Plant Biotechnology Reports
https://doi.org/10.1007/s11816-021-00674-5

Online ISSN 1863-5474
Print ISSN 1863-5466

ORIGINAL ARTICLE

Comparative transcriptome analysis to identify genes involved in terpenoid biosynthesis in *Agriophyllum squarrosum*, a folk medicinal herb native to Asian temperature deserts

Xiaoyue Yin[1,2] · Xia Yan[3] · Chaoju Qian[1] · Shanshan Zhou[1,2] · Tingzhou Fang[1,2] · Xingke Fan[1,2] · Yuan Gao[4] · Yuxiao Chang[5] · Jian Yang[6] · Xiao-Fei Ma[1,3]

Received: 14 February 2021 / Revised: 7 April 2021 / Accepted: 26 April 2021

Abstract

Agriophyllum squarrosum is a folk Mongolian medicine with pleiotropic pharmacological and ecological economic importance endemic to Asian temperature deserts. Terpenoids play critical roles in biotic and abiotic stresses due to their antioxidative activities. Based on non-targeted metabolomic analysis, we detected eight terpenoids enriched in the above-ground tissues of *A. squarrosum*, however, the molecular mechanism underlying terpenoids biosynthesis in this desert medicinal plant is rarely understood. Here, a comparative transcriptome analysis of different tissues in *A. squarrosum* was conducted to identify 84 unigenes encoding key enzymes in the upstream backbone biosynthesis and 53 unigenes encoding the downstream enzymes for terpenoid diversification. Most of the upstream genes exhibited significant high expression levels in leaf, and some of which were validated by qRT-PCR. Phylogenetic analysis showed that two downstream gene families OSCs (oxidosqualene cyclases) and TPSs (mainly in terpene synthases -*g* subfamily) had undergone notable gene expansions in *A. squarrosum* comparing with the other Amaranthaceae plant species and Arabidopsis. Nevertheless, most members from these two gene families showed the tissue-specific expression in *A. squarrosum*, which supported the diversification and tissue-specific enrichment of terpenoids across above-ground tissues. Considering to the habitat characteristics of *A. squarrosum*, we proposed that the enrichment of terpenoids and the functional diversification of terpenoids biosynthesis enzymes were more or less involved into its adaptation to stressful environments of deserts. These results expand the available genetic information underlying terpenoid biosynthesis in *A. squarrosum*, and contribute to deeper researches on pharmaceutical and eco-agricultural applications in this desert medicinal plant.

Keywords *Agriophyllum squarrosum* · Transcriptome · Terpenoid biosynthesis · Oxidosqualene cyclases · Terpene synthases · Gene expansion

✉ Xiao-Fei Ma
maxiaofei@lzb.ac.cn

[1] Key Laboratory of Stress Physiology and Ecology in Cold and Arid Regions, Gansu Province, Department of Ecology and Agriculture Research, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, Gansu, China

[2] University of Chinese Academy of Sciences, Beijing 100049, China

[3] School of Life Science, Nantong University, Nantong 226019, Jiangsu, China

[4] Faculty of Materials and Chemical Engineering, Yibin University, Yibin 644000, Sichuan, China

[5] Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 450002, China

[6] State Key Laboratory Breeding Base of Dao-Di Herbs, National Resource Center for Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijng 100700, China

Published online: 24 May 2021

Introduction

Agriophyllum squarrosum (L.) Moq. (Amaranthaceae, also called "sand rice") is an annual herbaceous plant recorded as a traditional Mongolian medicine. As a pioneer plant of the mobile and semi-mobile sand dunes in arid and semi-arid regions, *A. squarrosum* is widely distributed in the northwest and northeast of China, Mongolia, Russia and Kazakhstan of central Asia with an altitude range of 50 ~ 4000 m (Chen et al. 2014a; Qian et al. 2016; Genievskaya et al. 2017). With a long edible history dating from the Tang Dynasty (AD 618-907), its seeds have equivalent nutritions comparable to its relative species quinoa (*Chenopodium quinoa* Willd.) (Chen et al. 2014a). Its seeds and drier aboveground parts can be used as folk medicine for antipyretic, detoxification, diuretic, analgesia and plagues dispelling (Zhu 2000). Recent studies also found that the *A. squarrosum* plant extracts could significantly decrease blood glucose levels and improve glucose tolerance for resistance to diabetic in mice and work in treatment of hyperlipidemia in rats (Bao et al. 2016; Bao 2017; Saqier et al. 2019). Actually, various medicinally active compounds including flavonoids, terpenoids, coumarins, sterols, alkaloids have been identified in the aboveground tissues of *A. squarrosum*, which supported the putative basis for its ethnobotanical usage (Gong et al. 2012; Liu et al. 2013; Yin et al. 2018). On the other hand, the fresh aboveground parts of *A. squarrosum* are covered dense trichomes to protect it against abiotic and biotic stresses (Zhang et al. 2018), and emit a distinct and unique odour in the wild, which suggest glandular trichomes may produce and accumulate volatile compounds, such as volatile mono- and sesquiterpenoids. In addition, considering that its species phylogeny and nutrition compositions are close to quinoa, whose seeds layers significantly accumulate saponins (Fiallos-Jurado et al. 2016), *A. squarrosum* could also contain diverse terpenoids in different tissues. However, the molecular mechanism of terpenoids biosynthesis still remains far from clear in this folk Mongolian medicinal herb.

To date, triterpenes, such as oleanane-type triterpenoid saponin, lupeol, β-amyrin, eupatoric acid, pseudoginsenoside RT_1, diosgenin, spinasterol and stellasterol, have been isolated from the aboveground parts of *A. squarrosum* by modern phytochemical methods (Gong et al. 2012; Liu et al. 2013; Kong et al. 2018). Among the wealth of plant terpenoids, triterpenes including sterols are one of the most abundant and structurally diverse classes widely distributed in plant (Shang and Huang 2019). They are extensively implicated in pharmacological effects of many folk medicinal plants. A large of plant triterpenes also show multi-function in biological processes. For instance, simple triterpenes enhance membrane functionality and signaling, whereas complex triterpenes glycosides (saponins) contribute to plant defense to pathogen and herbivore (Thimmappa et al. 2014). Besides triterpenes, mono- and sesquiterpenes are the most commonly studied because of their roles in ecological and commercial applications (Abbas and Yu 2017), as their antimicrobial activity also benefits to clinical industries and crop protection. Plenty of pharmaceutical studies have revealed that terpenoids in the folk medicine exhibit pharmacological properties, including anti-diabetes, anti-hyperlipidemia, anti-inflammation, anti-bacteria, anti-cancer, anti-hepatoma, anti-anaphylaxis and immunomodulator (Park et al. 2010; Rios 2010; Yadav et al. 2010; Nazaruk and Borzym-Kluczyk 2015; Kishikawa et al. 2017; Xu et al. 2018). This kind of marketing demand had enhanced the research and development of terpenoids in diverse plant species, such as *Pogostemon cablin*, *Panax ginseng*, *Salvia miltiorrhiza* (Yang et al. 2015). Nevertheless, terpenoids accumulation varied in different tissues and genotypes (Chen et al. 2018; Liu et al. 2019). For example, the tissue- and habitat-specific patterns of terpenoids have been discussed in quinoa (Szakiel et al. 2011). As the level of these compounds can be affected by different stages of growth and environmental abiotic and biotic stimuli (Nazaruk and Borzym-Kluczyk 2015), we proposed that the extreme habitats stress (UV-B, drought, heat, poor soil) and outstanding aboveground biomass may guarantee *A. squarrosum* to accumulate numerous diverse terpenoids. As a representative edible and medicinal resource plant in the extreme regions, *A. squarrosum* has the potential to be developed as a renewable non-model plant for studying terpenoid production.

Generally, there are two independent pathways in plants, the primarily cytosolic mevalonic acid (MVA) pathway and the plastidial methylerythritol phosphate (MEP) pathway, which are subjected to produce isopentyl diphosphate (IPP) and dimethylallyl diphosphat (DMAPP) (Tholl 2006; Vranová et al. 2013). Diverse terpenoids are biosynthesized from precursors IPP and DMAPP. A series of branch-points enzymes, including geranyl diphosphate synthase (GPPS), farnesyl diphosphate synthase (FPPS) and geranylgeranyl pyrophosphate synthase (GGPPS), transformed these compounds into different linear C_{5n} prenyl diphosphates intermediates, which further formed carbocyclic skeleton of terpenoids(Nagegowda and Gupta 2020). Subsequently, the enzymes in cyclization, oxidation, hydroxylation, dehydrogenation, acylation or glycosylation are conductive to diversify chemical structures of terpenoids (Chen et al. 2011). Terpene synthases (TPSs) family is a major driver to transform prenyl diphosphate substrate into structurally distinct linear or cyclic terpenoids classes, including isoprene (C5), mono-(C10), sesqui-(C15), diter-(C20), tri-(C30) and tetra-(C40) terpenes (Fujihashi et al. 2018). *TPSs* gene family

contains around 30~100 members in each plant and was divieded into seven subfamilies (*TPS*-a to h). Taking the case of triterpenoids, the conversion of precursor farnesyl pyrophosphate (FPP) to 2,3-oxidosqualene is catalyzed by squalene synthase (SS) and squalene epoxidase (SE) (Misawa 2011). Then, the cyclization of skeleton 2,3-oxidosqualene catalyzed by oxidosqualene cyclases (OSCs)/triterpene synthases (tTPSs) is the first structure diversifying step to form backbone of triterpenoid (Shang and Huang 2019). OSCs gene family contains beta-amyrin synthase, lupeol synthase, cycloartenol synthase, lanosterol synthase and muti-function OSCs (Thimmappa et al. 2014). The ultimate diversification of chemical structure and biological activities is attributed to cytochrome P450 monooxygenases and UDP-glycosyltransferases, which modify the backbone of triterpenoid through oxidation and glycosylation (Sawai and Saito 2011). Genomic investigation and functional characterization had been conducted in Arabidopsis and Oryza, which suggested that TPSs family have experienced functional diversity by gene duplication and divergence in their evolutionary history (Chen et al. 2014b; Karunanithi and Zerbe 2019). However, the diversification of TPSs is very little known in non-model plants especial for Amaranthaceae.

Recent breakthroughs in sequencing promoted exploration to genome of more non-model medicinal herbs and discovery of functional genes clusters. It holds many researchers' interests in folk medicinal plant and builds the way for bioengineering of active secondary metabolites production (Lo et al. 2012). For example, RNA-seq may rapidly and powerfully offer a shortcut for the deep investigation of metabolic pathways and genes combined with non-targeted metabolome approaches in no genome reference folk herbs. For some folk medicinal plant, comparative transcriptome analysis have been used to identify candidate genes associated with targeting metabolites biosynthesis through investigating different tissues or cultivars or species (Amini et al. 2019; Testone et al. 2019; Tai et al. 2020). Despite transcriptome analysis linked to ecological traits were occasionally documented in *A. squarrosum* (Zhao et al. 2014, 2017), comparative transcriptome researches on different tissues especially related to secondary metabolites characterization has not been conducted.

In this study, based on non-targeted metabolomic profiling, eight terpenoids were found to be enriched in the aboveground tissues of *A. squarrosum.* To further elucidate the potential pathways underlying terpenoid biosynthesis, we then employed comparative transcriptome across different tissues in *A. squarrosum* to identify putative functional genes and profile their expression patterns. With the other representative species from Amaranthaceae and Arabidopsis, we also constructed phylogenetic relationships of key gene families involved in terpenoids biosynthesis and clarified their functional diversification. This effort will provide genetic information to illuminate further molecular mechanism of terpenoid biosynthesis in *A. squarrosum* and other Amaranthaceae species, which will promote pharmacological applications and serve a foundation for local evolution and ecology researches of this folk medicinal herb.

Materials and methods

Plant materials

All fresh plant samples were collected from wild habitats. The samples were harvested from Ertala, Gonghe county, Qinghai province (36° 11′ 39.48″ N, 100° 31′ 28.26″ E, 2917 m) on the Qinghai-Tibet Plateau. Leaves, stems and spikes were picked from the three different plants in the growth period and immediately chilled in liquid nitrogen and stored at -80 °C ultra-cold storage freezer. *A. squarrosum* is widely distributed in the arid and semi-arid desert regions of China and is not an endangered or protected species. We collected fresh samples for this study with no need for specific permits. Voucher specimens were deposited in Key Laboratory of Stress Physiology and Ecology in Cold and Arid Regions, Gansu Province, Department of Ecology and Agriculture Research, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences.

Non-targeted metabolomic analysis

The freeze-dried leaves, stems and spikes were crushed in using a mix mill (MM400, Ratsch) for 1.5 min at 30 Hz. The dried 100 mg power were weighted and extracted in 1.0 ml 70% aqueous methanol for overnight at 4 °C. Following centrifugation at 10,000 × g for 10 min, the extracts were absorbed (CNWBOND Carbon-GCB SPE Cartridge, 250 mg, 3 ml; ANPEL, Shanghai, China) and filtered (SCAA-104, 0.22 μm pore size; ANPEL). The samples for preparation were detected by LC-ESI-Q TRAP-MS/MS systems at the Metware company (Wuhan, China). The differentially enriched metabolites among different tissues were estimated by VIP (variable importance in the projection) value ≥ 1 and fold change ≥ 2 or ≤ 0.5.

RNA extraction and RNA-sequencing

Each sample was weighed approximate 100 mg to extract total RNA following as the operator's manual of RNAprep Pure Plant Kit (Tiangen Biotech Co., LTD, Bejing, China). The quality and content of RNA samples were evaluated by 1% agarose gel electrophoresis, NanoDrop 2000^{TM}micro-volume spectrophotometer (Thermo Scientific, Waltham, MA, USA) and Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Total RNA was

isolated from three biological replicates and equal amount of them was mixed together for RNA-seq. mRNA library prepare-action was carried out at an Illumina Hiseq^{TM}3000 platform of double end sequencing model of homogenization treatment of high throughput cDNA sequencing. The raw sequenced data have been submitted to the NCBI database with the accession number PRJNA659807.

De-novo assembly and Annotation

To acquire the clean reads, Trim Galore software dynamically removed adaptor sequence fragment from the 3' end of sequencing data, empty reads and low-quality fragments, and FastQC software was used to analyze the quality control of the pre-processed data. The Trinity assembled the clean reads and constructed the transcripts by connecting the contigs according to the pair-end information of the sequences (Grabherr et al. 2011). To maximize the integrity of homologs and minimize the redundancy of isoforms, only the longest transcript from one unigene was included into annotation. To identify genes related to terpenoid biosynthesis, we annotated the functions of all unigenes using BLASTx (E-value < 10^{-5}). Subsequently, all unigenes sequences alignment to several kinds of public proteinand/or nucleotide databases, including the non-redundant protein (Nr) database, the nucleotide sequence database (Nt), the SwissProt database, the EuKaryotic Orthologous Groups (KOG) database and KEGG pathway database. Blasted unigenes hits with E-value lower than 10^{-5} were kept into the further analysis. Blast2 GO program allowed us to obtain GO annotation (Götz et al. 2008). KOG/COG databases predicted the function classification of annotated sequences to further characterize the protein functions. KEGG pathway database allowed the analysis of functional classification and metabolic pathway mapping for each annotated unigenes.

Differentially expressed genes identification and analysis

The relative expression level of unigenes was normalized by the FPKM (Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced) value. Differentially expressed genes (DEGs) analysis in different tissues pairwise comparisons was assessed by DESeq2 R package (version 1.10.1) utilizing a negative binomial distribution model (Anders and Huber 2010). The FDR value less than 0.05 and |fold change| over 2 were regarded to identify significantly different expression of the unigenes (Love et al. 2014). Visualization of the clusters of the DEGs expression pattern was performed using Tbtools. Subsequently, enrichment KEGG pathways analysis of DEGs was obtained by KOBAS software (version 2.0.12). To get a better understanding of the DEGs, they were further searched for significantly enriched KEGG terms with the hypergeometric Fisher exact test, in which corrected p value < 0.05 was considered as significant enrichment. Transcription factors (TFs) were annotated by Hmmsearch and classified into TF families with characteristics by PlnTFDB database.

Quantitative real-time PCR (qRT-PCR) validation

Total RNA of different tissues was isolated by the same method mentioned above. The first-strand cDNA from total RNA was synthesized by FastKing RT Kit with gDNase (Tiangen Biotech Co., LTD, Bejing, China). The cDNA was diluted to 50 ng/μL used as template for qRT-PCR. The qRT-PCR was performed in Agilent Technologies Stratagene M × 3000P by SuperReal Premix Plus (SYBR Green) Kit (Tiangen Biotech Co., LTD, Bejing, China) following manufacturer's protocol. The Actin gene was used as an internal reference to normalize the data. The selected gene primers are listed in Supplementary Table S1. Each gene was conducted using two biological replicates and three technical replicates. The relative gene expression fold was calculated using the $2^{-\Delta\Delta Ct}$ method.

Co-expression analysis

To explore the expression state of interacted genes, we analyzed the correlation coefficient between a pair of DEGs related to terpenoid biosynthesis and used correlation coefficient to estimate the relationship of genes expression. We hypothesized that candidate TFs may exhibit expression patterns congruent with that of the key enzyme genes of the upstream terpenoids backbone biosynthesis. The thresholds (> 0.98 or < -0.98) were determined to screen positive and negative relationship on the co-expression network analysis. The network was constructed by Cytoscape v.3.7.2 and the edge and node represented the regulation possibility between genes. Three topological features of the network (Degree, Betweenness and Closeness) and the k-core algorithm were used to screen candidate genes. We also used CytoHubba add-in to find hub genes by different algorithms evaluation. These values can evaluate the protein's topological importance in the network. The k-core value was used for confirming the stability of subnetwork and complexity of gene relationship (Huber et al. 2007).

Construction of phylogenetic tree of OSCs and TPSs

Based on the results of the functional annotation, the candidate *A. squarrosum* sequences of OSCs and TPSs were obtained. Together with several sequenced Amaranthaceae species (*Chenopodium quinoa*, *Beta vulgaris*, *Spinacia oleracea*) and Arabidopsis (*Arabidopsis thaliana*), we screened

the full length CDS sequences of OSCs and TPSs genes in *A. squarrosum* from genome sequences (unpublished data). The multiple amino acids sequences were aligned with MAFFT 7.037 and viewed in Jalview 2.10.5. We analyzed their conserved motifs by MEME (http://meme-suite.org/tools/meme) with the default parameters and visualized in TBtools. MEGA 6.06 constructed phylogenetic trees of OSCs and TPSs genes. All positions containing gaps and missing data were eliminated. The phylogeny was inferred using the Neighbor-Joining method and the bootstrap values were performed with 1000 replicates to analyze evolutionary relationship of the taxa. Trees were visualized by Evolview v2.

Results

Differential enriched terpenoids in the above-ground tissues of *A. squarrosum*

Based on UPLC-MS non-targeted metabolomic profiling, the active terpenoid compounds included 2 monoterpenoids, 3 diterpenoids, 2 steroids and 1 carotenoid (Fig. 1). Among them, diosgenin was the most abundant compound. Sabinene (monoterpenoid) and 18-Nor-4,15-dihydroxyabieta-8,11,13-trien-7-one (diterpenoid) were significantly enriched in stems, while polypodine B (steroid) was enriched in spikes. Apo-13-zeaxanthinone was one kind of carotenoids, which usually accumulated in leaves.

Illumina sequencing and the de-nova assembly of the *A. squarrosum* transcriptome

As lack of reference genome sequences, we randomly selected 6 Gbps raw reads from each biological repeat samples and *de-novo* assembled the final transcriptome. A total of 897,386,640 raw reads were obtained and 884,051,398 clean reads with a total of 128.96 Gb nucleotides constituted the final data file after trimming the adapter and low-quality sequences. In short, the *de-novo* transcriptome contained 135,686 unigenes representing 165,720 transcripts with an N50 contig size of 933 nucleotides, with a mean length of 690.38 bp and GC contents of 39.73% (Supplementary Table S2). The total accumulated size of the assembled transcripts was approximately 13.76 Mb. Of all these unigenes, 81,733 unigenes were 200–500 bp, 30,593 unigenes were 500–1000 bp, 15,527 unigenes were 1000–2000 bp and 7833 unigenes were more than 2 kb. Due to alternative splicing

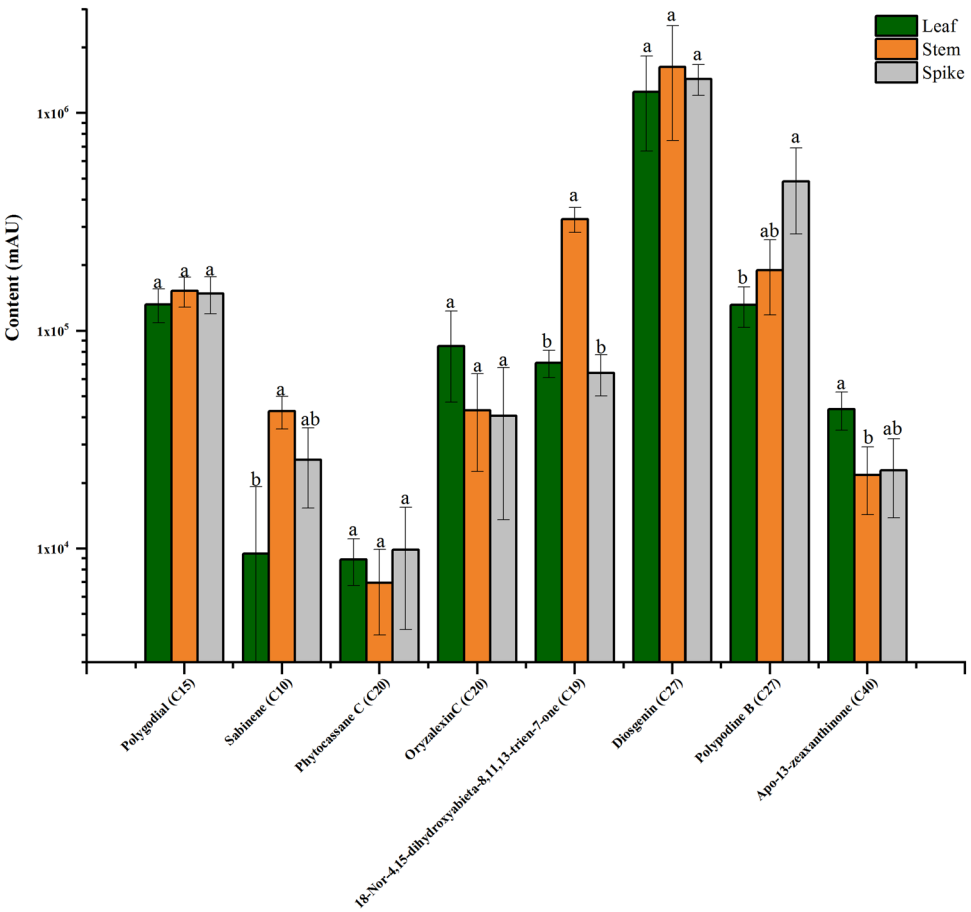


Fig. 1 The relative content of several terpenoids across different above-ground tissues in *A. squarrosum*

form, alleles, close paralogs and orthologs, some unigenes included several transcripts, and the longest transcripts of unigenes were chosen and utilized to analyze the functions and expression for these unigenes.

Functional classification

All unigenes were blasted and annotated against the public databases, including Nr, SwissProt, KOG, KEGG, GO, Interpro databases. As a result, total number of unigenes annotated by any databases was 69,977 accounting for 51.57% of all unigenes with protein sequences or domains, while only 3.71% of the unigenes could be available from all databases for strict annotation (Table 1). The Nr database (61,751 unigenes, 45.51%) had the highest annotation rate and the KEGG database (13,990 unigenes, 10.31%) had the lowest annotation. Overall, 11.29% (15,319) of the unigenes was characterized into 25 functional KOG groups. Among them, unigenes classified into "general function prediction only" occupied the highest part (4198 unigenes, 27.40% of KOG hits), followed by the group "signal transduction mechanisms" (2706 unigenes) and "posttranslational modification, protein turn over, chaperones" (2351 unigenes) (Fig. 2a). The group "secondary metabolites biosynthesis, transport and catabolism" (822 unigenes) may contribute to screen putative candidate genes in this study. Derived from sequence homology, a total of 38,680 (28.51%) unigenes were annotated with 62 GO terms, predicting putative function in 3 domains, 25 terms in biological process, 19 terms in cellular component and 18 terms in molecular functions. Among them, the most representative and abundant categories were "metabolic process" (30,952 unigenes) and "cellular process" (30,640 unigenes) in the biological process term. In the cellular component term, genes involved in "cell, cell part and organelle" were more than that in other classifications. Under the molecular functions, the top three categories were "binding", "catalytic activity" and "protein binding transcription factor activity" (Fig. 2b).

Table 1 Statistics of annotation for unigenes in different databases

Annotated databases	Number of unigenes	Percentage (%)
SwissProt	40,157	29.6
Nt	45,505	33.54
Nr	61,751	45.51
KOG	15,319	11.29
KEGG	13,990	10.31
Interpro	23,496	17.32
GO	38,680	28.51
Any databases	69,977	51.57
All databases	5040	3.71

In the light of KEGG annotations, unigenes putatively involved in biological pathways were predicted in *A. squarrosum*. A total of 1178 unigenes were annotated and mapped to KEGG secondary metabolism pathways system (Table 2). Unigenes associated with "phenylpropanoid biosynthesis" were the most prevailing. In addition, 457 unigenes were categorized as "Metabolism of terpenoids and polyketides", including "terpenoid backbone biosynthesis" (151 unigenes), "monoterpenoid biosynthesis" (38 unigenes), "diterpenoid biosynthesis" (30 unigenes), "sesquiterpenoid and triterpenoid biosynthesis" (52 unigenes), "carotenoid biosynthesis" (63 unigenes), "limonene and pinene degradation" (37 unigenes), "brassinosteroid biosynthesis" (14 unigenes), "zeatin biosynthesis" (29 unigenes).

Identification of DEGs and KEGG enrichment analysis

To assess differential expression level of genes in aboveground tissues of *A. squarrosum*, the number of DEGs was compared in two pairwise combinations, including leaf and stem, leaf and spike, stem and spike, respectively (Fig. 3a). The results showed the number of DEGs between leaf and spike pairs (6692 up-regulated and 10,791 down-regulated genes) was accounted for the highest part. Following by the annotation, DEGs among different tissues comparison group were mapped to KEGG pathway and a large portion of them was related to biosynthesis of secondary metabolites (Supplementary Table S3). Especially, the DEGs identified in the leaf vs. stem comparison were significantly enriched in diterpenoid biosynthesis, brassinosteroid biosynthesis, phenylpropanoid biosynthesis and flavonoid biosynthesis.

Identifying terpenoid biosynthesis genes and their expression patterns in *A. squarrosum*

Generally, terpenoid biosynthesis contains an array of core enzymes that constructed the elementary structure for isoprenoids via MVA and MEP pathways and then transformed to terpenoid intermediates via branch points (Fig. 3b). Based on KEGG pathway annotations combined with annotations from other public annotation databases, we found 84 annotated unigenes related to the upstream terpenoid backbone biosynthesis (Supplementary Table S4). In detail, 40 unigenes encoded the 6 key enzymes in MVA pathway; 19 unigenes encoded 7 key enzymes in the MEP pathway; and 25 unigenes encoded 4 key enzymes responsible to convert IPP to terpenoid intermediates in *A. squarrosum* (Table 3). In the MVA pathway, 12 unigenes encoded acetyl-CoA C-acetyltransferase (AACT), 14 unigenes encoded hydroxymethylglutaryl-CoA synthase (HMGS) and 8 unigenes encoded hydroxymethylglutaryl-CoA reductase (HMGR) were identified. In the MEP pathway, relative little number of unigenes

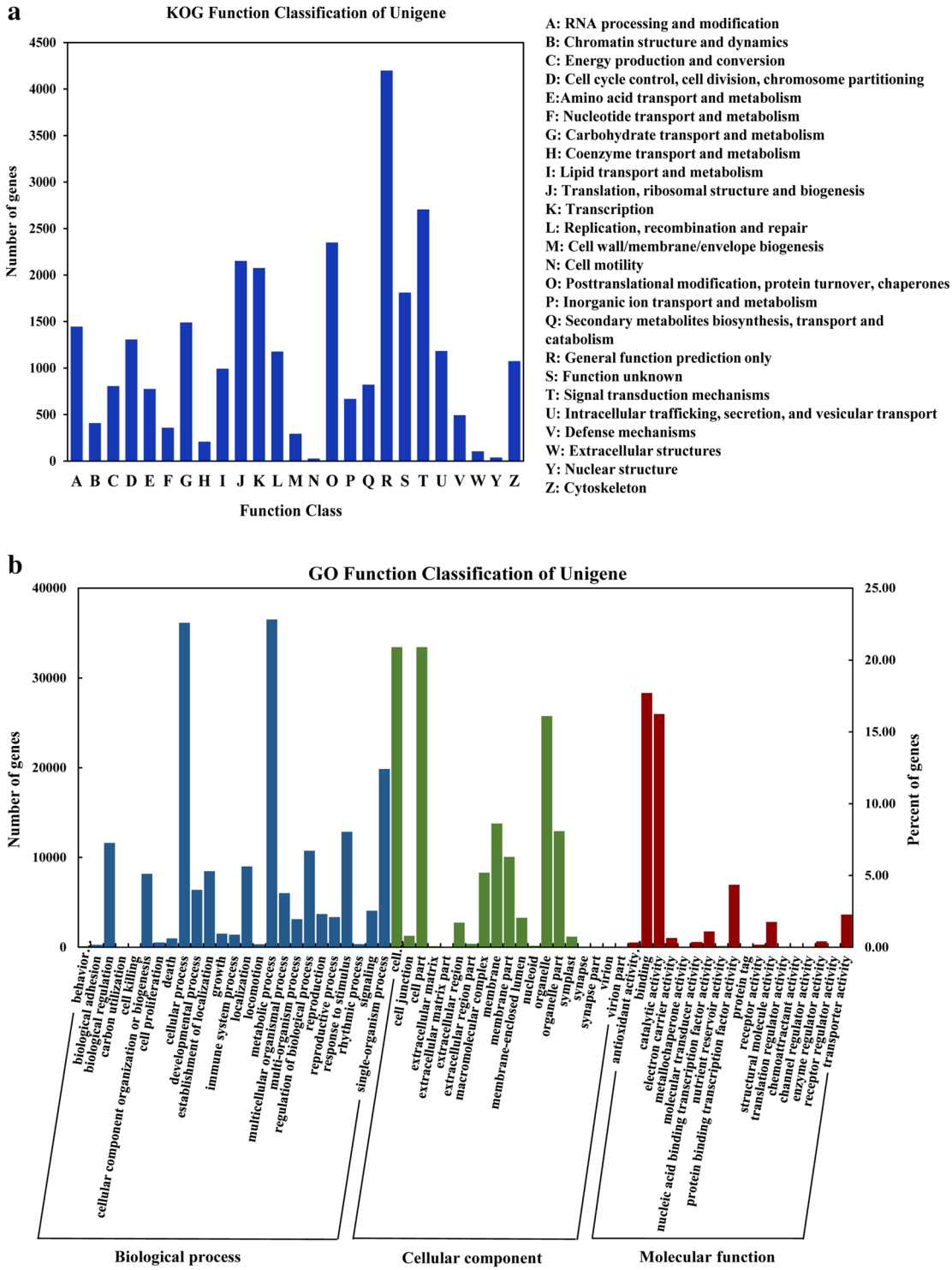


Fig. 2 Gene annotations of unigenes of *A. squarrosum.* **a** Histogram presentation of EuKaryotic orthologous groups (KOG) classification. **b** GO function classification of unigenes. The results were classed into three categories: biological process, cellular component, molecular function

was identified to encode key enzymes, like only one unigene encoded 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (MCT), 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK), respectively. In the branch point, DMAPP and IPP were transformed into farnesyl diphosphate (FPP), geranyl diphosphate (GPP), and geranylgeranyl diphosphate (GGPP) by the enzymes FPPS, GGPS and GGPPS. We found six unigenes encoded FPPS, seven

Table 2 Pathways and the number of unigenes involved in secondary metabolites biosynthesis in *A. squarrosum*

Secondary metabolites biosynthesis pathway	Gene number	Pathway ID
Anthocyanin biosynthesis	1	ko00942
Betalain biosynthesis	6	ko00965
Brassinosteroid biosynthesis	14	ko00905
Biosynthesis of ansamycins	26	ko01051
Caffeine metabolism	9	ko00232
Carotenoid biosynthesis	63	ko00906
Diterpenoid biosynthesis	30	ko00904
Flavone and flavonol biosynthesis	1	ko00944
Flavonoid biosynthesis	63	ko00941
Geraniol degradation	4	ko00281
Glucosinolate biosynthesis	12	ko00966
Indole alkaloid biosynthesis	6	ko00901
Isoflavonoid biosynthesis	3	ko00943
Isoquinoline alkaloid biosynthesis	73	ko00950
Limonene and pinene degradation	37	ko00903
Monoterpenoid biosynthesis	38	ko00902
Novobiocin biosynthesis	14	ko00401
Phenylpropanoid biosynthesis	366	ko00940
Polyketide sugar unit biosynthesis	1	ko00503
Sesquiterpenoid and triterpenoid biosynthesis	52	ko00909
Stilbenoid, diarylheptanoid and gingerol biosynthesis	49	ko00945
Streptomycin biosynthesis	42	ko00521
Terpenoid backbone biosynthesis	151	ko00900
Tetracycline biosynthesis	10	ko00253
Tropane, piperidine and pyridine alkaloid biosynthesis	78	ko00960
Zeatin biosynthesis	29	ko00908
Total	1178	

unigenes encoded GPPS and two unigenes encoded GGPPS may be responsible to transform substrate into a variety of terpenoids. Besides, we also identified 53 unigenes encoding the upstream enzymes, including 6 unigenes encoding SS, 6 for SE, and 8 for OSCs (3 for BAS, 4 for LUS and 1 for CAS) for triterpenoid biosynthesis, 33 unigenes encoding TPS enzymes acted on terpenoid diversification in different biosynthetic pathways.

In total, 23 DEGs were found in the upstream terpenoids backbone biosynthesis (Table 4, Supplementary Table S5). Most of DEGs were recognized between leaf and spike and were significantly up-regulated in the leaf (Fig. 3c–e). In the MVA pathway, most of *AACTs, HMGSs* and four *HMGR*s showed higher expression in leaf, while one *MK* (mevalonate kinase) and two *MVDs* (diphosphomevalonate decarboxylase) showed higher expression in spike (Fig. 3c). However, most of unigenes in the MEP pathway highly expressed in leaf, while three *DXS*s (1-deoxy-D-xylulose-5-phosphate synthase) possessed high expression level in stem (Fig. 3d). At branch point, *IPPI9*, *IPPI10*, *FPPS3*, *FPPS5*, *FPPS6*, *GPPS2*, *GPPS5*, *GPPS7*, *GGPPS1* and *GGPPS2* significantly highly-expressed in leaf compared with other two tissues (Fig. 3e). To confirm the RNA-seq results, eight putative terpenoid backbone biosynthesis genes were selected for qRT-PCR validation. Except for *HMGS1 and HMGR1*, the qRT-PCR expression pattern of other genes were consistent with the RNA-seq result, which all showed high expression level in leaf (Fig. 4). The tissue-specific expression levels of these genes may play a vital regulatory role in the downstream terpenoid products enrichment.

Putative transcription factor families involved in terpenoid biosynthesis

Here, a total of 1177 unigenes encoding TFs were annotated and classified into 53 families in *A. squarrosum*. Among these TFs, bHLH (Basic helix–loop–helix) family contained 109 memebers (9.26%) and was the most abundant transcription factor family, followed by ERF (Ethylene-Responsive-Factor, 83, 7.05%), bZIP (Basic Leucine Zipper, 67, 5.69%), MYB (65, 5.52%), WRKY (59, 5.01%) and NAC (58, 4.93%) families (Fig. 5a).

Based on the expression data of DEGs in four tissues, differently expressed TFs were screened and the co-expression

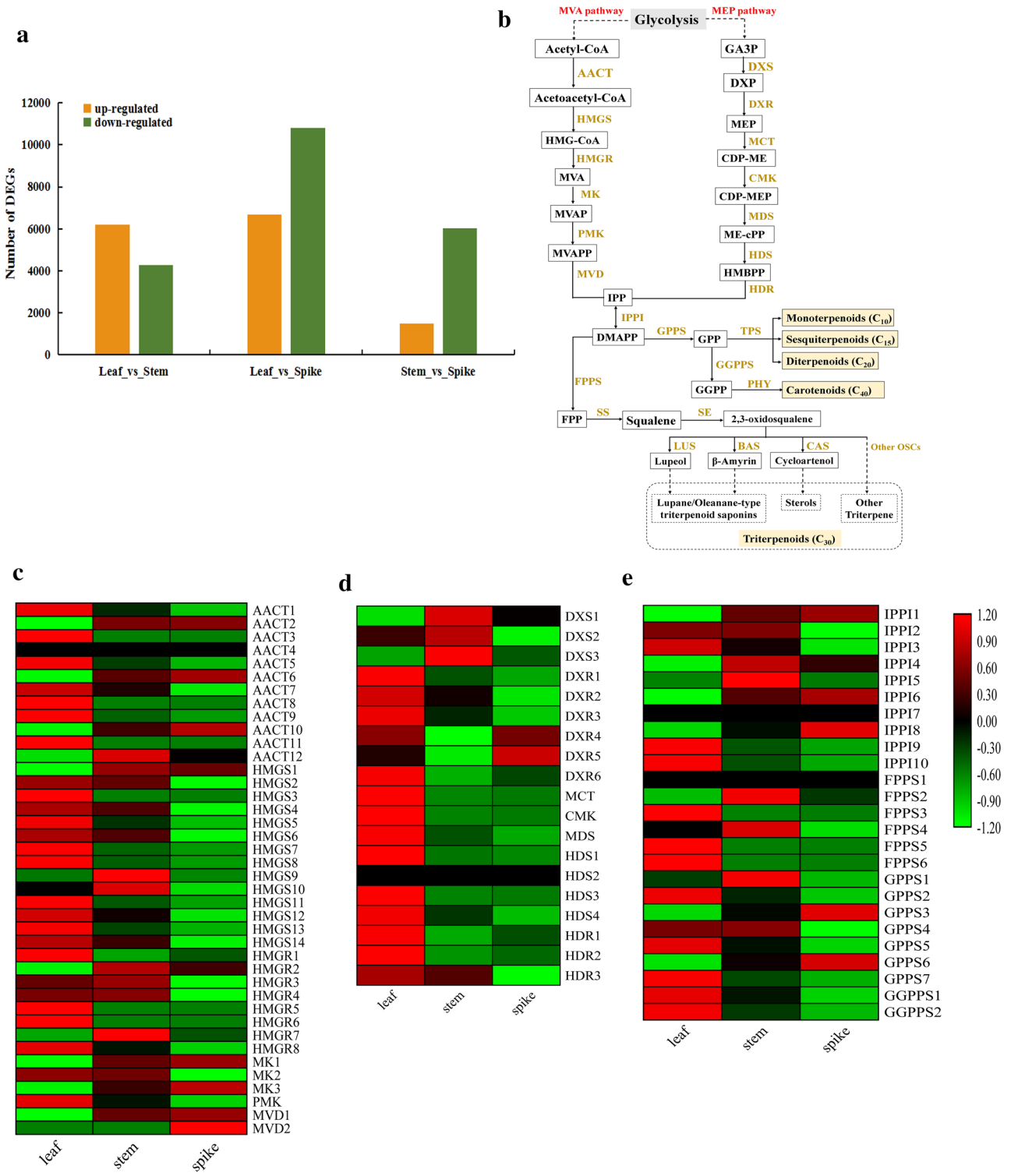


Fig. 3 Expression analysis of DEGs related to terpenoid biosynthesis pathways of *A. squarrosum*. **a** The number of different expressed genes among different tissues. **b** Proposed pathways for terpenoid biosynthesis in *A. squarrosum*. **c** The expression of genes involved in the MVA pathway. **d** The expression of genes involved in the MEP pathway. **e** The expression of genes involved in branch points of the various terpenoid biosynthesis pathway

Table 3 The unigenes encoding core enzymes involved in terpenoid biosynthesis in *A. squarrosum*

Pathway	Gene	Enzyme No	KO	No. of Unigenes
MVA	acetyl-CoA C-acetyltransferase (AACT)	EC:2.3.1.9	K00626	12
	hydroxymethylglutaryl-CoA synthase (HMGS)	EC:2.3.3.10	K01641	14
	hydroxymethylglutaryl-CoA reductase (HMGR)	EC:1.1.1.34	K00021	8
	mevalonate kinase (MK)	EC:2.7.1.36	K00869	3
	phosphomevalonate kinase (PMK)	EC:2.7.4.2	K00938	1
	Diphosphomevalonate decarboxylase (MVD)	EC:4.1.1.33	K01597	2
MEP	1-deoxy-D-xylulose-5-phosphate synthase (DXS)	EC:2.2.1.7	K01662	3
	1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR)	EC:1.1.1.267	K00099	6
	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (MCT)	EC:2.7.7.60	K00991	1
	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK)	EC:2.7.1.148	K00919	1
	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MDS)	EC:4.6.1.12	K01770	1
	4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (HDS)	EC:1.17.7.1	K03526	4
	4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase (HDR)	EC:1.17.7.4	K03527	3
Branch points	Isopentenyl-diphosphate delta-isomerase (IPPI)	EC:5.3.3.2	K01823	10
	Geranyl diphosphate synthase (GPPS)	EC:2.5.1.1	K00804	7
	Farnesyl diphosphate synthase (FPPS)	EC:2.5.1.10	K00795	6
	Geranylgeranyl pyrophosphate synthase (GGPPS)	EC:2.5.1.29	K13787	2

Table 4 The DEGs on the number of differential genes between different tissues of MVA and MEP terpenoid biosynthesis pathways

Comparison pairs	Number of DEGs	Up-regulated genes	Down-regulated genes
Leaf_vs_Stem	14	12	2
Leaf_vs_Spike	23	19	4
Stem_vs_Spike	7	7	0

network was further constructed to infer the correlations to the upstream key enzyme genes (Fig. 5b, Supplementary Table S6, S7). Among the expression network, we found the upstream genes *HMGS8*, *HMGS11* and *GGPPS2* habored the highest values of k-coreness and degree centrality (Supplementary Table S8), which suggested their central localization in terpenoid biosynthesis network. According to different algorithms screening hub genes, *HMGR5*, *HDS1* and *FPPS3* might also be important to construct k-core sub-networks of co-expression

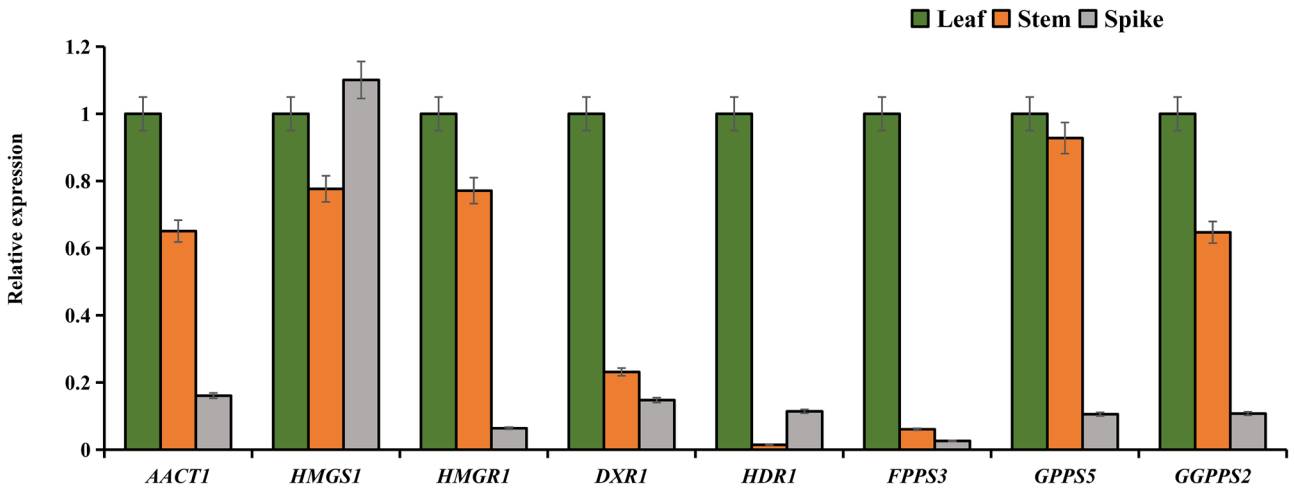


Fig. 4 The expression analysis of putative terpenoid backbone biosynthesis genes by qRT-PCR in three above-ground tissues in *A. squarrosum*. The relative expression level of each gene was calculated by Delta CT method. In each gene, the expression level of leaf samples was assigned a value of 1 and the expression levels of other tissues were transformed

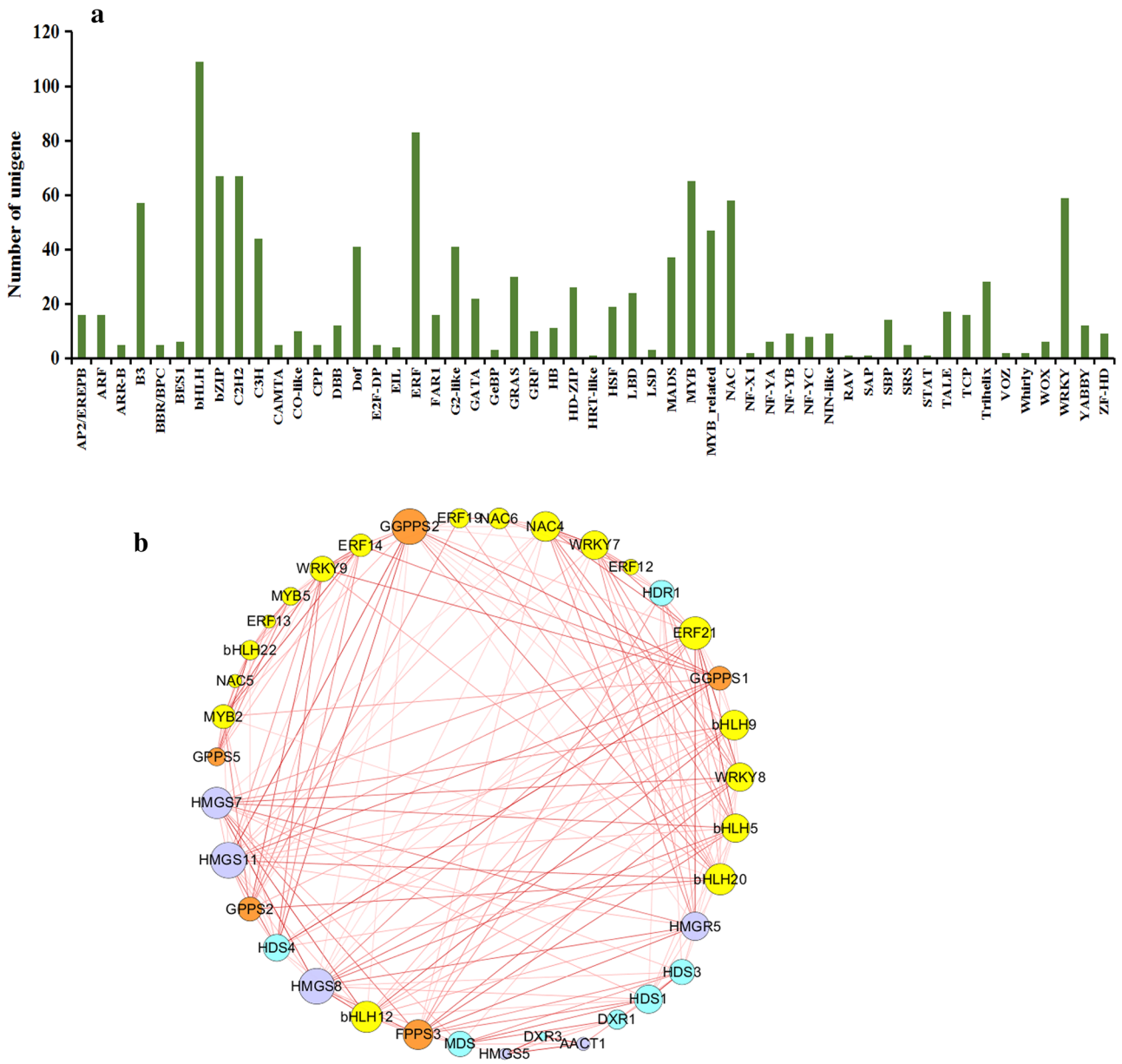


Fig. 5 Co-expression network. **a** Putative transcription factors families in *A. squarrosum*. **b** Genes co-expression network. Nodes represent genes and edges represent a significant expression level correlation between a pair of nodes. The different shades of color of edges represent the degree of correlation. The size of node circle represented degree value of the network. The lilac nodes, genes involved in MVA pathway; the lake blue nodes, genes involved in MEP pathway; the orange nodes, genes involved in the key branch points of terpenoid biosynthesis pathway; the yellow nodes, putative TFs involved in terpenoid biosynthesis pathway of *A. squarrosum*

network. These genes might be regulated by *bHLH20, bHLH9*, *ERF21*, *NAC4*, *bHLH5*, *WRKY7* and *WRKY8*, as they were embedded in the center of the network by hub genes screening.

Phylogenetic relationship and expression patterns of OSCs and TPSs family

From the phylogenetic tree illustrating the relatedness of OSCs in Amaranthaceae species (Fig. 6a), eight putative OSCs were identified in *A. squarrosum* (AsOSC), including β-amyrin synthase (BAS), lupeol synthase (LUS) and cycloartenol synthase (CAS), except that lanosterol synthases and multi-function OSCs were not found in Amaranthaceae family. In the BAS group, putative *AsBAS2* and *AsBAS3* shared a clade, while *CqBAS7*, *CqBAS8* and *CqBAS9* shared the other sister clade. In another clade, quinoa had more BAS gene copies and *A. squarrosum* had only *AsBAS1* belonging to this clade. In the LUS group,

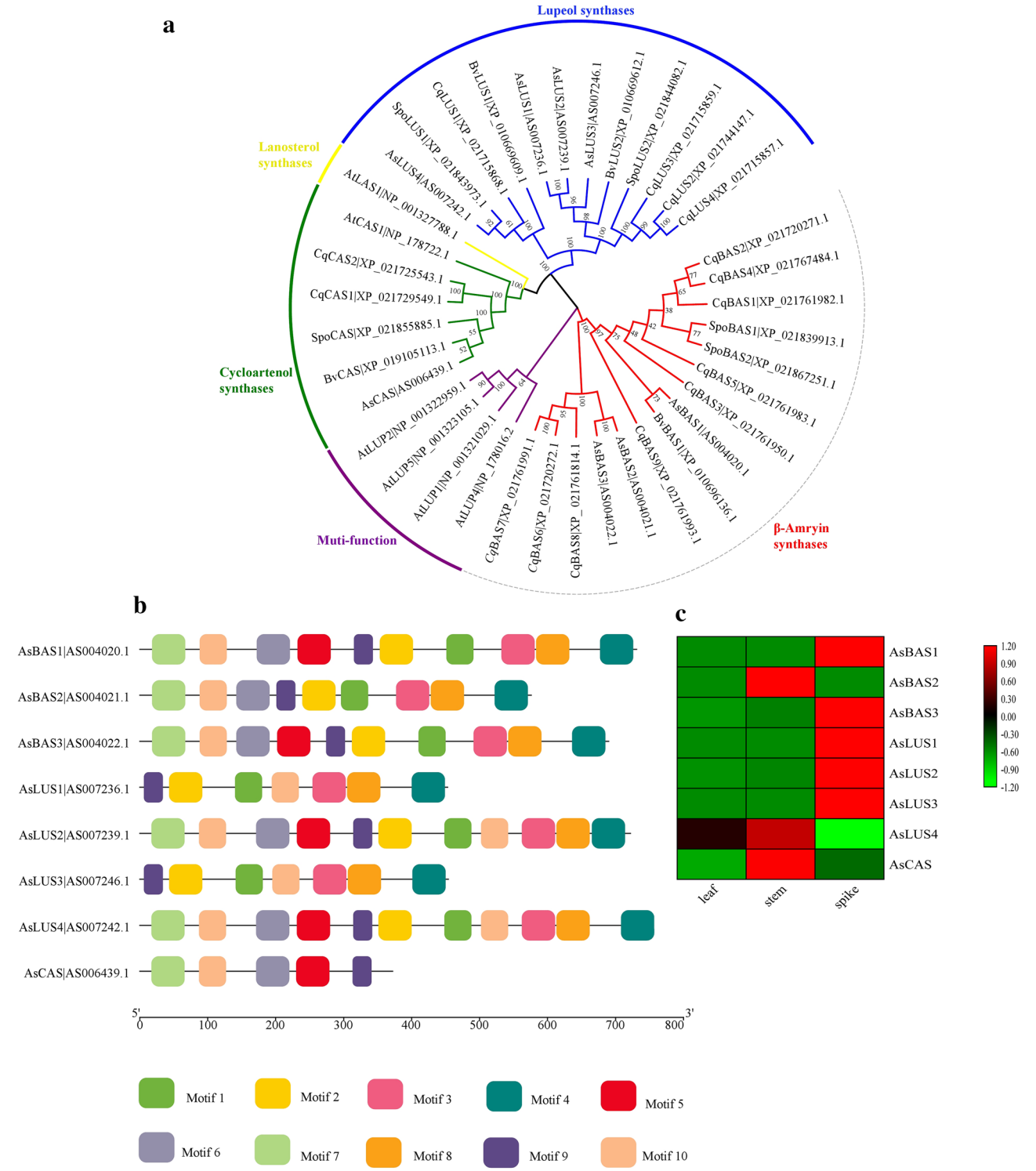


Fig. 6 Phylogenetic tree, conserved motif analysis and expression profiles of candidate OSCs in *A. squarrosum*. **a** Phylogenetic tree of OSCs candidates as compared to known enzymes from Arabidopsis and other candidates from three famous Amaranthaceae species. Cq, *Chenopodium quinoa*; Bv, *Beta vulgaris*; Spo, *Spinacia oleracea*. **b** Distributions of conserved motifs in OSCs genes. The 1–10 motifs were showed in different colored boxes. The sequence information for each motif was provided in Supplementary Table S10. **c** Expression profiles of *A. squarrosum* OSCs in different tissues

AsLUS1, *AsLUS2* and *AsLUS4* shared a clade. Compared with the other two Amaranthaceae species, *A. squarrosum* and quinoa increased gene copies in the one LUS clade. The gene expansion phenomenon in BAS and LUS suggested the synthesis of lupane/oleanane-type triterpenoid was extraordinarily active in *A. squarrosum* and quinoa.

As the predicted conserved motifs shown in Fig. 6b, motif 9 and motif 10 existed in the all candidate AsOSCs. OSCs in one group shared more similar motifs compositions. Among *AsBASs*, motif 5 was lack in *AsBAS2*. Among *AsLUSs*, *AsLUS1* and *AsLUS3* possessed the same number of motifs compositions (seven motifs), which was remarkably less than *AsLUS2* and *AsLUS4* (ten motifs). The expression patterns derived from different tissues transcriptome of all *AsOSC* genes were displayed in Fig. 6c. Most of the transcripts of AsBAS, AsLUS were all highly expressed in spike, while the transcripts of AsCAS expressed highly in stem.

Together with TPSs of Arabidopsis, candidate TPS-like genes from Amaranthaceae were selected to construct the phylogenetic tree (Fig. 7a, Supplementary Table S9). With the exception of TPS-d from gymnosperm *Picea abies* as an outgroup, TPS genes could be divided into five subfamilies. Among TPS-related genes subfamilies of *A. squarrosum*, TPS-a clade responsible to sesquiterpenes and diterpenes biosynthesis had the largest gene numbers (16 genes). Similar to 22 genes in Arabidopsis (Zhou and Pichersky 2020), both *A. squarrosum* and quinoa have more gene copies in this clade. Intriguingly, we discovered ten gene copies of TPS-g clade in *A. squarrosum*, whereas only one representative monoterpene synthase AtTPS14 was found in Arabidopsis. Among the Amaranthaceae species, the TPS-g genes only dramatically expanded in *A. squarrosum.*

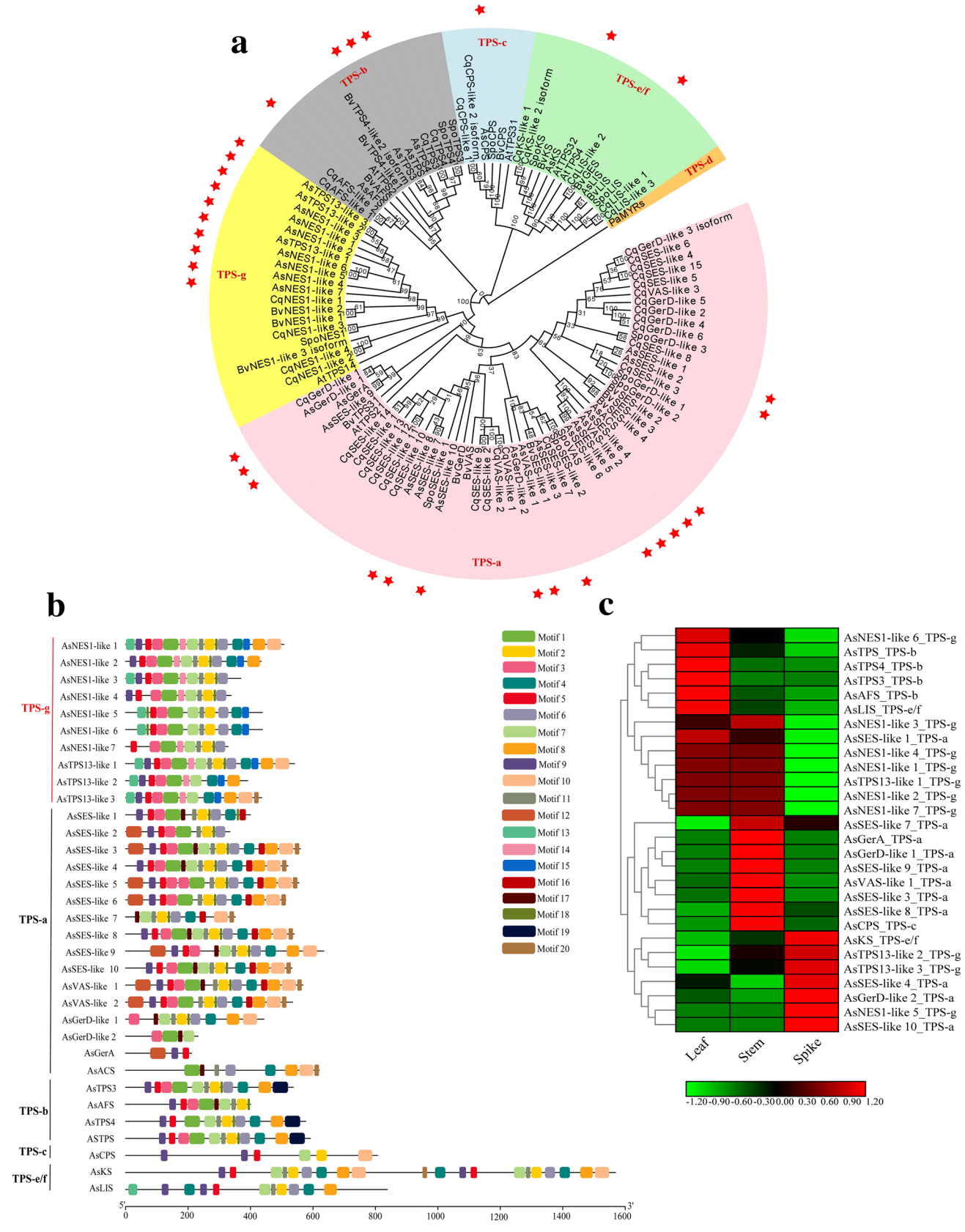
A schematic of predicted motifs in all AsTPSs exhibited in Fig. 7b. Based on MEME database, we identified 20 conserved motifs in AsTPSs. Most of motifs were widely distributed in the AsTPS family. The similar motifs were distributed within one subfamily, which indicated protein structure was conserved in one subfamily and function of the specific motif in one subfamily may be important for the gene function. Motif 12 was specific to TPS-a clade, while motif 14 and motif 15 were two unique motifs in TPS-g clade proteins. Among the TPS-g clade genes expression patterns, *AsNES1-like* genes exhibited high expression level in leaf and stem, except for *AsNES1-like 5*, while *AsTPS13-like* mainly highly expressed in spikes (Fig. 7c). The most of TPS-a clade genes were found high expression level in stem, while the TPS-b clade genes expressed highly in leaf.

To be mentioned, we found the relationship between gene expression levels and the accumulation of various terpenoids across different tissues of *A. squarrosum*. As showed in Fig. 8a, the significant accumulation of polypodine B (C27) in spikes (Fig. 1) was positively correlated with the expression levels of *AsBAS1*, *AsBAS2, AsLUS1, AsLUS2, AsLUS3*, all of which were highly expressed in spikes. Sabinene (C10) and 18-Nor-4,15-dihydroxyabieta-8,11,13-trien-7-one (C19) were significantly enriched in stems, both of which were positively correlated with high expression levels of TPS-a subfamily genes in stems. These positive correlations suggested tissue-specific expression levels could regulate the terpenoids accumulation across different tissues at the transcriptional levels.

Discussion

A. squarrosum is a folk herbal medicine native to Asian temperature deserts regions, and its terpenoids of above-ground tissues are believed as one of main medicinal components. Its dense trichomes and specific odor emitted by the above-ground parts suggested this herb might contain rich volatile terpenoids. On the other hand, the extreme habitats stresses of alpine desert could also be associated with the accumulation of terpenoids. However, apart from part of triterpenoid components identified by pharmacognosy researches (Gong et al. 2012; Liu et al. 2013; Kong et al. 2018), little is known to molecular mechanisms underlying terpenoids biosynthesis in *A. squarrosum*. To explore the genetic information of terpenoid biosynthesis in this non-model psammophyte and understand the biological and pharmacological significance, in this study, across different *A. squarrosum* tissues sampled from alpine desert, we firstly found differences in accumulation of various terpenoids by non-targeted metabolomic profiling and then conducted comparative transcriptome analysis. We obtained a total of 135, 686 unigenes and searched against Nr, Nt, Swiss-Prot, KEGG, KOG and GO public databases using BLAST algorithm (E-value $< 10^{-5}$). Ultimately, we identified a total of 84 unigenes involved in the upstream terpenoid backbone biosynthesis and 53 unigenes involved in the terpenoid functional diversification. Based on analysis on tissue-specific expression patterns and phylogenetic relationship, the potential roles of the genes involved in terpenoids biosynthesis and their functional diversity were fully discussed in this desert herb.

The MVA and MEP pathways are usually regulated by multiple levels, including the transcriptional regulation, translation/post-translational levels, feedback regulation (Hemmerlin 2013; Tholl 2015), mainly from the evidences of gene expression analysis. In the MVA pathway, seven genes encoding AACTs, eleven genes encoding HMGSs and three genes encoding HMGRs were significantly up-regulated in leaf. To be mentioned, the rate-limiting enzyme HMGR is co-regulated with HMGS, which has been reported in many eukaryotic organisms (Liao et al. 2020). We proposed that this expression module of *HMGR* and *HMGS* plays an important role in the tissue-specific accumulation of terpenoids in *A. squarrosum*. In the MEP


a
TPS-b
TPS-c
TPS-e/f
TPS-d
TPS-g
TPS-a
b
TPS-g
TPS-a
TPS-b
TPS-c
TPS-e/f
AsNES1-like 1
AsNES1-like 2
AsNES1-like 3
AsNES1-like 4
AsNES1-like 5
AsNES1-like 6
AsNES1-like 7
AsTPS13-like 1
AsTPS13-like 2
AsTPS13-like 3
AsSES-like 1
AsSES-like 2
AsSES-like 3
AsSES-like 4
AsSES-like 5
AsSES-like 6
AsSES-like 7
AsSES-like 8
AsSES-like 9
AsSES-like 10
AsVAS-like 1
AsVAS-like 2
AsGerD-like 1
AsGerD-like 2
AsGerA
AsACS
AsTPS3
AsAFS
AsTPS4
ASTPS
AsCPS
AsKS
AsLIS
Motif 1
Motif 2
Motif 3
Motif 4
Motif 5
Motif 6
Motif 7
Motif 8
Motif 9
Motif 10
Motif 11
Motif 12
Motif 13
Motif 14
Motif 15
Motif 16
Motif 17
Motif 18
Motif 19
Motif 20
5'
3'
0
200
400
600
800
1000
1200
1400
1600
c
AsNES1-like 6_TPS-g
AsTPS_TPS-b
AsTPS4_TPS-b
AsTPS3_TPS-b
AsAFS_TPS-b
AsLIS_TPS-e/f
AsNES1-like 3_TPS-g
AsSES-like 1_TPS-a
AsNES1-like 4_TPS-g
AsNES1-like 1_TPS-g
AsTPS13-like 1_TPS-g
AsNES1-like 2_TPS-g
AsNES1-like 7_TPS-g
AsSES-like 7_TPS-a
AsGerA_TPS-a
AsGerD-like 1_TPS-a
AsSES-like 9_TPS-a
AsVAS-like 1_TPS-a
AsSES-like 3_TPS-a
AsSES-like 8_TPS-a
AsCPS_TPS-c
AsKS_TPS-e/f
AsTPS13-like 2_TPS-g
AsTPS13-like 3_TPS-g
AsSES-like 4_TPS-a
AsGerD-like 2_TPS-a
AsNES1-like 5_TPS-g
AsSES-like 10_TPS-a
Leaf
Stem
Spike
-1.20 -0.90 -0.60 -0.30 0.00 0.30 0.60 0.90 1.20

◂**Fig. 7** Phylogenetic tree, conserved motif analysis and expression profiles of candidate TPS gene family in *A. squarrosum*. **a** Phylogenetic tree of TPS candidates as compared to known enzymes from Arabidopsis and other candidates from three famous Amaranthaceae species. The TPS-a, TPS-g, TPS-b, TPS-c and TPS-e/f clades are shading in pink, yellow, grey, blue and green, respectively. The *Picea abies* PaMYRs belonging to TPS-d clade was used to root the tree. TPS genes in *A. squarrosum* were marked by red star. Cq, *Chenopodium quinoa*; Bv, *Beta vulgaris*; Spo, *Spinacia oleracea*. **b** Distributions of conserved motifs in TPSs genes. TPS-g clade genes names were bolded by red. The 1–20 motifs were showed in different colored boxes. The sequence information for each motif was provided in Supplementary Table S11. **c** Expression profiles of *A. squarrosum* TPS gene family in different tissues

pathway, except that three genes encoding DXS expressed highly in stem, most of genes mainly expressed highly in leaf. Previous studies showed that DXS was the main rate-determining enzyme with the highest flux control force in the MEP pathway (Tong et al. 2015). The excessive expression of *DXS* can effectively enhance the terpenoid biosynthesis, which was associated with the biosynthesis of plastidial terpenes (Yuan et al. 2018). Compared to the other plant species, our findings on *DXS*s which were highly expressed in stem suggested they could also participate in the cross-talk between monoterpenoids and diterpenoids biosynthesis in the plastids and/or transportation.

At branch points, almost all of *IPPI*, *FPPS*, *GPPS* and *GGPPS* genes were detected to express across all the tissues. Among them, *IPPI9*, *IPPI10*, *FPPS3*, *GPPS2*, *GPPS5*, *GGPPS1* and *GGPPS2* were significantly up-regulated in leaf. FPPS is a branch-point enzyme directing carbon precursor to flow to triterpene pathway and regulates the metabolic flux (Dhar et al. 2013). GPPS catalyzed GPP synthesize precursor of monoterpenes, which may be associated with plant fragrance regulation (Kamran et al. 2020). Here, both of two *GGPPS* genes expressed highly in leaf, and GGPPSs family is known to act as the large subunit of heteromeric GPPS and control IPP flux towards pigments, phytohormones biosynthesis in green tissues (Barsain et al. 2020). This kind of expression patterns, that the upstream terpenoids biosynthesis genes highly expressed in the vegetative tissues, were also found in other terpenoid-riched plants (Biazzi et al. 2015). The tissue-specific expression of terpenoids backbone biosynthetic genes indicated that the accumulation of terpenoid precursors was largely influenced by plant development and its stress responses (Tholl 2015; Vranová et al. 2012).

To be mentioned, *GGPPS2* was one of the hub genes in the co-expression network of terpenoid biosynthesis in *A. squarrosum*, and several transcription factors, such as *bHLH20, bHLH9*, *ERF21*, *NAC4*, *bHLH5*, *WRKY7* and *WRKY8*, were strongly associated with its expression. Among them, *bHLH20,* a homologus gene of *bZIP1* responding to abiotic and biotic stresses in Arabidopsis (Sun et al. 2011) positively interacted with *GGPPS2* expression, which suggested that terpenoid biosynthesis could also be involved into stress responses for *A. squarrosum*. A homolog of *ERF21* was reported to take part in the ethylene-activated signaling pathway, which was closely related with stress responses in plants (Nagegowda and Gupta 2020). *WRKY7* and *WRKY8* are homologs to *WRKY42* and *WRKY4* of Arabidopsis, which participated in ethylene, salicylic acid signaling responses and defense responses (Lai et al. 2008; Niu et al. 2020). Remarkably, *bHLH5* is a homologous gene of *MYC2*, which acted as a regulatory factor in all kinds of secondary metabolism in Arabidopsis (Sui et al. 2018; Yang et al. 2020). Actually, terpenoid metabolism was regulated by AP2-EREBP, bHLH, WRKY, NAC, MYB and bZIP families, which have been identified in other species (Wang et al. 2016; Nagegowda and Gupta 2020). Thus, as most terpenoids biosynthesis is controlled primarily at transcriptional levels, we proposed that the terpenoid biosynthesis of *A. squarrosum* might be more or less regulated by transcription factors responding to abiotic and biotic stresses based on co-expression network in this study.

Since terpenoids were believed to protect plants from oxidative damage (ROS), temperature stress, UV-B and drought, even pests (Grassmann 2005; Yu and Wei 2019), the desert plants such as *A. squarrosum* could utilize the ecological functions of diverse terpenoids to adapt to the multiple severe stresses in their habitats. Hereby, the phylogenetic relationship and functional characteristics of the downstream gene family OSCs and TPSs improve our understanding of prediction to potential chemical diversity of terpenoids in *A. squarrosum*. OSCs were the first-step enzymes cyclizing the 2,3-oxidosqualene into diverse triterpene products (Shang and Huang 2019). Based on the function and motif compositions, plant OSC*s* can be classified into LUS, BAS, CAS and LAS. LAS was absent in Amaranthaceae species. Among eight OSC genes newly identified in *A. squarrosum*, we found three BAS genes and four LUS genes, all of which catalyzed the synthesis of triterpenoid saponins precursors. In this study, the transcripts encoding *AsBASs* and *AsLUSs* mainly expressed highly in spikes, which was significantly positively correlated with triterpnoids such as polypodine B accumulation in spikes (Fig. 8a). This kind of tissue-specific expression was not found in the relative species quinoa (Fiallos-Jurado et al. 2016). Notably, two *AsBAS* (1 and 3) and two *AsLUS* (2 and 4) shared ten motifs, while *AsBAS2* was lack of motif 5 and four motifs (5, 6, 7 and 10) were absent in *AsLUS1* and *AsLUS3* (Fig. 6b). We conjectured that the variation of motifs composition may result in function differentiation of genes and product diversification of proteins. In addition, the phylogenetic tree of OSCs family in Amaranthaceae illustrated that LUS and BAS genes had undergone expansions both in *A. squarrosum* and quinoa. Previous studies have figured out that these two subfamilies

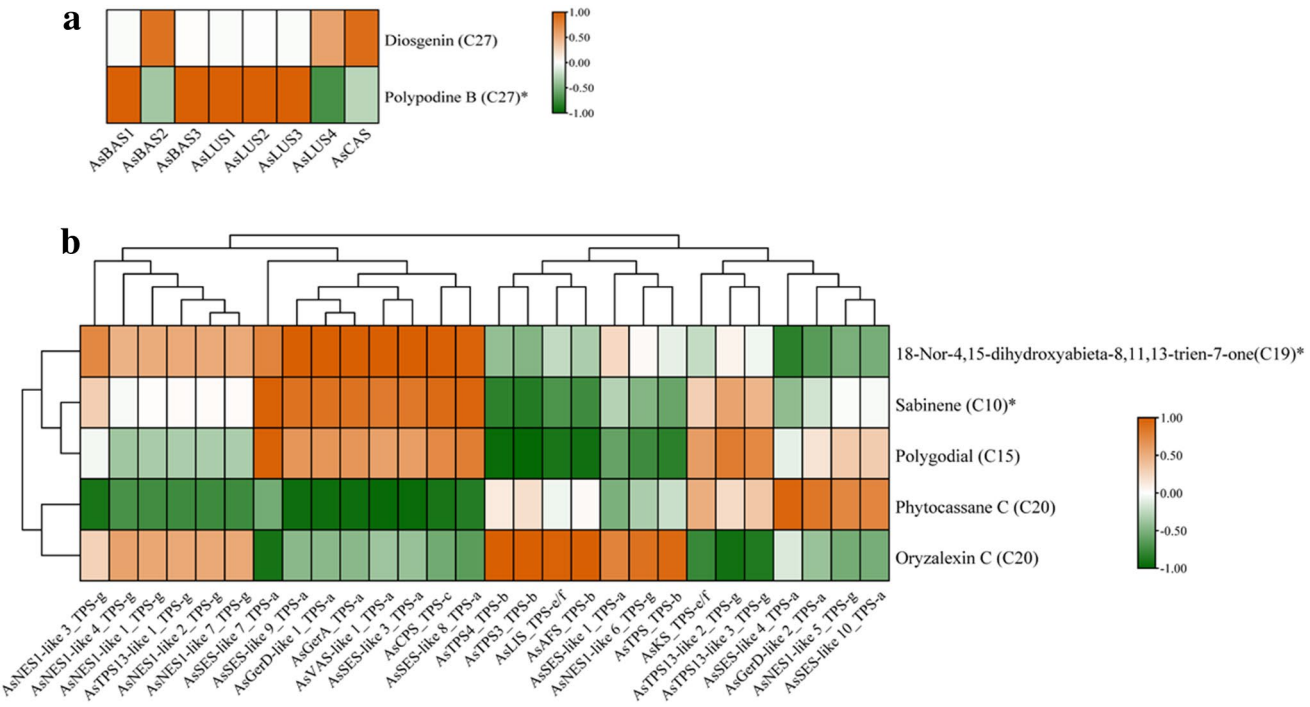


Fig. 8 Correlation analysis of OSCs (**a**) and TPSs (**b**) genes expression levels and various detected terpenoids content. The row represents a terpenoid metabolite, while the column represents a gene. The different colors represents the degree of the correlation coefficient. *Represents significantly differential enriched terpenoids in different tissues

were involved in the durability to the multiple abiotic stresses by the accumulation of triterpenoid saponins in quinoa (Adolf et al. 2012; Fiallos-Jurado et al. 2016; Jarvis et al. 2017). It had been reported that lupane/oleanane-type triterpenoid and steroidal saponins accumulation were influenced by abiotic stress, including humidity, poor nutrient, light and temperature (Moses et al. 2014). Similar to quinoa, the expansions and sub-functions of *AsBASs* and *AsLUSs* could also contribute to the multiple stresses tolerance for *A. squarrosum* through the tissue-specific accumulation of diverse triterpenoids.

For other kind of terpenoids biosynthesis, the functional diversification of TPSs family plays important roles in plant kingdom (Alicandri et al. 2020). TPS gene family has been assigned into seven subfamilies, TPS-a to TPS-h, and characterized classified functions (Chen et al. 2011). TPS-a, TPS-b and TPS-g subfamilies are specific to angiosperm. Based on protein sequences, the phylogentic relationship of TPSs gene family in Amaranthaceae species could be classed into five major subfamilies (TPS-a, TPS-b, TPS-c, TPS-e/f and TPS-g), most of which shared the conserved motifs within a subfamily. Interestingly, compared to the other Amaranthaceae species and Arabidopsis (only *AtTPS14*), significant expansions of TPS-g subfamily (seven NES1-like and three TPS13-like genes) were only found in *A. squarrosum*, suggesting the potential diversification and accumulation of monoterpenoid. Previous studies reported that TPS-g subfamily devoted to generate the acyclic monoterpene products with volatile properties (Dudareva et al. 2003). Monoterpenoids are famous for their strong odor and aromatic properties and are also employed in adaptation to biotic and abiotic stresses for plants (Nagegowda and Gupta 2020). They can derived from essential oils of aromatic plants and be widely used in perfume, food, and pharmaceutical industries (Ludwiczuk et al. 2017). The unique expansion of TPS-g subfamily genes in *A. squarrosum* may result in species-specific accumulation of the acyclic monoterpenes. Unfortunately, we have not detected the acyclic monoterpene in this study with the limitation of non-targeted metabolomic analysis. Further studies should be conducted to identify volatile terpenoids combined with dense trichomes and specific odor phenotypes in *A. squarrosum*. Remarkably, although *AsNES1-like 5* and *AsNES1-like 6* showed different expression pattern in stem and spike, the same motifs composition of them suggested these two genes might encode the same isozyme and their products might keep the same to each other. This kind of expression patterns were also found among the *AsSES-like* genes. Genes in TPS-a and TPS-b subfamilies, two subfamilies responsible for volatile terpenoids biosynthesis, mainly highly expressed in flowers, glandular trichomes and roots of multiple angiosperm species, which improved biotic and abiotic stresses

tolerance for plant adaptibility (Zhou and Pichersky 2020; Li et al. 2021). However, the high expression level of TPS-a subfamily genes in stem of *A. squarrosum* had a positive relationship with the significant accumulation of sabinene and 18-Nor-4,15-dihydroxyabieta-8,11,13-trien-7-one in stem (Fig. 8b). This tissue-specific expression patterns and the expansion of TPS-g subfamily genes indicated that the categories and functions of terpenoids were significantly differentiated during evolution process in *A. squarrosum*. These genetic resources will enhance our understanding of the molecular basis of medicinal properties and ecological functions of terpene-related biosynthesis pathways in this folk medicinal plant.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11816-021-00674-5.

Acknowledgements This research was supported by the National Key Research and Development Program of China (Grant No. 2016YFC0500903), the National Natural Science Foundation of China (Grant No.31901079 and 31770416), the Key Research and Development Program of Gansu Province (Grant No.18YF1NA123), the Xinjiang Production and Construction Crops Key Laboratory of Protection and Utilization of Biological Resources in Tarim Basin (Grant No. BRZD1809), "Light of West China" Program of CAS (Grant No. Y92972) and the Special Funds for Scientific and Technological Innovation and Development in Gansu Province (Research and Development Center of Desert Nutrition and Medicinal Plant Germplasm Innovation).

References

Abbas F, Yu R (2017) Volatile terpenoids: multiple functions, biosynthesis, modulation and manipulation by genetic engineering. Planta 246:803–816. https://doi.org/10.1007/s00239-020-09930-810.1007/s00425-017-2749-x

Adolf VI, Shabala S, Andersen MN, Razzaghi F, Jacobsen S-E (2012) Varietal differences of quinoa's tolerance to saline conditions. Plant Soil 357:117–129. https://doi.org/10.1007/s00239-020-09930-810.1007/s11104-012-1133-7

Alicandri E, Paolacci AR, Osadolor S, Sorgonà A, Badiani M, Ciaffi M (2020) On the evolution and functional diversity of terpene synthases in the Pinus species: a review. J Mol Evol 88:253–283. https://doi.org/10.1007/s00239-020-09930-8

Amini H, Naghavi MR, Shen T et al (2019) Tissue-specific transcriptome analysis reveals candidate genes for terpenoid and phenylpropanoid metabolism in the medicinal plant *Ferula Assafoetida*. G3 9:807–816. https://doi.org/10.1534/g3.118.200852

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:R106. https://doi.org/10.1007/s00239-020-09930-810.1186/gb-2010-11-10-r106

Bao M (2017) Study on hypolipidemic activity of *Agriophyllum*. Inner Mongolia University (**In Chinese**)

Bao S, Han S, Cheli-geer C-R, Ao W (2016) Effects of agiophyllum oligosaccharides on insulin resistance of Goto-Kakizaki rats. Chin Pharm Bull 32:403–409. https://doi.org/10.3969/j.issn.1001-1978.2016.03.021

Barsain BL, Purohit A, Kumar A, Joshi R, Hallan V, Yadav SK (2020) PkGPPS.SSU interacts with two PkGGPPS to form heteromeric GPPS in *Picrorhiza kurrooa*: molecular insights into the picroside biosynthetic pathway. Plant Physiol Biochem 154:115–128. https://doi.org/10.1016/j.plaphy.2020.05.029

Biazzi E, Carelli M, Tava A et al (2015) CYP72A67 catalyzes a key oxidative step in *Medicago truncatula* hemolytic saponin biosynthesis. Mol Plant 8:1493–1506. https://doi.org/10.1016/j.molp.2015.06.003

Chen F, Tholl D, Bohlmann J, Pichersky E (2011) The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. Plant J 66:212–229. https://doi.org/10.1111/j.1365-313X.2011.04520.x

Chen G, Zhao J, Zhao X, Zhao P, Duan R, Nevo E, Ma X (2014a) A psammophyte *Agriophyllum squarrosum* (L.) Moq.: a potential food crop. Genet Resour Crop Evol 61:669–676. https://doi.org/10.1007/s10722-014-0083-8

Chen H, Li G, Köllner TG, Jia Q, Gershenzon J, Chen F (2014b) Positive Darwinian selection is a driving force for the diversification of terpenoid biosynthesis in the genus Oryza. BMC Plant Biol 14:239. https://doi.org/10.1186/s12870-014-0239-x

Chen C, Zheng Y, Zhong Y, Wu Y, Li Z, Xu L, Xu M (2018) Transcriptome analysis and identification of genes related to terpenoid biosynthesis in *Cinnamomum camphora*. BMC Genomics 19:550. https://doi.org/10.1186/s12864-018-4941-1

Dhar MK, Koul A, Kaul S (2013) Farnesyl pyrophosphate synthase: a key enzyme in isoprenoid biosynthetic pathway and potential molecular target for drug development. New Biotech 30:114–123. https://doi.org/10.1016/j.nbt.2012.07.001

Dudareva N, Martin D, Kish CM et al (2003) (*E*)-β-Ocimene and myrcene synthase genes of floral scent biosynthesis in Snapdragon: function and expression of three terpene synthase genes of a new terpene synthase subfamily. Plant Cell 15:1227. https://doi.org/10.1105/tpc.011015

Fiallos-Jurado J, Pollier J, Moses T et al (2016) Saponin determination, expression analysis and functional characterization of saponin biosynthetic genes in *Chenopodium quinoa* leaves. Plant Sci 250:188–197. https://doi.org/10.1016/j.plantsci.2016.05.015

Fujihashi M, Sato T, Tanaka Y et al (2018) Crystal structure and functional analysis of large-terpene synthases belonging to a newly found subclass. Chem Sci 9:3754–3758. https://doi.org/10.1039/C8SC00289D

Genievskaya Y, Abugalieva S, Zhubanysheva A, Turuspekov Y (2017) Morphological description and DNA barcoding study of sand rice (*Agriophyllum squarrosum*, Chenopodiaceae) collected in Kazakhstan. BMC Plant Biol 17:177–177. https://doi.org/10.1186/s12870-017-1132-1

Gong B, Zhan K, Zhou Y-h, Zhang L, Hui Y-q, Li Y (2012) Separation and identification of chemical constituents from *Agriophyllum squarrosum* (L.) Moq. Modern Chin Med. 14:7–11. https://doi.org/10.13313/j.issn.1673-4890.2012.10.016

Götz S, García-Gómez J, Terol JM et al (2008) High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 36:3420–3435. https://doi.org/10.1093/nar/gkn176

Grabherr MG, Haas BJ, Yassour M et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652. https://doi.org/10.1038/nbt.1883

Grassmann J (2005) Terpenoids as plant antioxidants. In: Litwack G (ed) Vitamins & hormones, vol 72. Academic Press, pp 505–535. https://doi.org/10.1016/S0083-6729(05)72015-X

Hemmerlin A (2013) Post-translational events and modifications regulating plant enzymes involved in isoprenoid precursor biosynthesis. Plant Sci 203–204:41–54. https://doi.org/10.1016/j.plantsci.2012.12.008

Huber W, Carey VJ, Long L, Falcon S, Gentleman R (2007) Graphs in molecular biology. BMC Bioinformatics 8:S8. https://doi.org/10.1186/1471-2105-8-S6-S8

Jarvis DE, Ho YS, Lightfoot DJ et al (2017) The genome of *Chenopodium quinoa*. Nature 542:307–312. https://doi.org/10.1038/nature21370

Kamran HM, Hussain SB, Junzhong S, Xiang L, Chen L-Q (2020) Identification and molecular characterization of Geranyl Diphosphate Synthase (GPPS) genes in Wintersweet flower. Plants 9:666. https://doi.org/10.3390/plants9050666

Kishikawa A, Amen Y, Shimizu K (2017) Anti-allergic triterpenes isolated from olive milled waste. Cytotechnology 69:307–315. https://doi.org/10.1007/s10616-016-0058-z

Kong L, Zhang L, LI Y, (2018) Isolation and identification of oleanane type triterpenoid saponins from *Agriophyllum squarrosum*. J Shenyang Pharm Univ 35:443–447. https://doi.org/10.14066/j.cnki.cn21-1349/r.2018.06.002

Lai Z, Vinod K, Zheng Z, Fan B, Chen Z (2008) Roles of Arabidopsis WRKY3 and WRKY4 transcription factors in plant responses to pathogens. BMC Plant Biol 8:68. https://doi.org/10.1186/1471-2229-8-68

Li J, Wang Y, Dong Y et al (2021) The chromosome-based lavender genome provides new insights into Lamiaceae evolution and terpenoid biosynthesis. Hortic Res 8:53. https://doi.org/10.1038/s41438-021-00490-6

Liao P, Lung S-C, Chan WL, Bach TJ, Lo C, Chye M-L (2020) Overexpression of HMG-CoA synthase promotes Arabidopsis root growth and adversely affects glucosinolate biosynthesis. J Exp Bot 71:272–289. https://doi.org/10.1093/jxb/erz420

Liu Z, Li Y, Qu J, Li Y (2013) Triterpene saponins from the aerial parts of *Agriophyllum squarrosum*. Modern Chinese Med 15:936–939. https://doi.org/10.13313/j.issn.1673-4890.2013.11.013

Liu X, Hou X, Liu C et al (2019) De novo assembly and comparative transcriptome analysis: novel insights into terpenoid biosynthesis in *Chamaemelum nobile* L. Plant Cell Rep 38:101–116. https://doi.org/10.1007/s00299-018-2352-z

Lo H-Y, Li C-C, Huang H-C, Lin L-J, Hsiang C-Y, Ho T-Y (2012) Application of transcriptomics in Chinese herbal medicine studies. J Tradit Complem Med 2:105–114. https://doi.org/10.1016/S2225-4110(16)30083-9

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8

Ludwiczuk A, Skalicka-Woźniak K, Georgiev MI (2017) Chapter 11—terpenoids. In: Badal S, Delgoda R (eds) Pharmacognosy. Academic Press, Boston, pp 233–266. https://doi.org/10.1016/B978-0-12-802104-0.00011-1

Misawa N (2011) Pathway engineering for functional isoprenoids. Curr Opin Biotechnol 22:627–633. https://doi.org/10.1016/j.copbio.2011.01.002

Moses T, Papadopoulou KK, Osbourn A (2014) Metabolic and functional diversity of saponins, biosynthetic intermediates and semi-synthetic derivatives. Crit Rev Biochem Mol Biol 49:439–462. https://doi.org/10.3109/10409238.2014.953628

Nagegowda DA, Gupta P (2020) Advances in biosynthesis, regulation, and metabolic engineering of plant specialized terpenoids. Plant Sci 294:110457. https://doi.org/10.1016/j.plantsci.2020.110457

Nazaruk J, Borzym-Kluczyk M (2015) The role of triterpenes in the management of diabetes mellitus and its complications. Phytochem Rev 14:675–690. https://doi.org/10.1007/s11101-014-9369-x

Niu F, Cui X, Zhao P et al (2020) WRKY42 transcription factor positively regulates leaf senescence through modulating SA and ROS synthesis in *Arabidopsis thaliana*. Plant J. https://doi.org/10.1111/tpj.14914

Park S-B, Kim S-H, Suk K et al (2010) *Clinopodium gracile* inhibits mast cell-mediated allergic inflammation: involvement of calcium and nuclear factor-κB. Exp Biol Med 235:606–613. https://doi.org/10.1258/EBM.2010.009292

Qian C, Yin H, Shi Y et al (2016) Population dynamics of *Agriophyllum squarrosum*, a pioneer annual plant endemic to mobile sand dunes, in response to global climate change. Sci Rep 6:26613. https://doi.org/10.1038/srep26613

Rios J-L (2010) Effects of triterpenes on the immune system. J Ethnopharmacol 128:1–14. https://doi.org/10.1016/j.jep.2009.12.045

S KP, Philipp Z (2019) Terpene synthases as metabolic gatekeepers in the evolution of plant terpenoid chemical diversity. Front Plant Sci 10:1166. https://doi.org/10.3389/fpls.2019.01166

Saqier BS, Han S, Ao W (2019) Effects of *Agriophyllum squarrosum* extracts on glucose metabolism in KKAy mice and the associated underlying mechanisms. J Ethnopharmacol 241:112009. https://doi.org/10.1016/j.jep.2019.112009

Sawai S, Saito K (2011) Triterpenoid biosynthesis and engineering in plants. Front Plant Sci 2:25. https://doi.org/10.3389/fpls.2011.00025

Shang Y, Huang S (2019) Multi-omics data-driven investigations of metabolic diversity of plant triterpenoids. Plant J 97:101–111. https://doi.org/10.1111/tpj.14132

Sui X, Singh SK, Patra B, Schluttenhofer C, Guo W, Pattanaik S, Yuan L (2018) Cross-family transcription factor interaction between MYC2 and GBFs modulates terpenoid indole alkaloid biosynthesis. J Exp Bot 69:4267–4281. https://doi.org/10.1093/jxb/ery229

Sun X, Li Y, Cai H, Bai X, Ji W, Ding X, Zhu Y (2011) The Arabidopsis AtbZIP1 transcription factor is a positive regulator of plant tolerance to salt, osmotic and drought stresses. J Plant Res 125:429–438. https://doi.org/10.1007/s10265-011-0448-4

Szakiel A, Pączkowski C, Henry M (2011) Influence of environmental abiotic factors on the content of saponins in plants. Phytochem Rev 10:471–491. https://doi.org/10.1007/s11101-010-9177-x

Tai Y, Hou X, Liu C et al (2020) Phytochemical and comparative transcriptome analyses reveal different regulatory mechanisms in the terpenoid biosynthesis pathways between *Matricaria recutita* L. and *Chamaemelum nobile* L. BMC Genomics 21:169. https://doi.org/10.1186/s12864-020-6579-z

Testone G, Mele G, di Giacomo E et al (2019) Transcriptome driven characterization of curly- and smooth-leafed endives reveals molecular differences in the sesquiterpenoid pathway. Hortic Res 6:1. https://doi.org/10.1038/s41438-018-0066-6

Thimmappa R, Geisler K, Louveau T, O'Maille P, Osbourn A (2014) Triterpene biosynthesis in plants. Annu Rev Plant Biol 65:225–257. https://doi.org/10.1146/annurev-arplant-050312-120229

Tholl D (2006) Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. Curr Opin Plant Biol 9:297–304. https://doi.org/10.1016/j.pbi.2006.03.014

Tholl D (2015) Biosynthesis and biological functions of terpenoids in plants. In: Schrader J, Bohlmann J (eds) Biotechnology of isoprenoids. Springer International Publishing, Cham, pp 63–106. https://doi.org/10.1007/10_2014_295

Tong Y, Su P, Zhao Y et al (2015) Molecular cloning and characterization of DXS and DXR Genes in the terpenoid biosynthetic pathway of *Tripterygium wilfordii*. Int J Mol Sci 16:25516–25535. https://doi.org/10.3390/ijms161025516

Vranová E, Coman D, Gruissem W (2012) Structure and dynamics of the isoprenoid pathway network. Mol Plant 5:318–333. https://doi.org/10.1093/mp/sss015

Vranová E, Coman D, Gruissem W (2013) Network analysis of the MVA and MEP pathways for isoprenoid synthesis. Annu Rev Plant Biol 64:665–700. https://doi.org/10.1146/annurev-arplant-050312-120116

Wang Q, Reddy VA, Panicker D et al (2016) Metabolic engineering of terpene biosynthesis in plants using a trichome-specific transcription factor MsYABBY5 from spearmint (*Mentha spicata*). Plant Biotechnol J 14:1619–1632. https://doi.org/10.1111/pbi.12525

Xu J, Wang S, Feng T, Chen Y, Yang G (2018) Hypoglycemic and hypolipidemic effects of total saponins from *Stauntonia chinensis*

in diabetic db/db mice. J Cell Mol Med 22:6026–6038. https://doi.org/10.1111/jcmm.13876

Yadav V, Prasad S, Bokyung S, Kannappan R, Aggarwal B (2010) Targeting inflammatory pathways by triterpenoids for prevention and treatment of cancer. Toxins 2:2428–2466. https://doi.org/10.3390/toxins2102428

Yang L, Yang C, Chenyi L, Zhao Q, Liu L, Xin F, Chen X-Y (2015) Recent advances in biosynthesis of bioactive compounds in traditional Chinese medicinal plants. Sci Bull. https://doi.org/10.1007/s11434-015-0929-2

Yang Z, Li Y, Gao F et al (2020) MYB21 interacts with MYC2 to control the expression of terpene synthase genes in flowers of *Freesia hybrida* and *Arabidopsis thaliana*. J Exp Bot 71:4140–4158. https://doi.org/10.1093/jxb/eraa184

Yin X, Wang W, Qian C, Fan X, Yan X, Yan G, Ma X (2018) Analysis of metabolomics in *Agriophyllum squarrosum* based on UPLC-MS. Chin J Exp Tradit Med Form 24:51–56. https://doi.org/10.13422/j.cnki.syfjx.20181503

Yu C, Wei J (2019) Research advance on the effects of environmental factors on the biosynthesis of terpenoids in plants. Acta Bot Boreal-Occident Sin 39:1710–1710. https://doi.org/10.7606/j.issn.1000-4025.2019.09.1701

Yuan Y, Yu M, Jia Z, Xe S, Liang Y, Zhang J (2018) Analysis of *Dendrobium huoshanense* transcriptome unveils putative genes associated with active ingredients synthesis. BMC Genomics 19:978–978. https://doi.org/10.1186/s12864-018-5305-6

Zhang J, Zhao P, Zhao J, Chen G (2018) Synteny-based mapping of causal point mutations relevant to sand rice (*Agriophyllum squarrosum*) trichomeless1 mutant by RNA-sequencing. J Plant Physiol 231:86–95. https://doi.org/10.1016/j.jplph.2018.09.003

Zhao P, Capella-Gutiérrez S, Shi Y, Zhao X, Chen G, Gabaldón T, Ma X-F (2014) Transcriptomic analysis of a psammophyte food crop, sand rice (*Agriophyllum squarrosum*) and identification of candidate genes essential for sand dune adaptation. BMC Genomics 15:872. https://doi.org/10.1186/1471-2164-15-872

Zhao P, Zhang J, Qian C, Zhou Q, Zhao X, Chen G, Ma X (2017) SNP discovery and genetic variation of candidate genes relevant to heat tolerance and agronomic traits in natural populations of sand rice (*Agriophyllum squarrosum*). Front Plant Sci 8:536–536. https://doi.org/10.3389/fpls.2017.00536

Zhou F, Pichersky E (2020) The complete functional characterisation of the terpene synthase family in tomato. New Phytol 226(5):1341–1360. https://doi.org/10.1111/nph.16431

Zhu Y (2000) In botanical pharmacopeias of inner Mongolia, vol 1. Inner Mongolia People's Publishing House (**In Chinese**)