the Illumina HiSeq 2000 platform. Raw sequencing data from Illumina HiSeq were cleaned and assembled *de novo* using Trinity software to generate transcripts and unigenes. To enhance the accuracy and completeness of the transcriptome, FLNC sequences from PacBio sequencing were integrated with the Illumina-assembled transcripts. Errors in the Illumina data were corrected using long-read information from PacBio, facilitated by tools such as LoRDEC, thereby improving the overall quality of the transcriptome assembly. Gene expression of fragments per kilobase of transcript per million mapped reads (FPKM) was calculated using StringTie v2.2.3. Based on three biological replicates, differentially expressed genes (DEGs) were identified via the DESeq2 R package using a threshold of false discovery rate (FDR) < 0.05 and |log2 (Fold Change)| >1. Then, for in-depth functional analysis of transcripts, multiple databases were used to annotate their functions. These include NR (NCBI-non-redundant protein sequences), egg NoG (evolutionary genealogy of genes: Non-supervised Orthologous Groups), KEGG (Kyoto Encyclopedia of Genes and Genomes), GO (Gene Ontology) and Swiss-Pro. The annotation of these databases provides insight into the function of the proteins encoded by the transcripts.

Identification of the *AlbHLH* gene family

To identify members of the *AlbHLH* gene family, we retrieved the HMM profile for the conserved bHLH domain (PF00010) from Pfam database (http://pfam.xfam.org/) based on the full-length PacBio sequencing reads. Using TBtools software, we filtered transcriptome data with a threshold of E-value < e^{-5} to isolate relevant protein sequences. A reliable reference framework was established by extracting the *A. thaliana* bHLH gene family from the Arabidopsis Information Resource (https://www.arabidopsis.org/). A BLAST search was conducted against *A. lancea* proteins using *A. thaliana* bHLH protein sequences as reference sequences. To further analyze bHLH domains, eliminate redundant or incomplete bHLH domain sequences, and deduce the *A. lancea* bHLH family, we used the Batch CDD search database and plantRegMap database. Moreover, the basic characteristics of AlbHLH protein family, including amino acid length, molecular weight, isoelectric point, grand average of hydrophobicity, instability index and aliphatic index were analyzed using the ExPasy website (https://web.expasy.org/). Furthermore, WoLF PSORT website (https://wolfpsort.hgc.jp/) was utilized for subce llular localization predictions.

Phylogenetic analysis and gene structure of the *AlbHLH* gene family

Multiple sequence alignments of the full-length protein sequences of AlbHLHs and AtbHLHs were generated using ClusterW in MEGA X. Phylogenetic trees were constructed using the Maximum-Likelihood (ML) method, allowing free end gaps and a bootstrap value of 1000. AlbHLHs were classified according to AtbHLHs gene family. MEME Suite 5.4.1 (http://meme-suite.org/meme/tools/meme) was used to identify conserved motifs in AlbHLH protein sequences with parameters set to 6–50 amino acids and 20 motifs.

Gene expression analysis and RT-qPCR validation among different tissues in *A. lancea*

AlbHLH genes were selected from the identified DEG based on PF00010, and TBtools was used to map the expression of these DEGs of *AlbHLH* across the four different tissues. The real-time quantitative PCR (RT-qPCR) was performed to validate the expression of 12 highly expressed key genes in rhizomes or rhizome buds, randomly selected from DEGs of *AlbHLHs*. The *glyceraldehyde-3-phosphate dehydrogenase* (*GAPDH*) gene served as the internal control. All the amplification primers used are list in Table**S1**. The RT-qPCR was conducted using an iCycler iQ Real-Time PCR Detection System (Bio-Rad Laboratories, USA) with a Quant One Step RT-PCR (SYBR Green) kit (Tiangen, Beijing, China). Amplification reactions were performed according to the following program: 50 °C for 30 min, 95 °C for 2 min, followed by 40 cycles of 94 °C for 20 s, 55 °C for 20 s, and 68 °C for 20 s. The expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method with three replicates.

Subcellular localization analyses

The open reading frame (ORF) (stop codon-excluded) of *AlbHLH3* was amplified with primers (Table**S1**) and subsequently cloned into the pCNG-GFP via homologous recombination to construct fusion expression vector pCNG-AlbHLH3-GFP. The empty of pCNG-GFP was used as a control. The construct and empty vector control were transformed into *Agrobacterium tumefaciens* GV3101 using chemical transfection. Transformants were selected on LB plates containing 50 μg/mL kanamycin and 50 μg/mL rifampicin, with colony PCR verification. For infiltration, cultures were grown to OD600 = 0.6–0.8 in LB medium with appropriate antibiotics at 28 ℃, 200 rpm, then resuspended in infiltration buffer (10 mM MES, 10 mM $MgCl_2$, 100 μM acetosyringone) at a ratio of 1:1. Four-week-old *N.benthamiana* leaves were pressure-infiltrated using 1 mL needleless syringes. Infiltrated plants were incubated in a growth chamber maintained at 25 ℃, under a photoperiod of 16 h light and 8 h dark for 3 days. The *N. benthamiana* leaves were cut into 1 cm × 1 cm pieces incubated with 4',6-diamidino-2-phenylindole (DAPI) for 10 min, which was used for nuclear staining at room temperature, followed by three 5-min PBS washes. Confocal imaging was performed using a Leica TCS SP5 laser scanning confocal microscope (Leica Microsystems, Germany). GFP signals were acquired using 488 nm excitation with emission collected at 500–550 nm, while DAPI was excited at 364 nm with emission at 420–480 nm.

Yeast one-hybrid assays (Y1H)

Based on our previously cloned *AlHMGR* genes, we isolated the 5'-flanking regions of *AlHMGR3* using a Genome Walking Kit (TaKaRa, Beijing, China) according to the manufacturer's protocol. Specific primers were designed to facilitate the full-length cloning of the amplified promoter sequences (Table**S1**). Putative cis-regulatory elements were predicted using PlantCARE and PLACE databases. In plants, the vast majority of bHLH family members belong to the Group B. The Group B bHLH family typically interacted with the E-box containing CANNTG motif [23]. The segment of *AlHMGR3* (458 bp) which based on the E-box and the ORF of *AlbHLH3* (1773 bp) were amplified using primers (Table**S1**). The *AlHMGR3* promoter fragment was inserted into the pAbAi vector to create bait vectors pAbAi- *AlHMGR3*-Pro, and the ORF of *AlbHLH3* was cloned into the pGADT7 to create the prey recombinant plasmid pGADT7-AlbHLH3. Yeast self-activation assay

determined an AbA selection concentration of 100ng/ml. The prey and bait plasmids were transformed into Y1H-Gold yeast cells, cultured on SD/-leu and SD/-leu selection medium supplemented with 100 ng/ml AbA, and yeast colonies on the plates was observed to assess interactions. The empty pGADT7 served as negative control.

Electrophoretic mobility shift assay (EMSA)

The ORF of *AlbHLH3* was cloned into the pet28a-MBP vector. Both the recombinant construct (pET28a-MBP-AlbHLH3) and empty vector control were transformed into *E. coli* expression strain Rosetta (DE3). The recombinant proteins expression was induced with 0.5 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) at 16 ℃ for 20 h. Cell were then harvested by centrifugation (5000 rpm, 4 ℃, 15 min) and lysed by sonication for 30 min in chilled extraction buffer (50mM $NaH_2PO_4{\cdot}2H_2O$, 300mM NaCl, and 10mM imidazole). After centrifugation at 12,000 rpm for 15 min, the supernatant was collected and filtered through a 0.45 μm filter. The Ni-NTA Beads (Smart-Lifescience, China) was used for affinity purification. After column equilibration with lysis buffer, the supernatant was loaded at 1 mL/min. Weakly bound contaminants were removed with wash buffers containing 20mM and 50mM imidazole, and the fusion protein was eluted with elution buffer (50mM $NaH_2PO4{\cdot}2H_2O$, 300mM NaCl, and 250mM imidazole). The eluted protein was concentrated using a 30-kDa MWCO ultrafiltration device and analyzed by SDS-PAGE to confirm purity and molecular weight. EMSA was performed using the Chemiluminescent EMSA Kit (Beyotime, China). EMSA probe containing E-box was synthesized and labeled with biotin at the 5′ end. The probes were incubated with AlbHLH3 protein for 20 min. The mixture was resolved on a 6% (w/v) non-denaturing polyacrylamide gel, and then transferred onto a nylon membrane, crosslinked under UV light subsequently, and detected using ChemiDoc XRS (Bio-Rad Laboratories, USA). The probe sequence is list in Table**S1**.

Transient transformation assays in *A. lancea*

The transient transformation assay was adapted by Cao et al. [33]. The ORF of *AlbHLH3* was cloned into the 35S: RUBY to create AlbHLH3-RUBY fusion construct. The amplification primers used are list in Table**S1**. Both this construct and the empty 35S: RUBY vector (negative control) were independently transformed into *A.rhizogenes* strain Ar.A4. Transformed bacteria were cultured in TY liquid medium at 28 ℃ until reaching an OD600 of 0.8, harvested by centrifugation (4000 rpm, 15 min) and resuspended in infiltration buffer (10 mM MES, 10 mM $MgCl_2$, 100 μM acetosyringone) at a ratio of 1:1. For infection, 1-month-old sterile *A. lancea* seedlings were transplanted to soil for 3 weeks. After gently removing soil by washing, the main roots and lateral roots were excised, and a fresh cut was made near the soil-surface contact point. Seedlings were the fully submerged in the *Agrobacterium* suspension and subjected to vacuum infiltration for 5 min. Subsequently, seedlings were transferred to a seedling tray containing pure vermiculite and maintained under controlled conditions (25 ℃, 16 h light/8 h dark cycle). After about 2 months, positive hairy roots which express red signal were collected and the expression of terpenoid biosynthetic pathway genes was analyzed via RT-qPCR, as detailed in the "Gene expression analysis and RT-qPCR validation among different tissues in *A. lancea*" section. The primers used are list in Table**S1**.

Results

Analysis of PacBio SMRT sequencing and RNA-seq analysis of different tissues

The mixed samples including R, RB, L, and S of the *A. lancea* were sequenced using the PacBio SMRT platform (Fig. 1A). The sequencing analysis generated 6,560,183 subreads, which were processed to yield 352,632 CCS reads. Among these, 250,071 reads were successfully classified as FLNC. The FLNC/CCS percentage was 70.91%, which confirms that the library construction efficiency was satisfactory. Followed that, a total of 128,515 consensus sequences were generated. Besides, we conducted transcriptome sequencing on four tissues: R, RB, L, and S (Fig. 1B). Complementary RNA-seq analysis of individual tissues demonstrated exceptional data quality, with Q20 and Q30 scores exceeding 97% and 93% respectively (Table S2). Differential expression analysis revealed 18,945 significantly DEGs (Fig. 1B). In details, there were 10,054 DEGs in R vs. L with 5567 downregulated and 5387 upregulated, 10,971 DEGs in RB vs. L with 5980 downregulated and 4991 upregulated, 7505 DEGs in R vs. S with 4421 downregulated and 3084 upregulated, 6195 DEGs in L vs. S with 3092 downregulated and 3103 upregulated, and 543 DEGs in R vs. RB with 383 downregulated and 160 upregulated (Fig. **S1**). In the pairwise comparisons of transcriptomic profiles, the contrasts between RB vs. L and R vs. L exhibited the highest number of DEGs, whereas the R vs. RB comparison yielded the fewest DEGs. The GO and KEGG enrichment analysis on the DEGs was shown in Fig.**S1** and Fig. 1. Given the substantial transcriptomic differences observed in RB vs. L and R vs. L comparisons, we focused subsequent analyses on these tissue pairs. KEGG enrichment analysis revealed significant enrichment for sesquiterpenoid and terpenoid biosynthesis pathways in both comparisons (Fig. 1C). The RB vs. L comparison additionally showed enrichment in terpenoid backbone biosynthesis. These findings suggest that the RB vs. L and R vs. L comparisons are particularly enriched in terpenoid metabolites, with

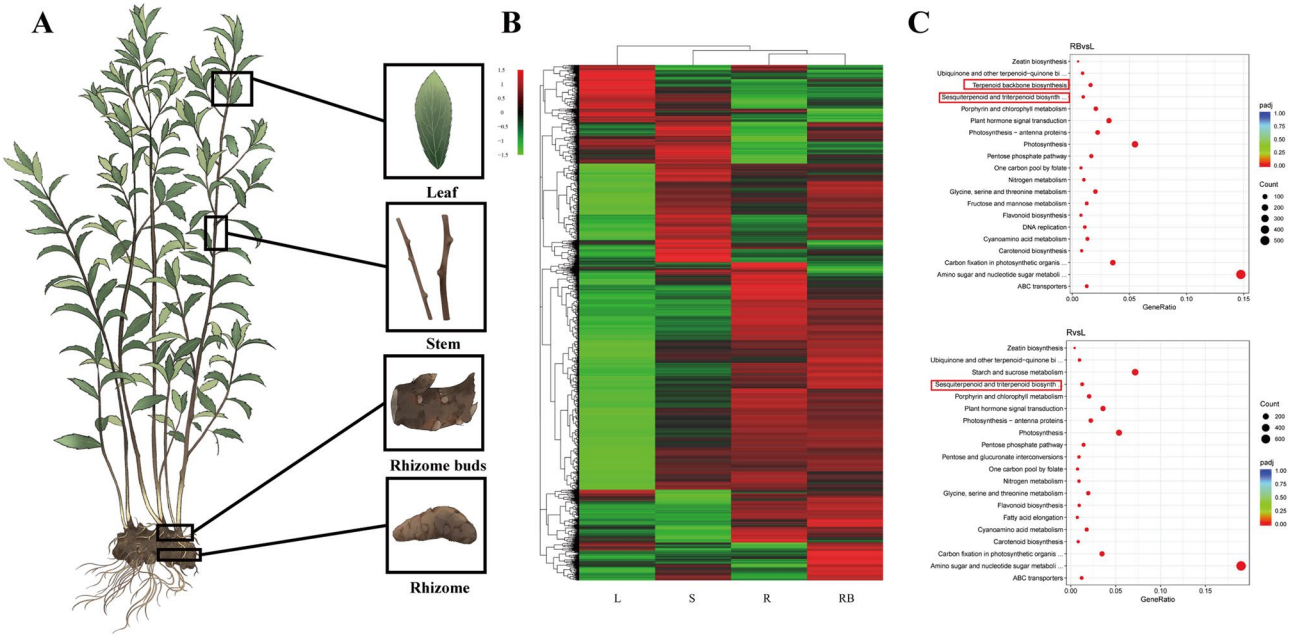

Fig. 1 The RNA-seq analysis of different tissues. **(A)** The different tissues of *A. lancea*. **(B)** The DEGs of L, S, R, and RB. **(C)** The KEGG enrichment of RB vs L and R vs L

a notable emphasis on sesquiterpenoids. This pattern aligns with the traditional Chinese medicine perspective, where both RB and R are recognized as medicinal sites rich in bioactive terpenoids, especially sesquiterpenoids.

Identification of AlbHLH genes in *A. lancea*

In our study, a 176 putative *AlbHLHs* were identified and sequentially named as *AlbHLH1* to *AlbHLH176*. To elucidate the fundamental attributes of these *AlbHLH* genes, the ExPasy and WoLF PSORT were employed, yielding comprehensive data summarized in Table S3. This translated into a wide spectrum of molecular weights, spanning from 10.99 kDa to 79.41 kDa. The predicted proteins encoded by *AlbHLHs* exhibited considerable variability in their amino acid lengths, ranging from a minimum of 100 amino acids in AlbHLH153 and AlbHLH173 to a maximum of 718 amino acids in AlbHLH13. Correspondingly, their molecular weights ranged from 10.99 kDa to 79.41 kDa. The predicted isoelectric point (PI) of these proteins varied significantly, spanning from 4.58 (AlbHLH115) to 10.02 (AlbHLH129), with 126 proteins classified acidic proteins (PI < 7) and 51 basic proteins (PI > 7). The hydrophobicity values ranged from − 1.042 (AlbHLH136) and 0.267 (AlbHLH153), indicating most AlbHLH proteins possess a hydrophilic nature, as evidenced by their predominantly negative hydrophobicity values. Among these proteins, AlbHLH129 and AlbHLH133 were stable proteins, while the remaining proteins were classified as unstable proteins. The aliphatic index ranged from 51.89 to 123.7. According to subcellular localization predictions, the majority of AlbHLH proteins are localized in the nucleus, which is consistent with their role as transcription factors. However, a few proteins were predicted to be localized outside the nucleus, including 6 AlbHLHs predicted to localize to the chloroplast (AlbHLH1, 50, 141, 147, 151 and 152), 5 to the Golgi apparatu (AlbHLH100, 109, 110, 111 and 150), and 4 to the cytoplasm (AlbHLH101, 138, 139, 146) (Table S3).

Phylogenetic analysis of the AlbHLHs

To elucidate the phylogenetic relationships among AlbHLHs and other species, and to predict their potential functions, a phylogenetic tree was constructed for the 176 AlbHLH proteins based on the evolutionary classification of AtbHLHs in *A. thaliana* [34]. The resulting phylogenetic tree revealed that AlbHLHs could be divided into 21 subfamilies, designated as I to XV (Fig. 2), consistent with the subfamily divisions observed in other plant species, including *A. thaliana* [34]. Among these, subfamilies XⅡ was the largest subfamilies with 34 AlbHLHs, followed by ⅢI (d + e) with 25 members. In contrast, subfamilies Ib, Ⅱ, XⅢ, XV, VI and VⅢb were the smallest subfamilies with only one AlbHLH protein. Given that AlbHLH proteins clustering within same clade characterized AtbHLH proteins are likely to share similar functions, the phylogenetic relationships between these proteins provide insights into their functional conservation and divergence. Therefore, the phylogenetic tree analysis serves as a valuable tool for predicting the functions of AlbHLH proteins based on their evolutionary relatedness to well-characterized AtbHLH proteins.

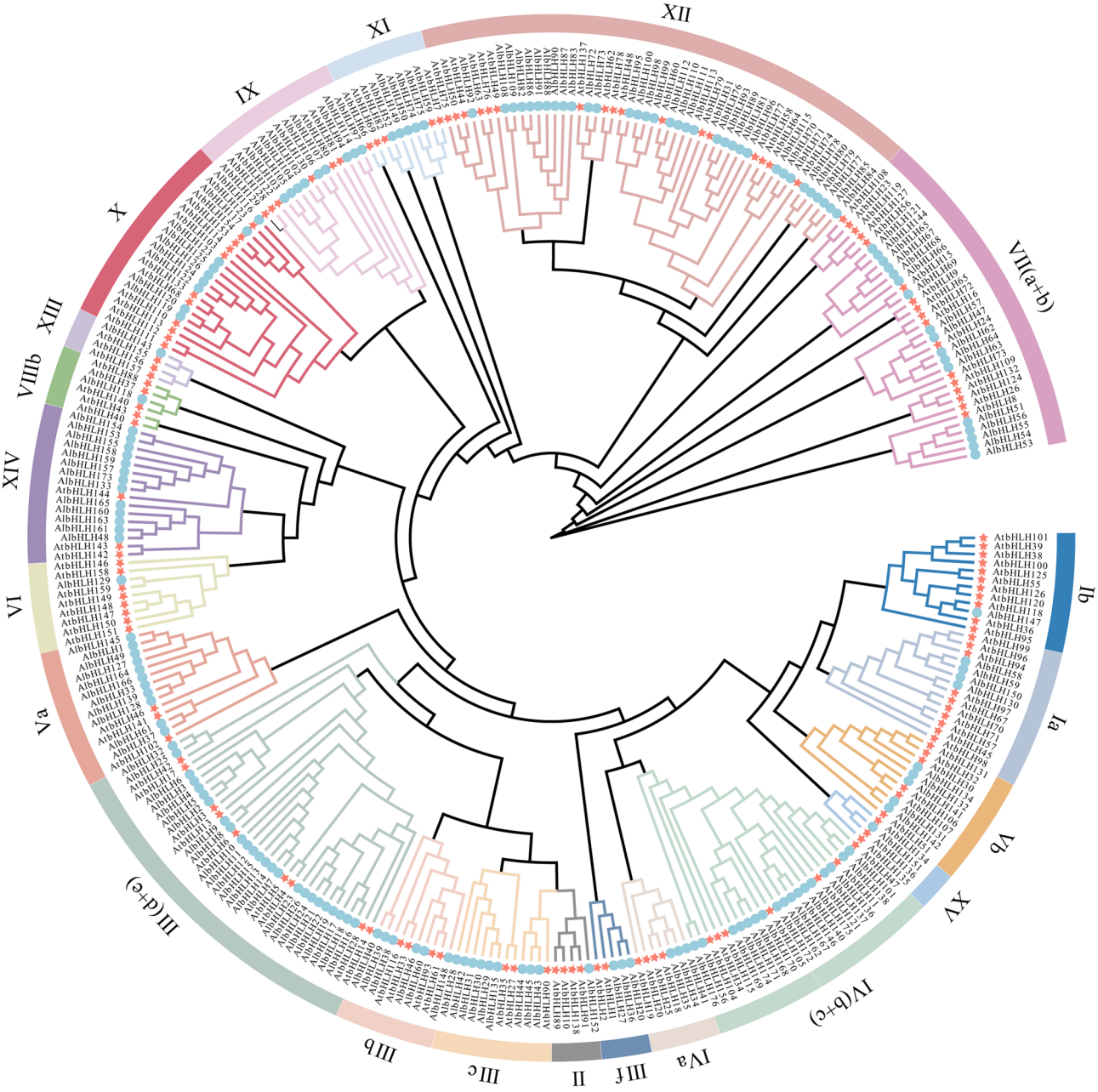

Fig. 2 Phylogenetic tree of the bHLH proteins from *A. thaliana* and *A. lancea*. MEGA X was used to construct an ML phylogenetic tree. The tree was constructed based on the complete protein sequences with 1000 bootstrap replicates. The different subfamilies represented by variously colored branches

Conserved motifs analysis

To further investigate the structural diversity and functional conservation of AlbHLH proteins, we utilized MEME and TBtools to identify and analyze conserved motifs (Fig. 3). Twenty conserved motifs, designated as motif 1–20, were represented by different colors. The motif 1 and motif 2 were ubiquitous in AlbHLH proteins across all subfamily, and are consistently adjacent to each other, collectively forming the conserved bHLH domain. Notably, despite motif differences, members of the same subfamily of AlbHLH typically exhibit similar motif compositions. For example, all five AlbHLHs of subfamily XI contained motif 1, 2, and 3. Similarly, the closely clustered of AlbHLH153/154/159/173/157/158/133/155 in subfamily XIV shared identical motif 2, 10 and 17. It was represented that AlbHLHs within the same subfamily are relatively conserved and their functions might be similar. However, significant differences in the types and quantities of conserved motifs were observed between different subfamilies, and some motifs were specific to individual subfamilies, potentially indicating functional diversification within the bHLH gene families. For

Fig. 3 Conserved motif analysis of 176 genes in *A. lancea*. 20 conservative motifs were identified using MEME, and differently colored boxes represent these motifs respectively

instance, subfamily III(d + e) exhibited the largest number of motif types, with each motif positioned relatively conservatively. Besides, Motif 5 was unique to subfamily III(d + e), and motif 9 was specific to subfamilies Va, XIV, and XⅡ. These results indicated that the specific functions of different subfamilies may be related to these specifically presented motifs. As Fig. 4 showed the bHLH conserved domain logos of AlbHLHs, which including two helixes, the basic region, and a loop region. It contains 50–60 relatively conserved amino acids. The motif 1 and motif 2 were located in the bHLH domain. All the AlbHLHs contained the conserved bHLH domain except AlbHLH33/37/38/69/135/170. Our analysis revealed that the 176 AlbHLH proteins could be categorized into two major groups based on their bHLH domain sequence profiles: 8 (4.5%) atypical AlbHLHs, which were non-DNA-binding proteins, and 168 (95.5%) typical AlbHLHs (Fig. 4). The typical AlbHLHs were further subdivided into two subcategories: 133 E-box-binding proteins and 35 non-E-box-binding proteins (Fig. S2). The sequences of the 168 typical AlbHLHs exhibited divergence in their basic regions but were relatively conserved in their HLH regions, particularly within the loop region (Fig. 4). A comparable pattern was observed for the non-E-box-binding proteins (Fig. S2), suggesting their close relationship to the atypical AlbHLHs. In contrast, the residues in the basic region of the E-box-binding proteins were more conserved than those in the HLH region.

Expression profile of *AlbHLHs* among different tissues in *A. lancea*

Four tissues (R, RB, S, L,) of *A. lancea* were selected for both transcriptomic and metabolic analysis. Among 176 *AlbHLH* genes in four different tissues of *A. lancea*, 36

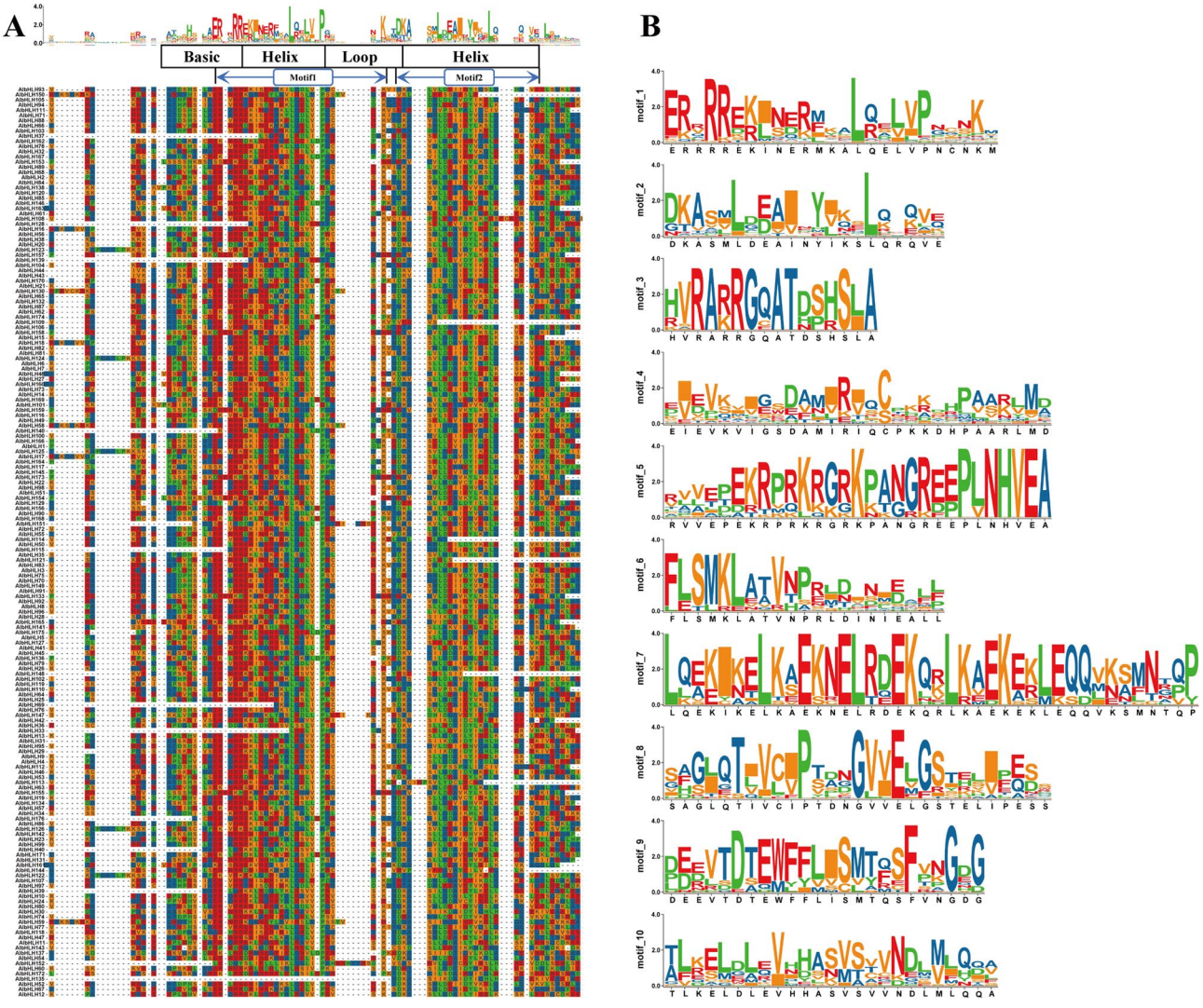

Fig. 4 The bHLH domain is highly conserved across all AlbHLH proteins. **(A)**The Multiple sequence alignment of AlbHLH proteins. **(B)** The schematic representation of ten motifs in AlbHLH proteins is elucidated by MEME

AlbHLHs were significantly differentially expressed. In details, 19 genes exhibited significantly expression in R and RB, 8 genes were highly expressed in L, and 9 genes shown the highest expression in S (Fig. 5A). To verify the accuracy of transcriptome data, RT-qPCR analysis was performed on 12 randomly selected *AlbHLHs* from the 19 aforementioned genes. The results were highly consistent with the transcriptome analysis, confirming the authenticity and reliability of transcriptomic data (Fig. 5B). Based on the GC-MS analysis of R, RB, S, L of *A. lancea*, both hinesol and atractylon exhibited the highest relative contents in the RB and R, especially in RB (Fig. 5C). The *β*-eudesmol displayed the highest concentration in the R, with the RB showing the second-highest level. Further correlation analysis between gene expression and metabolite content revealed that *AlbHLH55*, *AlbHLH148*, *AlbHLH150*, *AlbHLH73*, *AlbHLH117*, *AlbHLH3*, *AlbHLH84*, *AlbHLH102*, and *AlbHLH108* exhibited positive correlations with the sesquiterpenoid hinesol in *A. lancea*. Among these, *AlbHLH3* and *AlbHLH102* also showed positive associations with atractylon. Additionally, *AlbHLH41* was positively correlated with the content of *β*-eudesmol (Fig. 5D). The results imply that these transcription factors may play key regulatory roles in sesquiterpenoid biosynthesis in *A. lancea*.

Phylogenetic analysis and subcellular localization of AlbHLH3

To gain further insights into the phylogenetic relationships and potential functions of these candidate genes, a neighbor-joining phylogenetic analysis was performed. This analysis included the 19 AlbHLHs and functionally characterized bHLHs from other species (Fig. 6A). In the phylogenetic analysis, AlbHLH3 was found to cluster within a distinct clade that includes several functionally characterized bHLH proteins, such as CpbHLH3, CpMYC2, LaMYC4, and AtMYC2. This phylogenetic positioning suggests that AlbHLH3 may share similar biological functions with these well-studied proteins. Specifically, CpMYC2 and CpbHLH13 have been identified as key regulators of linalool and *β*-caryophyllene biosynthesis [26]. Overexpression of *LaMYC4* has been

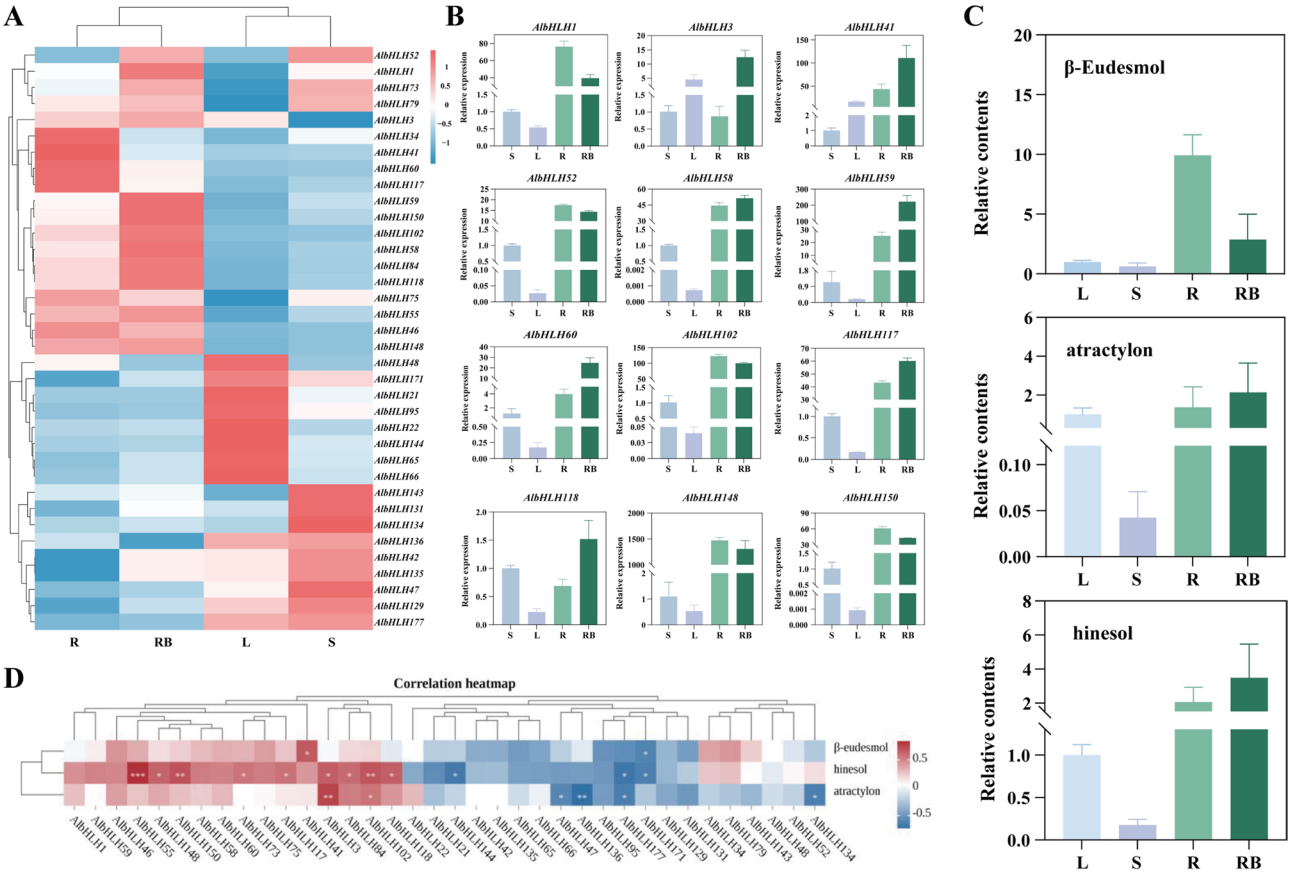

Fig. 5 Expression patterns of *AlbHLHs* in different tissues. **(A)** Cluster heatmap of differentially expressed *AlbHLHs* in four tissues. The horizontal coordinate represents the sample name, and the vertical coordinate is the normalized values of DEGs FPKM. The expression level is indicated by a change in color, blue indicates a lower expression level, red indicates a higher expression level. **(B)** Expression levels of 12 *AlbHLHs* detected by RT-qPCR in four different tissues of *A. lancea*. Error bars, mean ± standard deviation. n = 3 independent experiments were conducted. **(C)** The relative contents of *β*-eudesmol, hinesol and atractylon in different tissues. **(D)** The correlation heatmap of 36 *AlbHLHs* gene expression and relative contents of hinesol, atractylon and *β*-eudesmol

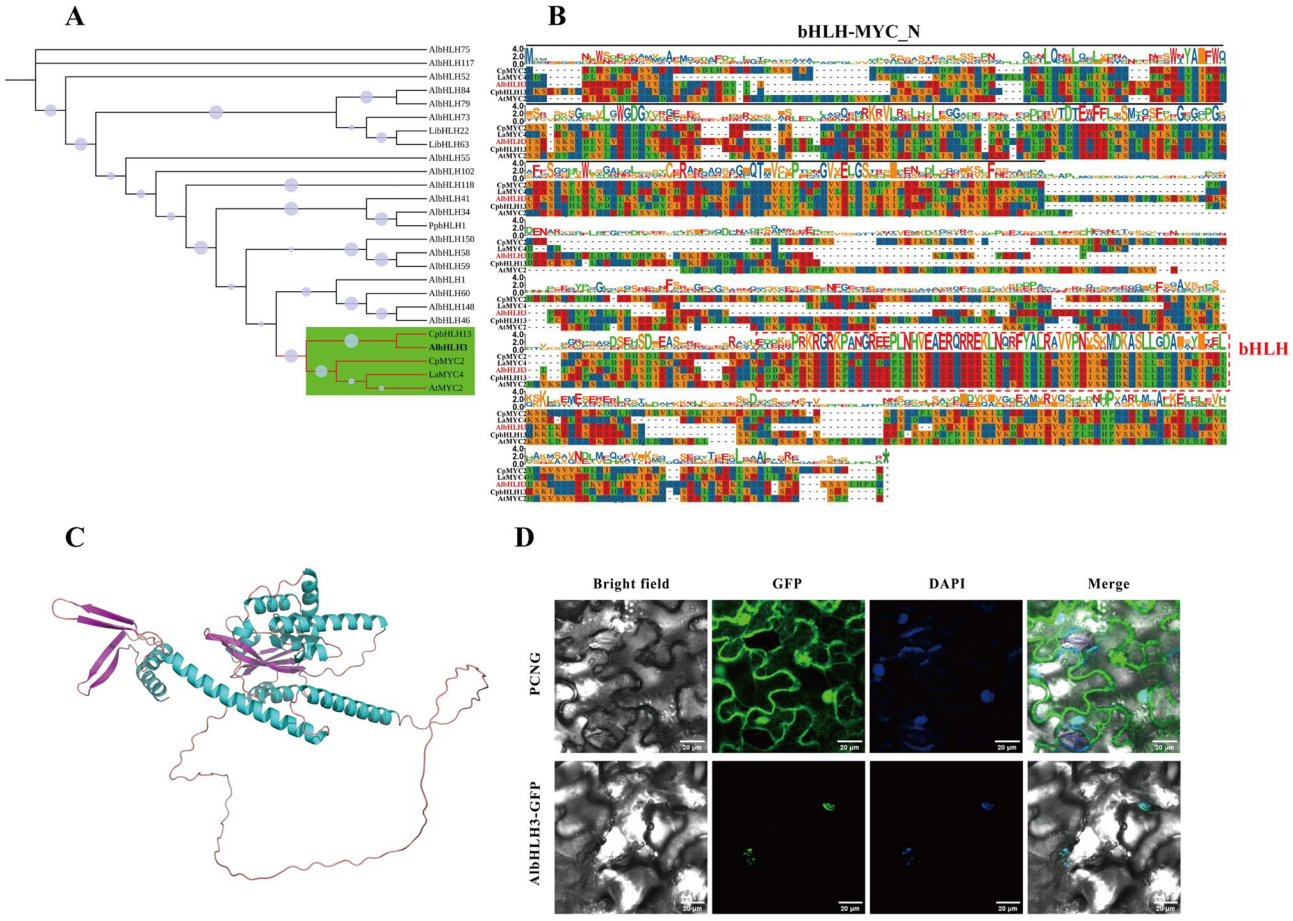

Fig. 6 Phylogenetic tree and subcellular localization of candidate AlbHLH proteins. **(A)** A neighbor-joining phylogenetic tree of candidate AlbHLHs and other bHLHs regulating terpenoid synthesis. Other plant species are as follows: Cp, *Chimonanthus praecox* L.; At, *Arabidopsis thaliana*; La, *lavender*; Pp, *Pyrus pyrifolia Nakai*; Li, *Lilium*. **(B)** The multiple sequence alignment of AlbHLH proteins and other species. The solid black line indicates the bHLH-MYC-N conservative domain, and the dotted red box indicates the bHLH conservative domain. **(C)** The protein structure of the AlbHLH3 protein. **(D)**Subcellular localization of AlbHLH3 (Scale Bars = 20 μm)

shown to significantly increase the levels of sesquiterpenoids, including caryophyllenes [25]. Additionally, AtMYC2 functions as a regulator that can activate the expression of sesquiterpene synthase genes [35]. While other AlbHLH proteins cluster with functionally characterized proteins such as PpbHLH1, LibHLH22, and LibHLH63, these are primarily associated with monoterpene biosynthesis [36, 37]. Given the phylogenetic relationship and functional context, AlbHLH3 was selected for further detailed analysis.

Structural characterization further supported AlbHLH3's functional annotation. Multiple sequence alignment confirmed the presence of conserved bHLH domains, including the characteristic bHLH-MYC_N motif, shared among AlbHLH3 and its functionally characterized orthologs (Fig. 6B). bHLH-MYC_N motif contains one JAZ (jasmonate ZIM-domain) interaction domain (JID) and one MYC transcriptional activation domain that interacts with the MED25 subunit of the mediator complex [38]. This domain is located at the N-terminus of MYC proteins and is critical for recognizing and binding to target gene promoters, thus regulating the expression of target gene [39]. Tertiary structure prediction using AlphaFold identified AF-A0A103XHY1-F1-v4 as the highest-confidence model, exhibiting strong structural homology to Myc-type bHLH domain-containing proteins (Fig. 6C). These computational predictions were consistent with the phylogenetic clustering, providing complementary evidence for AlbHLH3's classification within this functionally significant protein family.

Subcellular localization predictions using WoLF PSORT suggested that AlbHLH3 was localized to the nucleus. To experimentally validate this prediction, transient expression assays were conducted in *N.benthamiana*. As illustrated in Fig. 6D, while control GFP showed ubiquitous cytoplasmic and nuclear distribution, the AlbHLH3-GFP fusion protein exhibited exclusive nuclear localization, as demonstrated by precise co-localization with DAPI staining. This nuclear targeting pattern, characteristic of transcription factors,

supports AlbHLH3's predicted role in transcriptional regulation of sesquiterpenoid biosynthesis genes.

Transactivation of AlbHLH3 on promoters of sesquiterpenoid synthesis genes

HMGR, a key enzyme in the terpenoid biosynthesis pathway, has been reported to regulate terpenoid accumulation in *A. lancea*. Previous studies have demonstrated that bHLH transcription factors can modulate the transcriptional levels of core terpenoid biosynthetic genes, including *HMGR* [15]. Furthermore, our previous work identified four *AlHMGR* isoforms in *A. lancea*, with two novel isoforms (*AlHMGR3* and *AlHMGR4*) cloned in this study supplementing the two previously reported [40]. Tissue-specific expression analysis revealed *AlHMGR1* and *AlHMGR3* predominance in R and RB, suggesting isoform functional specialization. Since the expression level of *AlHMGR3* is higher in R and RB, it was selected as a candidate gene for further exploring (Fig. S3). Given that bHLH transcription factors typically regulate terpenoid biosynthesis by binding to E-box motifs located on the promoters of terpenoid synthesis genes, we cloned the promoter sequence of *AlHMGR3* to identify potential targets of *AlHMGR3*, which contains putative bHLH DNA-binding motifs (E-box elements) (Fig. 7A). Next, Y1H analysis and EMSA were performed to determine whether AlbHLH3 binds to the promoter of *AlHMGR3*. In the Y1H assay, the negative control group grew only on SD/-Ura. In contrast, the bait construct supported growth SD/-Ura in the absence of Aureobasidin A (AbA), but was suppressed by 100ng/mL AbA (Fig. S4). Based on this, the 100 ng/mL AbA was chosen for subsequent

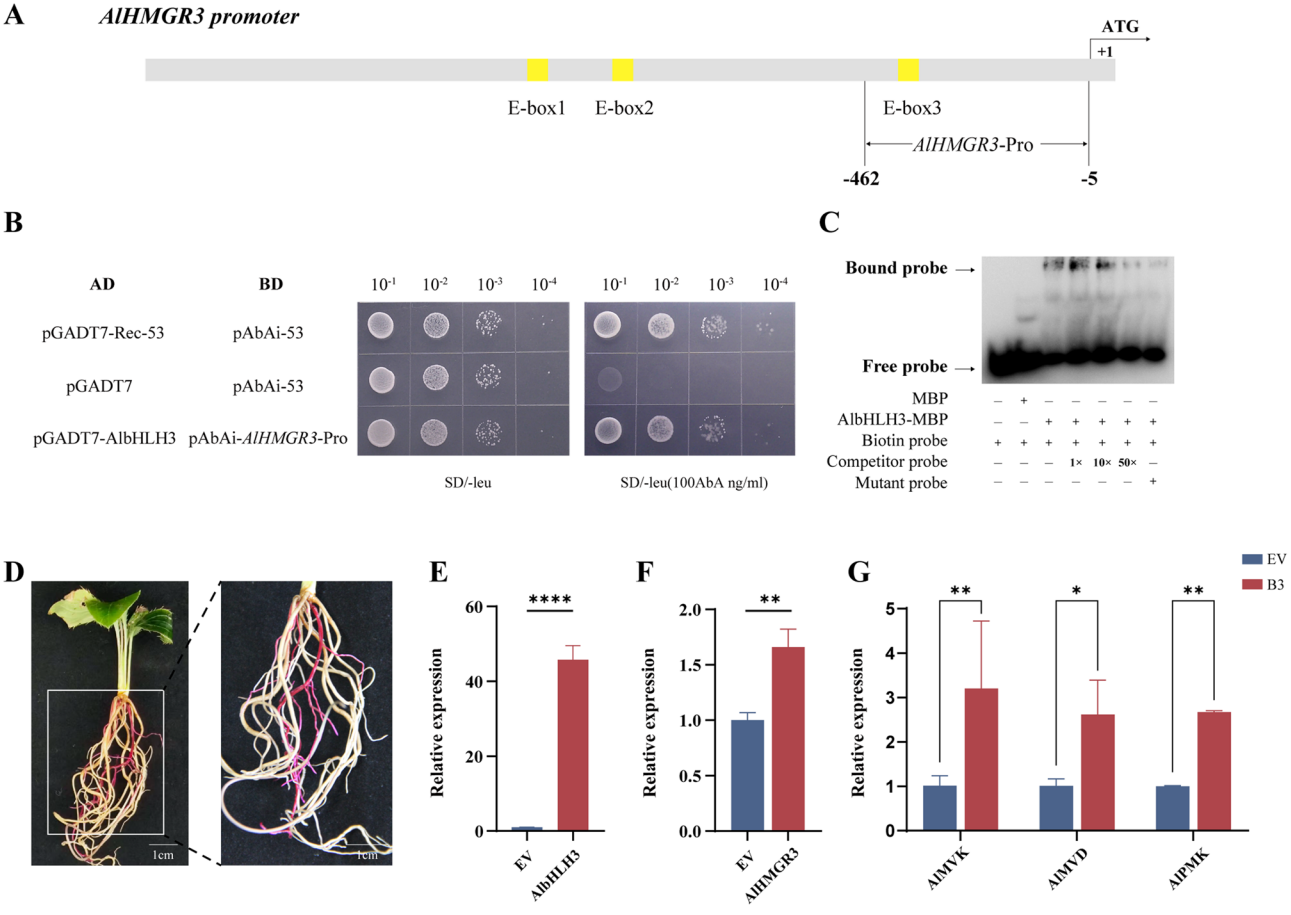

Fig. 7 Transcriptional activation assay of AlbHLH3. **(A)** Schematic representation of E-box motifs on the promoters of *AlHMGR3* genes. The ATG is shown at position +1. **(B)** Interaction of AlbHLH3 with promoters of *AlHMGR3* in yeast. Negative control transformed with empty AD vector; positive control transformed with pGADT7-Rec-53. 10^{-1},10^{-2},10^{-3},10^{-4} represented the dilution ratios of the yeast suspension. **(C)** EMSA showing the AlbHLH3 could bind to the E-box element on the promoter of *AlHMGR3*. AlbHLH3 could bind to biotin-labeled probe, and the binding strength gradually decreased with greater concentrations of unlabeled probes. **(D)** The positive roots of transgene (AlbHLH3-RUBY) in *A. lancea*. Scale Bars = 1 cm. **(E-F)** *AlbHLH3* and *AlHMGR3* expression in positive roots was examined following the overexpression of *AlbHLH3*. EV: empty vector. Error bars, mean ± standard deviation. n = 3 independent experiments were conducted. **indicate significant changes by p-value < 0.01, **** indicates significant changes by p-value < 0.0001. **(G)** Transient expression of the downstream terpenoid pathway genes. MVK, Mevalonate kinase; PMK, Phosphomevalonate kinase; MVD, Mevalonate decarboxylase. EV: empty vector, B3: AlbHLH3-RUBY. Asterisks indicate significance in two-way ANOVA vs. the negative control: *$p < 0.05$, **$p < 0.01$. Error bars, mean ± standard deviation. n = 3 independent experiments were conducted

experiments. As shown in Fig. 7B, the positive control group (pAbAi-53 + pGADT7-Rec-53) grew on both SD/-Leu and SD/-Leu with 100ng/mL AbA, whereas the negative control group (pAbAi-53 + pGADT7) only grew on SD/-Leu, did not grow on SD/-Leu with 100ng/mL AbA. This result indicated that the control groups were credible. Strikingly, only the pAbAi-*AlHMGR3*-Pro + pGADT7-AlbHLH3 transformants grew on SD/-Leu/100 ng/mL AbA media (Fig. 7B), confirming specific AlbHLH3-*AlHMGR3* promoter interaction. To further confirm that AlbHLH3 binds directly to the E-box motifs, we performed EMSA assays using the AlbHLH3-MBP fusion protein expressed in *E. coli.* SDS-PAGE analysis confirmed the successful expression and purification of the AlbHLH3-MBP fusion protein (~ 110 kDa), while the control MBP protein migrated at ~ 50 kDa (Fig. S5). For DNA-binding analysis, a distinct mobility shift was observed when the biotin-labeled probe was incubated with AlbHLH3-MBP (Fig. 7C, lane 3), while control reactions containing either free probe (lane 1) or MBP-only protein (lane 2) showed no complex formation. Competitive EMSA revealed dose-dependent binding, with increasing molar ratios (1×, 10×, 50×) of competitor probe progressively diminishing the shifted band intensity, with most competition achieved at 50× excess (lane 4–6). Additionally, AlbHLH3 exhibited weakly binding affinity to the mutated probe (lane7) compare to the wild-type probe (Fig. 7C). These results provide strong evidence that AlbHLH3 binds to E-box motifs in the native promoters of *AlHMGR3*. Additionally, we established an overexpression system in *A. lancea* in order to further explore the biological function of *AlHMGR3*. As shown in Fig. 7D and E, transgenic roots expressing the 35S: RUBY-AlbHLH3 construct exhibited strong red pigmentation and RT-qPCR analysis revealed a 45.7-fold increase in AlbHLH3 transcript levels compared to empty vector controls ($p < 0.0001$; Fig. 7E), validating the efficacy of our overexpression system. Importantly, *AlHMGR3* expression was significantly up-regulated in *AlbHLH3*-overexpressed group compared to the EV group (Fig. 7F), accompanied by coordinated induction of downstream sesquiterpenoid biosynthetic genes (*AlMVK*, *AlMVD* and *AlPMK*) (Fig. 7G). These results imply that AlbHLH3 could directly activate *AlHMGR3* by binding to the E-box motifs in the native promoters of *AlHMGR3*, which is involved in sesquiterpenoid biosynthesis, influencing the accumulation of sesquiterpenoid in *A. lancea*.

Discussion

Recent research has highlighted the pivotal role of TFs in influencing the epigenetic traits and enhancing medicinal quality of medicinal plants, suggesting their crucial involvement in the synthesis and accumulation networks of secondary metabolites [16]. Among these TFs, bHLH proteins have garnered extensive attention due to their widespread occurrence and functional diversity across various species, including *Gynostemma pentaphyllum* [41], *Ginkgo biloba* [42], and *Pinus thomsonii* [43]. However, the bHLHs involved in regulating terpenoid biosynthesis in *A. lancea* remain largely elusive. Despite the *A. lancea* genome has been sequenced [44], the lack of publicly available annotation files has hindered genome-based identification of *AlbHLH* genes. To address this gap, we constructed three generations of full-length transcriptomes to systematically identify *AlbHLH* members, ultimately uncovering 176 *AlbHLH* genes. Comparative analysis revealed that the bHLH family in *A. lancea* is substantially larger than in *Artemisia argyi* (52 genes) and *Dracaena cambodiana* (62 genes), but lower than in *A. thaliana* (162 genes), rice (167 genes), and tomato (159 genes) [21, 45, 46], underscoring the dynamic expansion and contraction of this gene family across plant lineages. Phylogenetic classification, based on *A. thaliana* as a reference, grouped the AlbHLH into 21 subfamilies. Notably, no AlbHLHs were classified into the families IVd, VIIIa, and VIIIc, suggesting potential gene loss in *A. lancea* during long-term evolution. Intriguingly, the highest abundance of AlbHLH was concentrated within subfamily XII, mirroring the observations made in *A. argyi*, a fellow member of the same plant family [46]. Members of the same subfamily are frequently performing similar functions. As previous studies have shown that bHLHs from the IVc subfamily in *A. thaliana* play an important role in Fe regulation, hinting that AlbHLHs in the same subfamily may have similar regulatory roles [47, 48]. Additionally, the bHLH proteins from III (d + e) subfamily involved in secondary metabolism, suggesting that AlbHLHs in III (d + e) subfamily might be associated with regulating the pathway of terpenoid metabolism biosynthesis [49].

The functional core of bHLH proteins lies in their conserved "helix-loop-helix" domain, which comprises approximately 50–60 amino acids, and can be divided into two parts: the basic region near the N-terminus, consisting of 10–15 amino acids, and the helix-loop-helix (HLH) region near the C-terminus, composed of 40 hydrophobic amino acids [17, 23, 50]. The conserved HLH domain serves as a pivotal structural and functional unit, mainly composed of motif 1 and motif 2, and possessed highly conserved DNA binding ability [23, 50]. They positioned in close proximity within the sequence, yet they were separated by a distinct gap. Specifically, Motif 1 is situated nearer to the N-terminus. Upon visual analysis of the bHLH domain architecture, it is evident that Motif 1 encompasses both the basic region and the first helix region, whereas Motif 2 corresponds to the part of loop and the second helix region. Except

6 AlbHLHs, all AlbHLH proteins contained the HLH domain, potentially due to the evolutionary process leading to the absence of this part in AlbHLHs.

The basic region specifically interacts with E-box (CANNTG) or G-box (CACGTG) motifs in the promoter region, while the HLH region forms homo- or heterodimers through interactions between two amphipathic α-helices, thereby cooperatively regulating the expression of downstream target genes [51, 52]. In our study, 133 AlbHLHs were classified as E-box binding proteins. This finding suggests that AlbHLHs may primarily interact with target genes through E-box motifs in their promoter regions.

Among the 176 *AlbHLH* genes identified, 36 exhibited significant expression across different tissues, with over half of these genes showing particularly high expression levels in the R or RB. Comparative transcriptome analysis revealed that DEGs were significantly enriched in pathways related to "Terpenoid backbone biosynthesis" and "Sesquiterpenoid and triterpenoid biosynthesis." This enrichment suggests that the accumulation of terpenoids in *A. lancea* is highly tissue-specific. These findings collectively indicate that *AlbHLH* genes with elevated expression in the R or RB may potentially play crucial roles in the terpenoid biosynthetic pathways. In addition, bHLH TFs have been proven to be involved in terpenoid biosynthesis in several species [16]. The content of terpenoids was enhanced by LibHLH22 and LibHLH63 [37]. Besides, *LaMYC4* overexpression increased the levels of sesquiterpenoids [25]. Given the conserved nature of gene families, the putative functions of *AlbHLH* genes may be inferred through comparison with their homologous genes. Therefore, we selected *AlbHLH3* as a candidate gene among the genes with high expression of R or RB for further study, and found AlbHLH3 was highly homologous to CpbHLH13 which was identified can regulate linalool and *β*-caryophyllene [26]. The bHLH transcription factors typically influence the accumulation of terpenoids by modulating the expression of key genes involved in the terpenoid biosynthetic pathway, such as *HMGR*, *DXR*, and *TPS*, among others [15]. For instance, PIFs (PHYTOCHROME INTERACTING FACTORs) were confirmed as negative regulators that interact with *AtHMGR2* [53]. Similarly, EkbHLH144 could interact with the *EkHMGR* to upregulate the synthesis of *β*-sitosterol [54]. The LibHLH22 and LibHLH63 which could promote the expression of *LiDXR* and *LiTPS2* to increase content of terpenoids [37]. In *Prunus persica*, PpbHLH1 could bind to the promoter of *PpTPS3* to enhance the level of linalool [36]. The GpbHLH15 and GpbHLH58 could activate the transcription of *GpFPS1*, *GpSS1* and *GpOSC1* which were the key enzyme-encoding genes in gypenoside biosynthetic pathway [41]. DobHLH4 could directly bind to the G-box element in the promoter of *DoTPS10* to regulate the biosynthesis pathway of linalool [30]. In *A. annua*, AaMYC3 could activate the transcription of *AaHD1* to initiate glandular trichome (GST) development and upregulate the expression of *CYP71AV1* and *ALDH1*, thereby promoting artemisinin production [31]. In *A. lancea*, AlHMGR serves as the rate-limiting enzyme in the mevalonate (MVA) pathway for sesquiterpenoid biosynthesis. Previous studies have demonstrated that hydrogen peroxide (H_2O_2) accumulation enhances both *AlHMGR* gene expression and *β*-eudesmol production [55]. Furthermore, under copper ion (Cu^{2+}) stress, *AlHMGR* expression positively correlates with the accumulation of three pharmacologically active sesquiterpenoids—atractylon, *β*-eudesmol, and atractylodin [56]. These results underscore the pivotal role of *AlHMGR* in sesquiterpenoid biosynthesis of *A. lancea*.

Based on the previous exploration of the *AlHMGR* gene and tisusse expression, we cloned the promoter of *AlHMGR3* to further validate practical function of AlbHLH3 using Y1H and EMSA. These results suggested that AlbHLH3 could bind directly to the E-box region of the *AlHMGR3* promoter. Functional validation revealed that *AlbHLH3* overexpression resulted in significant transcriptional activation of *AlHMGR3* and concomitant upregulation of downstream sesquiterpenoid pathway genes. These findings provide mechanistic insights into how AlbHLH3 modulates flux through the MVA pathway, ultimately influencing sesquiterpenoid accumulation in *A. lancea*. In summary, our study provides compelling evidence that AlbHLH3 may play a crucial role in regulating terpenoid synthesis in *A. lancea*. This finding not only enhances our understanding of the molecular mechanisms underlying terpenoid biosynthesis in this medicinal plant but also offers valuable insights for genetic improvement and the enhancement of medicinal quality through targeted manipulation of bHLH proteins.

Conclusion

In this study, we performed a comprehensive transcriptomic analysis of *A. lancea*, identifying 176 *AlbHLHs* and elucidating the fundamental characteristics, evolutionary relationships, and conserved motifs. This foundational resource will serve as a valuable resource for future functional studies aimed at deciphering the mechanisms underlying terpenoid biosynthesis and regulation in this medicinally significant plant. Through determination of active ingredients and gene expression profiling, we identified 19 *AlbHLHs* that were highly expressed in underground tissues, such as R and RB, suggesting their potential roles in terpenoid biosynthesis. Among them, *AlbHLH3* emerged as a particularly candidate due to its homology with known terpenoid-regulating transcription factors. Our Y1H, EMSA and overexpression

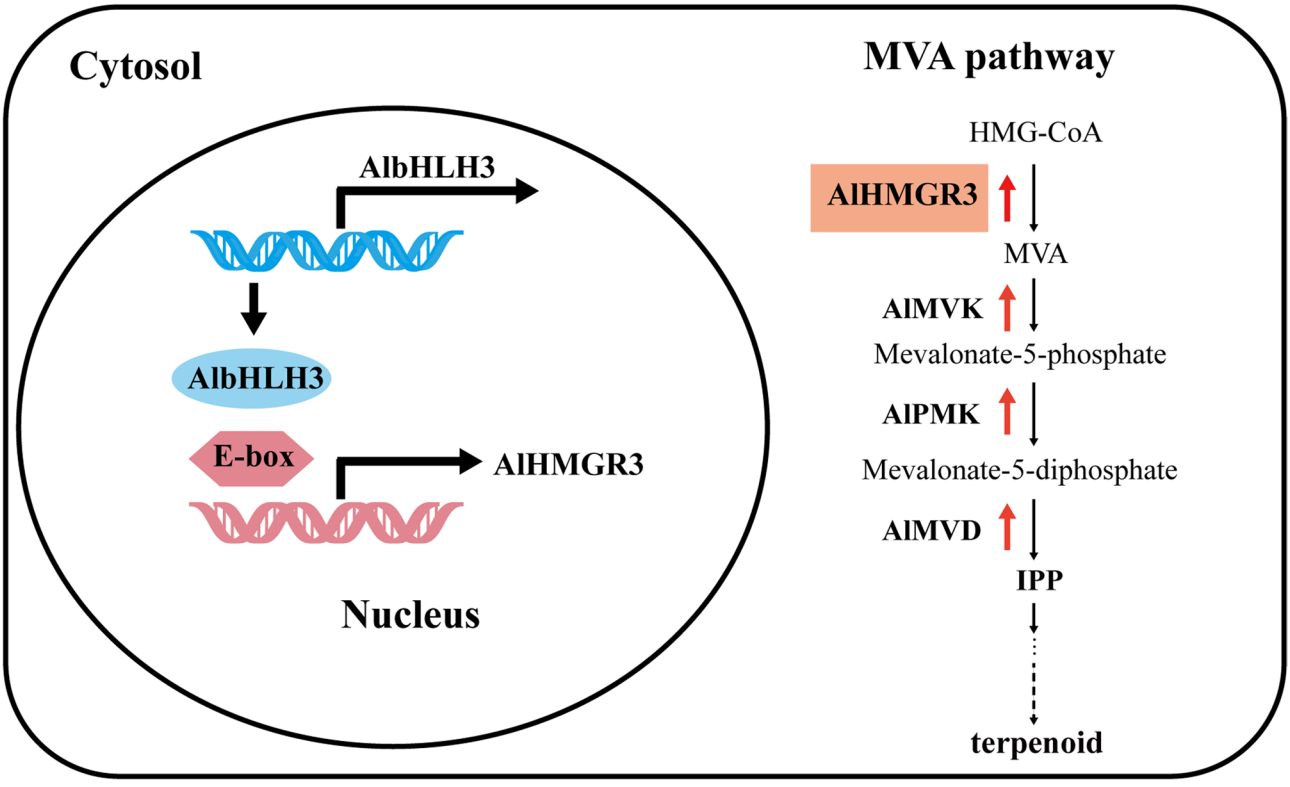

Fig. 8 A model diagram of AlbHLH3 in the regulation of terpenoid biosynthesis in *A. lancea*. AlbHLH3 directly binds to the E-box in the promoter of *AlHMGR3* to active its transcription. The black dotted arrows indicate the predicted regulatory pathway

experiments provided compelling evidence for the direct interaction between AlbHLH3 and the promoter of *AlHMGR3*, a pivotal rate-limiting enzyme in the terpenoid biosynthetic pathway. This interaction may lead to the transcriptional activation of *AlHMGR3*, thereby enhancing the expression of downstream genes involved in terpenoid biosynthesis (Fig. 8). These findings establish AlbHLH3 as a critical regulator of terpenoid biosynthesis in *A. lancea*, providing new insights into the molecular mechanisms underlying the production of these bioactive compounds.

Abbreviations

bHLH	Basic helix-loop-helix
MVA	Mevalonate
IPP	Isopentenyl diphosphate
DMAPP	Dimethylallyl diphosphate
AACT	Acetyl-CoA acetyltransferase
HMGS	3-hydroxy-3-methylglutaryl-CoA synthase
HMGR	HMG-CoA reductase
MVK	Mevalonate kinase
PMK	Phosphomevalonate kinase
MPDC	Mevalonate pyrophosphate decarboxylase
FPP	Farnesyl pyrophosphate
FPPS	Farnesyl pyrophosphate synthase
TPS	Terpenoid synthase
CYP450s	Cytochrome P450s
CCS	Circular consensus sequences
FLNC	full-length non-concatemer
FPKM	Fragments per kilobase of transcript per million mapped reads
DEG	Differentially expressed genes
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
EMSA	Electrophoretic mobility shift assay
Y1H	Yeast one-hybrid assays

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-07579-3.

Supplementary Material 1.

Acknowledgements
Not applicable.

Authors' contributions
Baiyi Zhang performed the experiments and wrote the original draft. Peina Zhou performed data analysis and wrote the original draft. Xiaoli Yang, Bingyan Gao, and Yucheng Cao contributed to project support and resource provision. Xiao Huang and Juan Deng participated in data curation and formal analysis. Kun Yu supervised the research. Ling Gong designed the experimental framework and revised the manuscript. All authors reviewed and approved the final version of the manuscript.

Funding
This work was supported by Natural Science Foundation of Hubei Province, China (2022CFB576), 71 st General Project of China Postdoctoral Science Foundation (2023AFB972), National Natural Science Foundation of China (32000254, 31670341 and 81891014, 82404814).

Data availability
The RNA-seq data has been submitted to the National Genomics Data Center (accession code: CRA025459) at https://ngdc.cncb.ac.cn/gsa.

Declarations

Ethics approval and consent to participate
The authors confirmed that the experimental research on plants including the collection of plant material comply with relevant institutional, national, and international guidelines and legislation. Plant material used in the study is not wild species and also permission for the plant material was obtained. The voucher specimen is stored in the Herbarium of hubei university of Chinese medicine, with the code 221007G, Xiuqiao Zhang identified it.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 26 April 2025 / Accepted: 14 October 2025
Published online: 12 November 2025

References

1. Wang MD, Chen PX, Yin M, Xu XX, Chen Y, Feng X, Guan FQ, Liao PH, Wang QZ. Phytochemical and chemotaxonomic study on *Atractylodes lancea*. Biochem Syst Ecol. 2023;111:104734.
2. Liu YQ, Zhang BX, Cai Q. Study on the pharmacodynamics and metabolomics of five medicinal species in *Atractylodes* DC. on rats with rheumatoid arthritis. Biomed Pharmacother. 2020;131:110554.
3. Li D, Ma XQ, Yang X, Li XY, Yao RY. Bioinformatic analysis of Ping-Wei-San Decoction for pulmonary malignancy based on UPLC-QE-Orbitrap-MS detection. Nat Prod Commun. 2024;19(5):1–9.
4. Chen LJ, Zhang SF, Wang YF, Sun HX, Wang SQ, Wang DH, Duan YZ, Niu JF, Wang ZZ. Integrative analysis of transcriptome and metabolome reveals the sesquiterpenoids and polyacetylenes biosynthesis regulation in *Atractylodes lancea* (Thunb.) DC. Int J Biol Macromol. 2023;253:127044.
5. Xu R, Wu JX, Zhang YZ, Jiang L, Yao JC, Zha LP, Xie J. Isolation, characterisation, and expression profiling of DXS and DXR genes in *Atractylodes lancea*. Genome. 2023;66(6):150–64.
6. Jun X, Fu P, Lei Y, Cheng P. Pharmacological effects of medicinal components of *Atractylodes lancea* (Thunb.) DC. Chin Med. 2018;13(1):59.
7. Sun ZJ, Zhang YT, Peng X, Huang SJ, Zhou HH, Xu J, Gu Q. Diverse sesquiterpenoids and polyacetylenes from *Atractylodes lancea* and their Anti-Osteoclastogenesis activity. J Nat Prod. 2022;85(4):866–77.
8. Cheng Y, Mai JY, Hou TL, Ping J, Chen JJ. Antiviral activities of atractylon from *Atractylodis rhizoma*. Mol Med Rep. 2016;14(4):3704–10.
9. He F, Wang W, Wu MC, Fang YP, Wang SZ, Yang Y, Ye C, Xiang F. Antioxidant and antibacterial activities of essential oil from *Atractylodes lancea* rhizomes. Ind Crop Prod. 2020;153:112552.
10. Tholl D. Biosynthesis and biological functions of terpenoids in plants. Adv Biochem Eng Biotechnol. 2015;148:63–106.
11. Yao SC, Tan XM, Huang D, Li LS, Chen JH, Ming RH, Huang RS, Yao C. Integrated transcriptomics and metabolomics analysis provides insights into aromatic volatiles formation in *Cinnamomum Cassia* bark at different harvesting times. BMC Plant Biolo. 2024;24(1):81.
12. Devi A, Masand M, Sharma B, Majeed A, Sharma RK. Integrated transcriptome and metabolome analysis decrypting molecular insights of specialized metabolism in *Valeriana Jatamansi* Jones. Ind Crop Prod. 2024;214:118504.
13. Singh S, Chhatwal H, Pandey A. Deciphering the complexity of terpenoid biosynthesis and its Multi-level regulatory mechanism in plants. J Plant Growth Regul. 2024;43(10):3320–36.
14. Wang X, Wang CY, Yang MK, Jie WC, Fazal A, Fu JY, Yin TM, Cai JF, Liu B, Lu GH, Lin HY, Han HW, Wen ZL, Qi JL, Yang YH. Genome-Wide comparison and functional characterization of HMGR gene family associated with Shikonin biosynthesis in *Lithospermum erythrorhizon*. Int J Mol Sci. 2023;24(15):12532.
15. Wei JC, Yang Y, Peng Y, Wang SY, Zhang J, Liu XB, Liu JJ, Wen BB, Li MF. Biosynthesis and the transcriptional regulation of terpenoids in tea plants (*Camellia sinensis*). Int J Mol Sci. 2023;24(8):6937.
16. Guo JR, Sun BX, He HR, Zhang YF, Tian HY, Wang BS. Current Understanding of bHLH transcription factors in plant abiotic stress tolerance. Int J Mol Sci. 2021;22(9):4921.
17. Lei P, Jiang YX, Zhao Y, Jiang MQ, Ji XM, Ma L, Jin GZ, Li JX, Zhang SB, Kong DX, Zhao XY, Meng FJ. Functions of basic Helix–Loop–Helix (bHLH) proteins in the regulation of plant responses to Cold, Drought, Salt, and iron deficiency: A comprehensive review. J Agric Food Chem. 2024;72(19):10692–709.
18. Yang LY, Miao LY, Gong Q, Guo JQ. Advances in studies on transcription factors in regulation of secondary metabolites in Chinese medicinal plants. PCTOC. 2022;151(1):1–9.
19. Li XX, Duan XP, Jiang HX, Sun YJ, Tang YP, Yuan Z, Guo JK, Liang WQ, Chen L, Yin JY, Ma H, Wang J, Zhang DZ. Genome-Wide analysis of Basic/Helix-Loop-Helix transcription factor family in *Rice* and *Arabidopsis*. Plant Physiol. 2006;141(4):1167–84.
20. Carretero-Paulet L, Galstyan A, Roig-Villanova I, Martínez-García JF, Bilbao-Castro JR, Robertson DL. Genome-Wide classification and evolutionary analysis of the bHLH family of transcription factors in *Arabidopsis*, *Poplar*, *Rice*, *Moss*, and *Algae*. Plant Physiol. 2010;153(3):1398–412.
21. Zhu JH, Xia DN, Xu J, Guo D, Li HL, Wang Y, Mei WL, Peng SQ. Identification of the bHLH gene family in *Dracaena Cambodiana* reveals candidate genes involved in flavonoid biosynthesis. Ind Crop Prod. 2020;150:112407.
22. Ali A, Khan NM, Jiang YQ, Zhou GZ, Wan YL. Comprehensive Genome-Wide identification and expression profiling of bHLH transcription factors in *Areca Catechu* under abiotic stress. Int J Mol Sci. 2024;25(23):12936.
23. Liu W, Tian XF, Feng Y, Hu J, Wang B, Chen SL, Liu D, Liu YF. Genome-wide analysis of bHLH gene family in *Coptis chinensis* provides insights into the regulatory role in benzylisoquinoline alkaloid biosynthesis. Plant Physiol Biochem. 2023;201:107846.
24. Dong YM, Zhang WY, Li JR, Wang D, Bai HT, Li H, Shi L. The transcription factor LaMYC4 from lavender regulates volatile terpenoid biosynthesis. BMC Plant Biol. 2022;22:189.
25. Dong YM, Wei ZL, Zhang WY, Li JR, Han MX, Bai HT, Li H, Shi L. LaMYC7, a positive regulator of Linalool and caryophyllene biosynthesis, confers plant resistance to *Pseudomonas syringae*. Hortic Res. 2024;11(4):uhae044.
26. Aslam MZ, Lin X, Li X, Yang N, Chen LQ. Molecular cloning and functional characterization of CpMYC2 and CpBHLH13 transcription factors from wintersweet (*Chimonanthus praecox* L). Plants. 2020;9(6):785.
27. Xiang L, He P, Shu GP, Yuan MY, Wen ML, Lan XZ, Liao ZH, Tang YL. AabHLH112, a bHLH transcription factor, positively regulates sesquiterpenes biosynthesis in *Artemisia annua*. Front Plant Sci. 2022;13:973591.
28. Zhang T, Chen XH, Xiong YP, Niu MY, Zhang YY, Yan HF, Li Y, Zhang XH, Ma GH. Identification and functional analysis of SabHLHs in *Santalum album* L. Life. 2022;12(7):1017.
29. Cao YY, Liu L, Ma KS, Wang WJ, Lv HM, Gao M, Wang XM, Zhang XC, Ren SX, Zhang N, Guo DY. The jasmonate-induced bHLH gene *SlJIG* functions in terpene biosynthesis and resistance to insects and fungus. J Integr Plant Biol. 2022;64(5):1102–15.
30. Yu ZM, Zhang GH, Teixeira da Silva JA, Zhao CH, Duan J. The Methyl jasmonate-responsive transcription factor DobHLH4 promotes *DoTPS10*, which is involved in Linalool biosynthesis in *Dendrobium officinale* during floral development. Plant Sci. 2021;309:110952.
31. Yuan M, Sheng Y, Bao J, Wu W, Nie G, Wang L, Cao J. AaMYC3 bridges the regulation of glandular trichome density and Artemisinin biosynthesis in *Artemisia annua*. Plant Biotechnol J. 2024;23(2):315–32.
32. Zheng JR, Liao YL, Ye JB, Xu F, Zhang WW, Zhou X, Wang L, He X, Cao ZY, Yi YW, Xue YS, Chen QW, Sun JX. The transcription factor MYC2 positively regulates terpene trilactone biosynthesis through activating GbGGPPS expression in *Ginkgo Biloba*. Hortic Res. 2024;11:uhae228.
33. Cao XS, Xie HT, Song ML, Lu JH, Ma P, Huang BY, Wang MG, Tian YF, Chen F, Peng J, Lang ZB, Li GF, Zhu JK. Cut–dip–budding delivery system enables genetic modifications in plants without tissue culture. Innov. 2023;4(1):100345.
34. Toledo-Ortiz G, Huq E, Quail PH. The *Arabidopsis* Basic/Helix-Loop-Helix transcription factor family. Plant Cell. 2003;15(8):1749–70.
35. Yang ZZ, Li YQ, Gao FZ, Jin W, Li SY, Kimani S, Yang S, Bao TT, Gao X, Wang L. MYB21 interacts with MYC2 to control the expression of terpene synthase genes in flowers of *Freesia hybrida* and *Arabidopsis Thaliana*. J Exp Bot. 2020;71(14):4140–58.
36. Wei CY, Liu HR, Cao XM, Zhang ML, Li X, Chen KS, Zhang B. Synthesis of flavour-related Linalool is regulated by PpbHLH1 and associated with changes in DNA methylation during Peach fruit ripening. Plant Biotechnol J. 2021;19(10):2082–96.
37. Feng Y, Guo ZY, Zhong J, Liang YL, Zhang P, Sun M. The LibHLH22 and LibHLH63 from *Lilium* 'Siberia' can positively regulate volatile terpenoid biosynthesis. Horticulturae. 2023;9(4):459.
38. Wang H, Ding CW, Du HP, Liu HL, Wang YL, Yu DY. Identification of soybean MYC2-like transcription factors and overexpression of *GmMYC1* could

stimulate defense mechanism against common cutworm in Transgenic tobacco. Biotechnol Lett. 2014;36(9):1881–92.
39. Li ZW, Huang YS, Shen ZW, Wu MF, Huang MJ, Hong SB, Xu LA, Zang YX. Advances in functional studies of plant MYC transcription factors. Theor Appl Genet. 2024;137(8):195.
40. Chen LN, Wan QY, Deng J, Ming SF, Gong L, Yu K, Xu JZ, Wu N. Cloning and analysis of two HMGR genes (AlHMGR) in *Atractylodes lancea*. Genomics Appl Biol. 2020;39:1732–40.
41. Qin YH, Li JM, Chen JH, Yao SC, Li LB, Huang RS, Tan Y, Ming RH, Huang D. Genome-wide characterization of the bHLH gene family in *Gynostemma pentaphyllum* reveals its potential role in the regulation of gypenoside biosynthesis. BMC Plant Biol. 2024;24(1):205.
42. Zhou X, Liao YL, Kim SU, Chen ZX, Nie GP, Cheng SY, Ye JB, Xu F. Genome-wide identification and characterization of bHLH family genes from *Ginkgo Biloba*. Sci Rep. 2020;10(1):13723.
43. Xiao L, Huang D, Wu ZD, Shang XH, Cao S, Zeng WD, Lu LY, Shi PL, Yan HB. Genome-wide identification of the bHLH transcription factor family and analysis of bHLH genes related to puerarin biosynthesis in *Pueraria lobata* var. *thomsonii* (Benth.). Plant Gene. 2023;33:100390.
44. Zhang CC, Wang S, Sun JH, Li XK, Wang HY, Guo XZ, Wang YF, Jiang DQ, Lyu CG, Kang CZ, Zhang Y, Xiang ZX, Yuan QJ, Yan BB, Qin M, Huang LQ, Guo LP. Genome resequencing reveals the genetic basis of population evolution, local adaptation, and rewiring of the rhizome metabolome in *Atractylodes lancea*. Hortic Res. 2024;11(8):uhae167.
45. Sun H, Fan HJ, Ling HQ. Genome-wide identification and characterization of the bHLH gene family in tomato. BMC Genomics. 2015;16:9.
46. Yi XZ, Wang XW, Wu L, Wang MY, Yang L, Liu X, Chen SL, Shi YH. Integrated analysis of basic helix loop helix transcription factor family and targeted terpenoids reveals candidate AarbHLH genes involved in terpenoid biosynthesis in *Artemisia argyi*. Front Plant Sci. 2022;12:811166.
47. Kobayashi T, Ozu A, Kobayashi S, An G, Jeon JS, Nishizawa NK. OsbHLH058 and OsbHLH059 transcription factors positively regulate iron deficiency responses in *rice*. Plant Mol Biol. 2019;101(4–5):471–86.
48. Sun KL, Wang HY, Xia ZL. The maize bHLH transcription factor bHLH105 confers manganese tolerance in Transgenic tobacco. Plant Sci. 2019;280:97–109.
49. Chu Y, Xiao SM, Su H, Liao BS, Zhang JJ, Xu J, Chen SL. Genome-wide characterization and analysis of bHLH transcription factors in *Panax ginseng*. Acta Phar Sin B. 2018;8(4):666–77.
50. Bano N, Patel P, Chakrabarty D, Bag SK. Genome-wide identification, phylogeny, and expression analysis of the bHLH gene family in tobacco (*Nicotiana tabacum*). Physiol Mol Biol Plants. 2021;27(8):1747–64.
51. Zuo ZF, Lee HY, Kang HG. Basic Helix-Loop-Helix transcription factors: regulators for plant growth development and abiotic stress responses. Int J Mol Sci. 2023;24(2):1419.
52. Huang X, Su LY, Xian BH, Yu QY, Zhang M, Fan J, Zhang CX, Liu YQ, He HZ, Zhong X, Li M, Chen SC, He YR, Li Q. Genome-wide identification and characterization of the sweet orange (*Citrus sinensis*) basic helix-loop-helix (bHLH) family reveals a role for CsbHLH085 as a regulator of citrus bacterial canker resistance. Int J Biol Macromol. 2024;267:131442.
53. Michael R, Ranjan A, Gautam S, Trivedi PK. HY5 and PIF antagonistically regulate HMGR expression and sterol biosynthesis in *Arabidopsis Thaliana*. Plant Sci. 2024;346:112168.
54. Wang M, Sun YW, Liu Y, Zhu L, Yang XJ, Xi RY, Yang YX, Cai X. Modulation of drought resistance by EkbHLH144-mediated β-sitosterol synthesis through EkHMGR. Ind Crop Prod. 2024;211:118271.
55. Tao JH, Wang DG, Pu XL, Zhao X, Jiang S. Promotion of HMGR activation and β-eudesmol biosynthesis in *Atractylodes lancea* suspension cell culture by hydrogen peroxide-mediated endopytic fungal elicitor. Chin Tradit Herb Drugs. 2013;44:5.
56. Lu QJ, Chao JG, Gu W, Zhang WM, Sang XH. Effects of copper stress on accumulation of three medicinal compositions and expression of two key enzyme genes in biosynthesis of *Atractylodes lancea*. Chin Tradit Herb Drugs. 2019;50:710–5.

Publisher's Note