



OPEN ACCESS

EDITED BY

Xiaozeng Yang,
Beijing Academy of Agricultural and
Forestry Sciences, China

REVIEWED BY

Anuj Kumar,
Dalhousie University, Canada
Jlabao Ye,
Yangtze University, China

*CORRESPONDENCE

Liming Jia

✉ jlm@bjfu.edu.cn

Zhong Chen

✉ zhongchen@bjfu.edu.cn

[†]These authors have contributed
equally to this work

SPECIALTY SECTION

This article was submitted to
Functional and Applied Plant
Genomics,
a section of the journal
Frontiers in Plant Science

RECEIVED 06 September 2022

ACCEPTED 19 December 2022

PUBLISHED 09 January 2023

CITATION

Xu Y, Liu J, Ji X, Zhao G, Zhao T,
Wang X, Wang L, Gao S, Hao Y, Gao Y,
Gao Y, Weng X, Jia L and Chen Z
(2023) Integrative analysis of
microRNAs and mRNAs reveals the
regulatory networks of triterpenoid
saponin metabolism in Soapberry
(*Sapindus mukorossi* Gaertn.).
Front. Plant Sci. 13:1037784.
doi: 10.3389/fpls.2022.1037784

COPYRIGHT

© 2023 Xu, Liu, Ji, Zhao, Zhao, Wang,
Wang, Gao, Hao, Gao, Gao, Weng, Jia
and Chen. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Integrative analysis of microRNAs and mRNAs reveals the regulatory networks of triterpenoid saponin metabolism in Soapberry (*Sapindus mukorossi* Gaertn.)

Yuanyuan Xu^{1,2,3†}, Jiming Liu^{1,2,3†}, Xiangqin Ji⁴,
Guochun Zhao^{1,2,3}, Tianyun Zhao^{1,2,3}, Xin Wang^{1,2,3},
Lixian Wang^{1,2,3}, Shilun Gao^{1,2,3}, Yingying Hao^{1,2,3},
Yuhan Gao^{1,2,3}, Yuan Gao⁵, Xuehuang Weng⁶,
Liming Jia^{1,2,3*} and Zhong Chen^{1,2,3,7*}

¹Key Laboratory of Silviculture and Conservation of the Ministry of Education, College of Forestry, Beijing Forestry University, Beijing, China, ²National Energy R&D Center for Non-food Biomass, Beijing Forestry University, Beijing, China, ³National Innovation Alliance of Sapindus Industry, Beijing Forestry University, Beijing, China, ⁴Bioinformatics Analysis Department, Hangzhou KaiTai Biotechnology Co., Ltd, Hangzhou, Zhejiang, China, ⁵Planning and Design Institute of Forest Products Industry, National Forestry and Grassland Administration, Beijing, China, ⁶Research and Development Department, Yuanhua Forestry Biological Technology Co., Ltd., Sanming, Fujian, China, ⁷Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, Beijing Forestry University, Beijing, China

Triterpenoid saponin are important secondary metabolites and bioactive constituents of soapberry (*Sapindus mukorossi* Gaertn.) and are widely used in medicine and toiletry products. However, little is known about the roles of miRNAs in the regulation of triterpenoid saponin biosynthesis in soapberry. In this study, a total of 3036 miRNAs were identified, of which 1372 miRNAs were differentially expressed at different stages of pericarp development. Important KEGG pathways, such as terpenoid backbone biosynthesis, sesquiterpenoid and triterpenoid biosynthesis, and basal transcription factors were highlighted, as well the roles of some key miRNAs, such as ath-miR5021, han-miR3630-3p, and ppe-miR858, which may play important roles in regulating triterpenoid saponin biosynthesis. In addition, 58 miRNAs might participate in saponin biosynthesis pathways by predicting the targets of those miRNAs to 53 saponin biosynthesis structural genes. And 75 miRNAs were identified to potentially play vital role in saponin accumulation by targeting transcript factor genes, bHLH, bZIP, ERF, MYB, and WRKY, respectively, which are candidate regulatory genes in the pathway of saponin biosynthesis. The results of weighted gene coexpression network analysis (WGCNA) suggested that two saponin-specific miRNA modules and 10 hub miRNAs may participate in saponin biosynthesis. Furthermore, multiple miRNA-mRNA regulatory networks potentially involved in saponin biosynthesis were generated, e.g.,

ath-miR5021–*SmIDI2/SmGPS5/SmbAS1/SmCYP71D-3/SmUGT74G-2*, han-miR3630-3p–*SmCYP71A-14/SmbHLH54/SmMYB135/SmWRKY32*, and ppe-miR858–*SmMYB5/SmMYB32*. qRT-PCR analysis validated the expression patterns of nine miRNAs and 12 corresponding target genes. This study represents the first comprehensive analysis of miRNAs in soapberry and lays the foundation for further understanding of miRNA-based regulation in triterpenoid saponin biosynthesis.

KEYWORDS

Sapindus mukorossi, miRNA, triterpenoid saponin, biosynthesis, coexpression network

Introduction

Soapberry (*Sapindus mukorossi* Gaertn.) is an important economic agricultural product, which is widely used in biomedicine, bioenergy, and cosmetics (Zhao et al., 2019). The whole plant, including its fruit, root, bark, and leaves, has been used in traditional medicine in China (Xu et al., 2018). The crude protein content of soapberry is about 22%, which is comparable with protein-rich vegetables such as *Phaseolus legumin* and the pulses seeds meal proteins. And the essential amino acids in its globulin and albumin could meet the needs of infants aged 2–5 years recommended by FAO/WHO. These make the proteins from soapberry seeds potentially a source of protein nutrition (Yin et al., 2011). And the kernels of soapberry contain about 40% fatty acids and could be used to produce biodiesel and high-grade lubricants (Sun et al., 2017). In addition, previous studies indicated that soapberry accumulates up to 27% (w/w pericarp dry weight) triterpenoid saponins (Sun et al., 2017). Triterpenoid saponins consist of non-polar aglycones (triterpene) and one or more glycone (monosaccharide) moieties, which make them hydrophilic and hydrophobic. These properties make soapberry saponins excellent emulsifiers and foaming agents, with antibacterial, antitumor, hepatoprotective, antihyperglycemic, antidyslipidemic, and insecticidal activities (Upadhyay and Singh, 2012). More than 70 variant triterpenoid saponins have been isolated from soapberry and identified as oleanane-type, dammarane-type, tirucullane-type and lupane-type saponins, respectively (Xu et al., 2018). The triterpenoid saponin biosynthesis pathway is divided into three stages. In the early stage, isopentenyl diphosphate and its isomer dimethylallyl diphosphate are synthesized *via* the mevalonic acid (MVA) pathway in the cytosol and the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway in the plastid. In the intermediate stage, isopentenyl diphosphate and dimethylallyl diphosphate are catalyzed by the corresponding enzymes to produce 2,3-oxidesqualene. The late stage involves the synthesis of triterpenoid saponins from 2,3-oxidesqualene (Augustin et al., 2011; Zhao and Li, 2018; Xu et al., 2021a).

MicroRNAs (miRNAs) are a class of endogenous non-coding small RNAs (sRNAs), approximately 18–25 nucleotides in length, that perform vital regulatory roles in numerous biological and metabolic processes not only in animals, but also in plants (Song et al., 2019). In plants, miRNAs regulate target gene expression mainly *via* direct endonucleolytic cleavage or translational inhibition at the posttranscriptional level (Bartel, 2004). Studies have shown that miRNAs are involved in various biological processes in plants, including development, biotic and abiotic stress responses, pathogen defense, and the biosynthesis of secondary metabolites (Bologna and Voinnet, 2014; Ye et al., 2020). miRNAs can regulate the biosynthesis of triterpenoid saponin by directly targeting genes encoding biosynthetic enzymes, TFs, or other regulators. For example, miR854e in *Panax ginseng* was shown to target farnesyl pyrophosphate synthetase (*FPS*) in ginsenoside biosynthesis, while miR1439b and miR1439h were shown to target β -amyrin synthase (β -AS) (Mathiyalagan et al., 2013). Moreover, Wei et al. (2015) and Kajal and Singh (2017) identified several miRNAs involved in the triterpenoid saponin biosynthesis pathway in *Panax notoginseng* and *Chlorophytum borivillianum* by high-throughput sequencing. Although miRNAs related to triterpene biosynthesis have been identified, there have been no reports to date of miRNAs in soapberry.

In this study, to comprehensively investigate the regulatory mechanisms of triterpenoid saponin biosynthesis in soapberry, 24 sRNA libraries were constructed from soapberry pericarps at eight stages of development. We used bioinformatics tools to predict known and novel miRNAs and their potential targets. Further, the miRNA expression profiles were determined during the pericarp development process, and differential expression analysis and weighted gene coexpression network analysis (WGCNA) were performed to identify the miRNAs associated with saponin biosynthesis. Furthermore, miRNA coexpression networks and miRNA–mRNA interaction networks were constructed to characterize the key regulators in the saponin

biosynthesis pathway. This study identified many miRNAs and their target genes, which not only provided the first comprehensive miRNA expression profile, but also built a foundation for future studies of miRNA-mediated regulation of saponin biosynthesis in soapberry.

Materials and methods

Plant materials

Soapberry fruits were harvested from Jianning County, Fujian Province, China (27°06'N, 117°25'E) at eight developmental stages (S1–8), covering the entire fruit development process. Three biological replicates were taken at each stage, for a total of 24 samples. For sRNA sequencing, pericarp samples were collected from three trees. The details of these materials were described in our previous report (Zhao et al., 2019; Xu et al., 2021b; Xu et al., 2022).

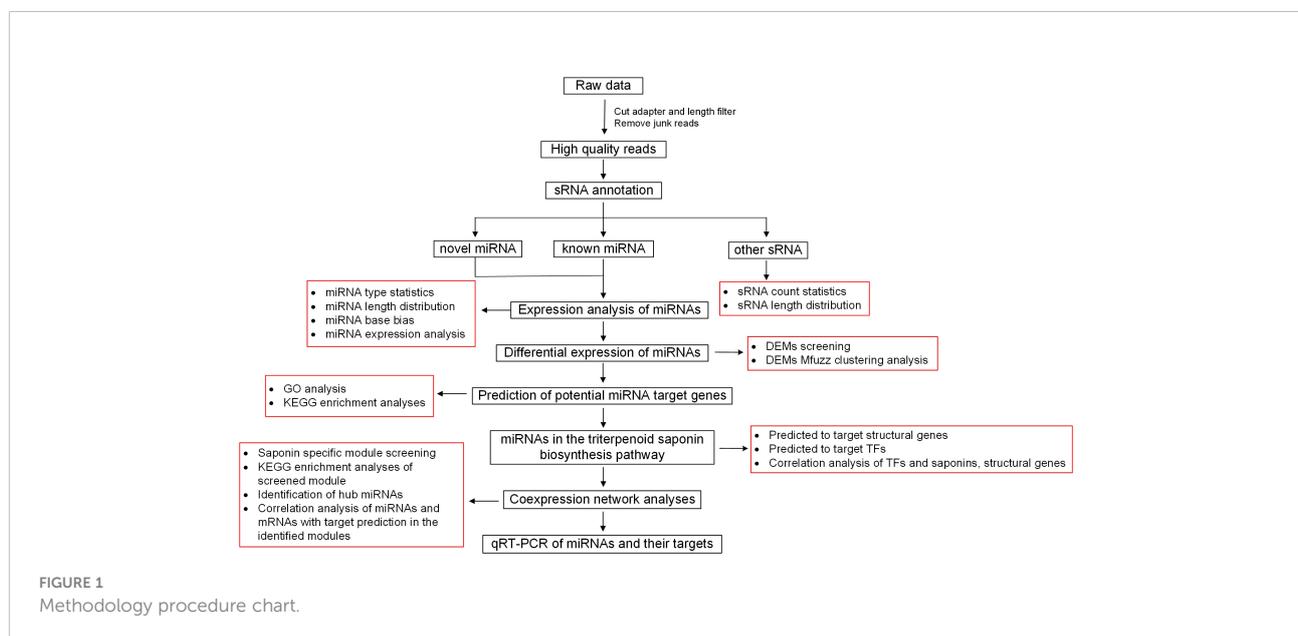
sRNA library construction and high-throughput sequencing

Total RNA was extracted from the 24 samples using the FinePure Plant RNA Kit (Genfine Biotech, Beijing, China) in accordance with the manufacturer’s protocol. RNA quality and quantity were examined using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and Nanodrop™ 2000 spectrophotometer (Thermo Fisher Scientific, Fremont, CA, USA). Sequencing libraries were constructed using the NEBNext® Multiplex Small RNA Library Prep Set for Illumina® (New England

Biolabs, Ipswich, MA, USA) in accordance with the manufacturer’s protocol, and single-end sequencing was then performed on the Illumina HiSeq X Ten System (Illumina, San Diego, CA, USA) at ORI-BIO Co., Ltd. (Beijing, China). The sRNA dataset was deposited in the National Center for Biotechnology Information Databank at the Sequence Read Archive (SRA) database under accession numbers PRJNA784546, and includes 24 accession items (SRR17065996-SRR17065973).

Identification of miRNAs and prediction of their targets

The procedure for methodology is summarized in Figure 1. Briefly, after removal of adapter sequences and low-quality reads, the clean 15–34 nt sRNAs were mapped to the soapberry reference genome. The mapped sRNAs were then compared with sequences in the Rfam database (<http://pfam.xfam.org/>), allowing no more than two mismatches and 15 valid alignments to remove RNA categories including mRNA, tRNA, rRNA, snRNA, and snoRNA (Burge et al., 2013). The known miRNAs from mapped sRNAs were identified by BLASTN (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>, Version 2.2.28+) search against the miRBase database (<http://www.mirbase.org/>) (Kozomara and Griffiths-Jones, 2014). To identify novel miRNAs, Bowtie2 (<http://bowtie-bio.sourceforge.net/>, Version 2.2.4) was used to map reads to the soapberry reference genome, and then the MIREAP program (<http://mireap.sourceforge.net/>, Version 0.2) was used to obtain all candidate precursors with hairpin-like structures (Langmead and Salzberg, 2012). Target genes of miRNAs, including known and novel miRNAs, were predicted using the MiRanda (<http://www.microrna.org/microrna/>, Version 3.3a; score=150, energy=-14)



(Enright et al., 2003) and RNAhybrid (<http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/>, Version 2.1.2; E -value $< 1 \times 10^{-30}$) (Rehmsmeier et al., 2004) programs from the soapberry transcriptome derived from the same eight developmental stages of soapberry pericarp, and the intersections of the predicted results were used as the final target gene prediction results. Target gene prediction was performed using the default parameters. In addition, annotation information for these target genes was extracted from the Soapberry Transcriptome Annotation Project (Xu et al., 2022). The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases were used for gene annotation using emapper.py (EggNOG v5.0: <http://eggno5.embl.de/#/app/home>) (Hernández-Plaza et al., 2022). In addition, GO and KEGG enrichment analyses were conducted using the clusterProfiler package in R.

Differential expression analysis of miRNAs

Quantification of miRNA expression levels was based on reads per million (RPM): $\text{RPM} = (\text{mapped read counts}/\text{total read counts}) \times 1000000$. Differential expression analysis of miRNAs in samples from two pericarp development stages was performed using DESeq2 (Love et al., 2014). The P -values were calculated by displacement test and then corrected using the false discovery rate to calculate the Q -values. miRNAs with a Q -value < 0.05 and $|\log_2(\text{fold change})| > 1$ were identified as differentially expressed miRNAs (DEMs). Then, the online tool MISA-web was used to predict the simple sequence repeats (SSRs) in pre-mature sequences of identified DEMs using default parameters (Beier et al., 2017).

miRNA coexpression network analysis

Coexpression networks were constructed using the WGCNA package in R (Langfelder and Horvath, 2008) and miRNA expression data from 24 libraries of soapberry. miRNAs with RPM values > 0 in more than two duplications were chosen for the WGCNA (Hollender et al., 2014). A gene expression adjacency matrix was generated to analyze the network topology with a soft threshold of $\beta = 4$, mergeCutHeight of 0.30, and minModuleSize of 30. The blockwiseModules were used with the default settings to obtain the modules. The first principal component module characteristic genes (MEs) were calculated to assess the expression levels of the gene modules. Then, module-trait associations were estimated by calculating the Pearson's correlations between the MEs and the differentially accumulated saponin (DAS) content obtained from the pericarp materials from the same eight developmental stages determined previously by our group previously (Table S1). Integrative analysis was also performed to determine correlations between miRNA MEs and mRNA MEs from the same samples (Xu et al.,

2022). Subsequently, we also performed Pearson's correlation analysis to detect negatively associated miRNA-mRNA pairs. Cytoscape software (v. 3.8.2) was used to generate the networks (Shannon et al., 2003).

qRT-PCR based expression analysis of miRNAs and their predicted targets

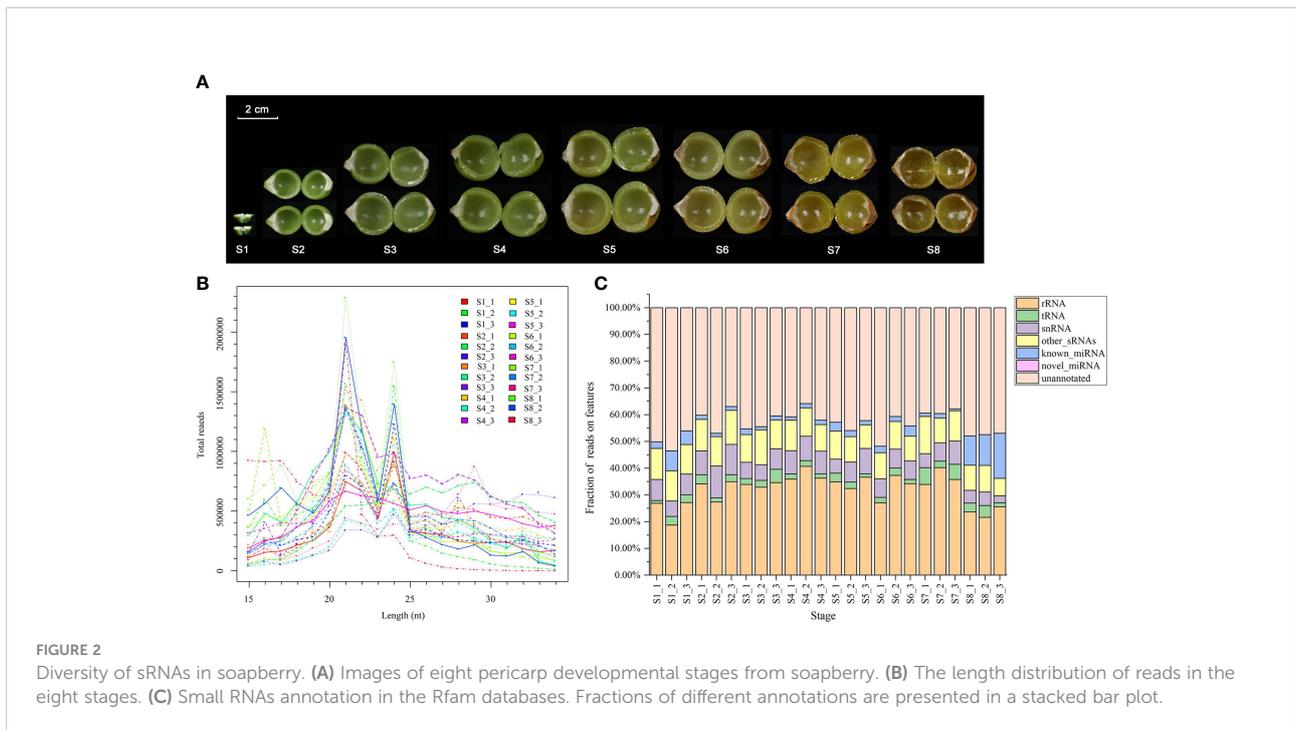
qRT-PCR was performed to validate the miRNAs identified using deep sequencing and to analyze their expression patterns. Total RNA was prepared as described above. For miRNA quantification, total RNA was polyadenylated and reverse transcribed into cDNA using *TransScript*[®] miRNA RT Enzyme Mix and 2×TS miRNA Reaction Mix (TransGen Biotech, Beijing, China). To quantify the expression of specific miRNA sequences, the cDNA samples were prepared with PerfectStart[™] Green qPCR SuperMix (TransGen Biotech), ROX Reference Dye I (Thermo Fisher Scientific), Universal miRNA qPCR Primer (TransGen Biotech), and our miRNA-specific 5' primer(s) and subjected to qPCR using the ABI StepOnePlus Real-Time PCR system (Applied Biosystems, Foster City, CA, USA). For target mRNA expression analysis, reverse transcription was performed using the Goldenstar RT6 cDNA Synthesis Kit ver. 2 (Tsingke, Beijing, China), followed by qPCR using TaKaRa TB Green[™] Premix Ex Taq[™] (Tli RNaseH Plus; TaKaRa, Kyoto, Japan). Soapberry snRNA U6 (*Samuk01G0008400*) and *SmACT* (*Samuk13G0061200*) were used for normalization of the relative expression of miRNAs and mRNAs, respectively. The primers were designed using Oligo 7.0 software (Table S2 and S3) and synthesized by Tsingke Biotech. Three biological and technical replicates were performed. The expression levels of miRNAs and their target genes were determined by calculating the fold change in selected stages of pericarp samples relative to stage S1 using the $2^{-\Delta\Delta C_t}$ method.

Results

High-throughput sequencing of sRNAs

Twenty-four sRNA libraries were constructed from soapberry pericarp samples and sequenced (Figure 2A). The raw reads yielded by deep sequencing of the 24 libraries ranged from 8.724 million (M) to 18.286 M (Table S4). After removing low-quality sequences and adaptor sequences, we obtained 5.576 M–14.134 M clean reads with average lengths of 19.2 nt–27.6 nt (Table S4). Q20 and Q30 base percentages of each library were $> 99.0\%$ (Table S4). The data obtained by sequencing were reliable and could be used for further experimental analyses.

The length distributions (15–34 nt) of all sRNA in the 24 samples revealed a major peak at 21 nt and a minor peak at 24 nt (Figure 2B). Unannotated sRNAs (35.85%–53.54%) and rRNAs (18.67%–40.65%) accounted for a considerable proportion of the



total reads. A total of 57462–1359491 (0.72%–16.83%) sequences were produced as candidate miRNAs from the 24 libraries, including 57420–1358629 known miRNAs and 19–862 novel miRNAs (Figure 2C).

Identification of miRNAs in soapberry

The size distribution of the predicted miRNAs differed among the 24 pericarp libraries, but the majority were between 15 and 34 nt in length, with 21 nt being the most common size followed by 20 nt (Figure S1A). Nucleotide bias analysis showed that these candidate miRNAs had a strong preference for adenosine (A) at position 15, for guanosine (G) at position 4 and 9, for cytidine (C) at position 19, and for uridine (U) at position 1 (Figure S1B). In total, 3036 miRNAs (2221 known miRNAs and 815 novel miRNAs) were predicted (Table S5; Figure S2A). Most of the identified miRNAs showed high degrees of identity among 60 plant species, including *Glycine max*, *Oryza sativa*, *Brachypodium distachyon*, *Medicago truncatula*, *Gossypium raimondii*, and *Arabidopsis lyrata* (Figure S2B). Particularly, 190 miRNAs identified in soapberry showed homology to those of *G. max*. Based on the sequence similarity, we grouped these known miRNAs into 1249 miRNA families. The largest families identified were miR156 with 59 members, followed by miR166 and miR169, with 44 and 41 members, respectively. Of the remaining 1246 families, 183 families were represented by 2–39 members, while 1063 were represented by only a single member (Table S6). The size of a miRNA family may represent its function.

The average expression level of known miRNAs (449.62 RPM/miRNA) was higher than that of novel miRNAs (1.70 RPM/miRNA), indicating that most novel miRNAs are weakly expressed and imprecisely processed (Cuperus et al., 2011). The expression levels of the known miRNAs markedly from < 5 to > 650000 RPM. In addition, 1266 known miRNA families were expressed during at least one pericarp development stage (Table S7), of which miR159 was the most abundant miRNA family. The expression levels of a few other miRNAs, including miR166, miR2916, miR482, miR396, miR162, miR894, and miR398, were also very high in both stages examined (RPM >2000). In addition, 56.99% of known miRNA families showed RPM values < 100. Meanwhile, some members of the same miRNA family also differed in expression level. For example, the RPM ranges of pta-miR159a, bdi-miR159b-3p.1, and lus-miR159c, belonging to the miR159 family, were 1.93–69.16, from 14.28–9215.03, and from 311.52–30461.96, respectively, in the eight development stages, indicating functional divergence of the different development stages of soapberry pericarp. In addition, the expression levels of novel miRNAs were low, and only novel-m1318-5p and novel-m0486-5p had RPM values > 100.

Differential expression of miRNAs in soapberry

A total of 1372 DEMs were identified by pairwise comparisons, comprising 1171 known miRNAs (belong to 691 families) and 201 novel miRNAs. These DEMs belonged mainly to the miR156, miR166, miR159, miR396, and miR169 families.

Compared with the previous stage, the number of downregulated miRNAs was more than that of upregulated miRNAs in each pairwise comparison (Figure S3). The 1372 DEMs were divided into 12 groups by Mfuzz clustering analysis based on their expression levels (Figure S4, Table S8). Among them, most miRNAs were preferentially expressed at a certain stage, e.g., cluster 2 (147 miRNAs) at stage S4 and cluster 3 (161 miRNAs) at stage S6. In addition, the expression of miRNAs in other clusters varied significantly among the different stages, e.g., cluster 1 (61 miRNAs) at stages S6 and S7 and cluster 6 (67 miRNAs) at stages S2 and S4 (Figure S4). Among the 1372 DEMs, only 12 miRNAs were predicted to contain SSR loci in their pre-mature sequences (Table S9).

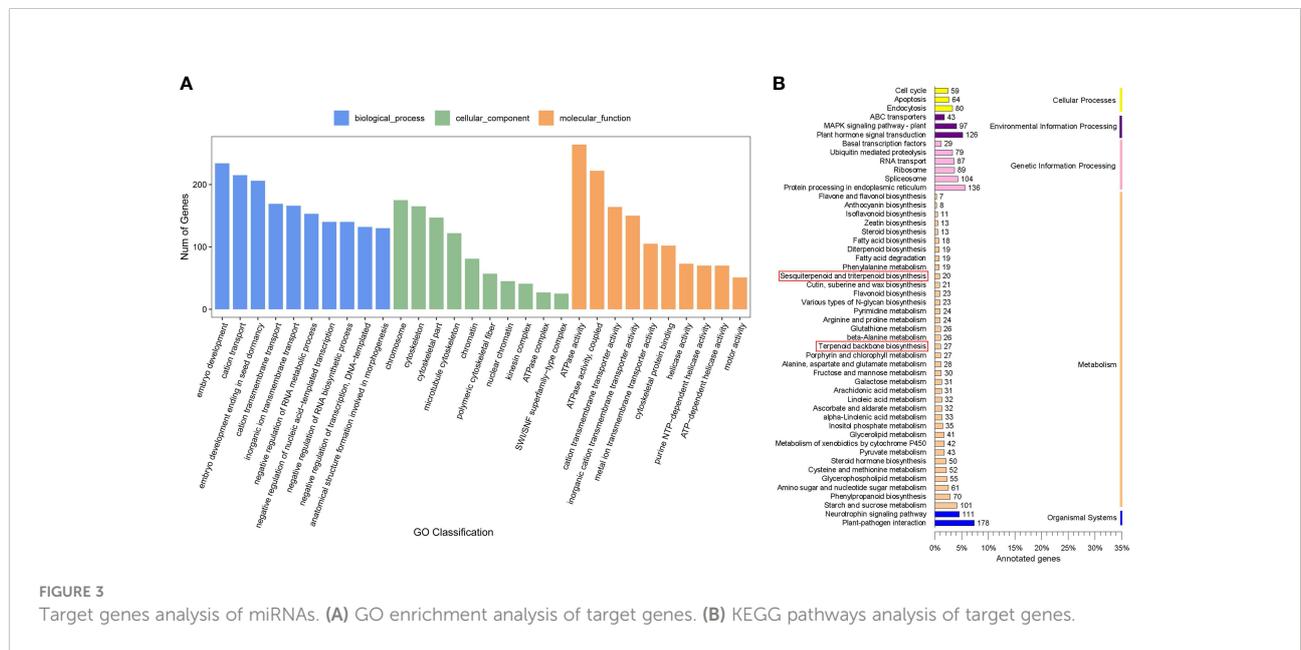
Prediction and annotation of potential miRNA target genes

A total of 12750 genes were predicted to be the targets of 2188 miRNAs among the 3036 identified miRNAs, with an average of 5.83 targets per miRNA. These target genes were further subjected to GO analysis and KEGG pathway analyses. The target genes were classified into 2438, 392, and 653 GO terms for biological process, cellular component, and molecular function ontology, respectively. The main biological processes of miRNA target genes were cation transport, cation transmembrane transport, and negative regulation of RNA metabolic process. The molecular functions associated with the predicted miRNA targets were ATPase activity and cation transmembrane transporter activity, and chromosome was the main cellular component of miRNA target genes (Figure 3A). In addition, GO terms related to UDP-

glycosyltransferase activity, UDP-glycosyltransferase activity, and transcription factor binding were also associated with the target genes. These observations suggested that the miRNA target genes related to transmembrane transport, UDP-glycosyltransferase activity, and transcription factor binding may play important roles in triterpenoid saponin biosynthesis. KEGG pathway enrichment analysis suggested that 47 mRNAs of miRNA target genes may function in terpenoid backbone biosynthesis and sesquiterpenoid and triterpenoid biosynthesis (Figure 3B). Among them, 46 mRNAs belonged to the target genes of DEMs such as ath-miR5021, han-miR3630-3p, and ppe-miR858, etc. (Figure S5). Furthermore, a variety of pathways associated with saponin transcription and transport, such as ABC transporters, plant hormone signal transduction, and basal transcription factors, were also found in this study (Figure S5), suggesting that some of the DEMs may play important roles in saponin biosynthesis in soapberry.

Involvement of miRNAs in the triterpenoid saponin biosynthesis pathway

In the present study, 53 structural genes from 13 gene families involved in the saponin biosynthesis pathway were screened from miRNA target genes (Figure S6, Table S10). These candidate genes were predicted as targets of 58 miRNAs, totaling 91 miRNA-miRNA pairs (Table S11). Among them, 32 and 10 miRNAs were found to regulate terpenoid backbone biosynthesis and sesquiterpenoid and triterpenoid biosynthesis pathways, respectively (Figure 4). In addition, 26 miRNAs were predicted to



target CYP450s and UGTs. Our results suggested that *csi-miR3951*, *ath-miR5021*, and *ptc-miR7817b*, each of which targets more than five saponin biosynthesis candidate genes in the miRNA–mRNA network (Figure S6), may be important regulators.

Furthermore, many TFs from bHLHs, bZIPs, ERFs, MYBs, and WRKYs, which are considered to play important roles in saponin biosynthesis, were also identified among the miRNA target genes (Figure S7, Table S10). In this study, 659 miRNA–TF pairs were obtained, including 314 miRNAs and 315 target TFs (including 43 bHLHs, 103 WRKYs, 26 bZIPs, 39 ERFs, and 104 MYBs) (Table S11). In addition, *ppe-miR858*, *ptc-miR7817b*, *ath-miR5021*, and *csi-miR3951* were suggested to

be important regulators, each of which targeted more than 20 TFs in the miRNA–mRNA network (Figure S7).

According to the FPKM and DAS content in the eight developmental stages of pericarp of soapberry, the correlation between the expression level of the selected TFs and structural genes or DAS content was analyzed. The 18 TFs which included 1 bHLH, 4 bZIPs, 4 ERFs, 4 MYBs, and 5 WRKYs ($r > 0.8$ or $r < 0.8$, $P < 0.05$; Table S12) were significantly correlated with DAS content. In addition, 63 TFs (11 bHLHs, 4 bZIPs, 7 ERFs, 22 MYBs, and 19 WRKYs) associated with the regulation of 30 structural genes ($r > 0.8$ or $r < 0.8$, $P < 0.05$; Table S12). After taking out the redundancy, 75 TFs that were associated with the biosynthesis of saponin.

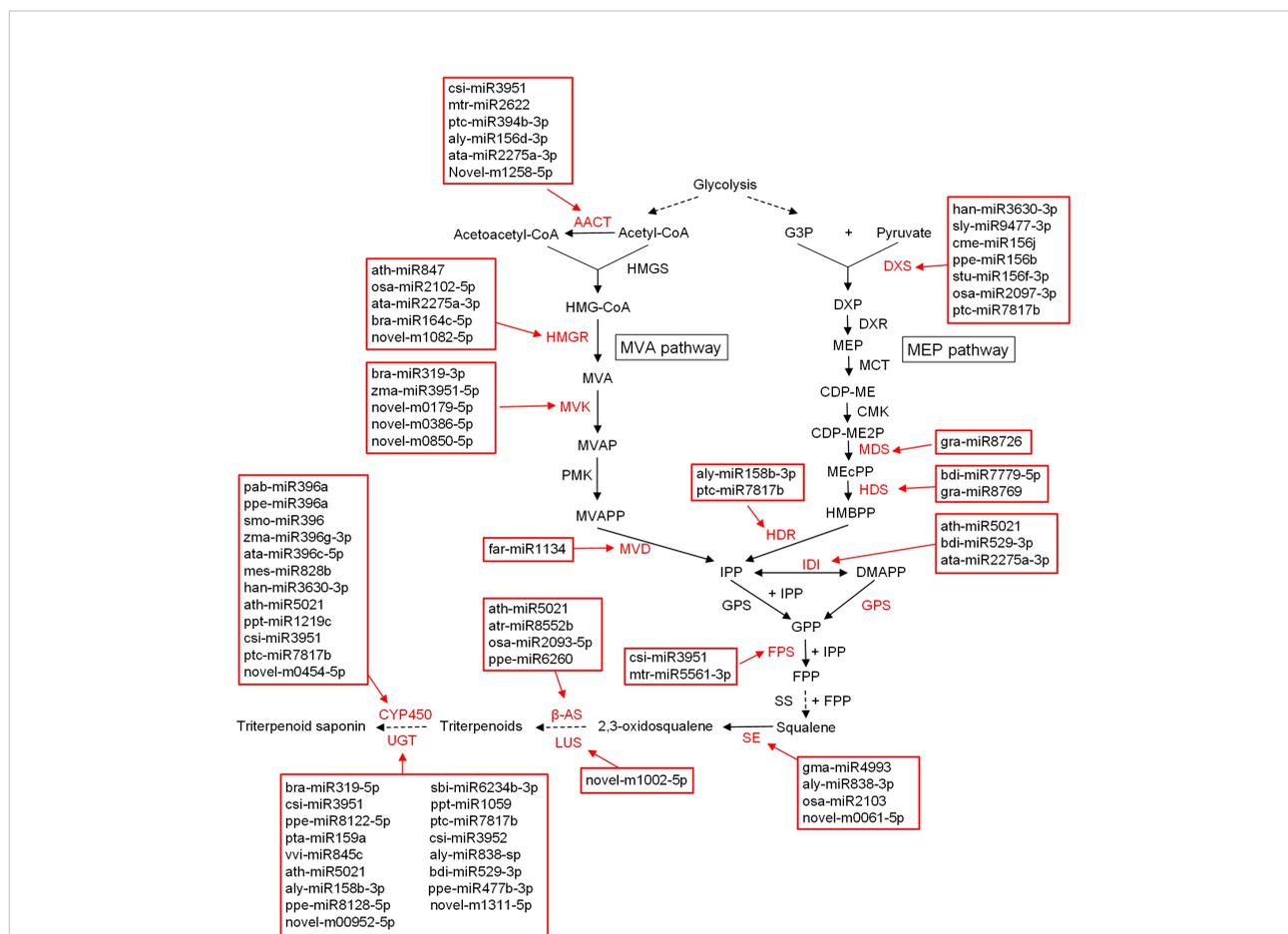


FIGURE 4

Hypothetical model of miRNAs regulation in triterpenoid biosynthesis pathway. AACT, acetyl-CoA C-acetyltransferase; HMGS, 3-hydroxy-3-methylglutaryl-CoA synthase; HMG-CoA, 3-hydroxy-3-methylglutaryl-CoA; HMGR, 3-hydroxy-3-methylglutaryl-CoA reductase; MVA, mevalonate; MVK, mevalolate kinase; MVAP, 5-methylvalic acid phosphoric; PMK, phosphoric acid valate kinase; MVAPP, 5-methylvalic acid pyrophosphoric acid; MVD, 5-hydroxyvalerate decarboxylase; G3P, 3-glyceraldehyde phosphate; DXS, 1-deoxy-D-xylulose-5-phosphate synthetase; DXP, 1-deoxy-D-xylulose-5-phosphate; DXR, 1-deoxy-D-xylulose-5-phosphate reductoisomerase; MEP, 2-C-methyl-D-erythritol-4-phosphATE; MCT, 4-cytidine diphosphate-2-C-methyl-D-erythritol synthetase; CDP-ME, 4-cytidine diphosphate-2-C-methyl-D-erythritol; CMK, 4-cytidine diphosphate-2-C-methyl-D-erythritol kinase; CDP-ME2P, 4-cytidine diphosphate-2-C-methyl-D-erythritol-2-phosphate; MDS, 2-C-methyl-D-erythritol-2,4-cyclic phosphate synthetase; MEcPP, 2-C-methyl-D-erythritol-2,4-cyclic phosphate; HDS, 4-hydroxy-3-methylbut-2-enyl diphosphate synthase; HMBPP, 4-hydroxy-3-methylbut-2-enyl pyrophosphate; HDR, 4-hydroxy-3-methylbut-2-enyl diphosphate reductases; IDI, isopentenyl pyrophosphate isomerase; IPP, isopentenyl pyrophosphate; DMAPP, dimethylallyl pyrophosphate; GPP, geranyl pyrophosphate; GPS, geranylpyrophosphate synthetase; FPS, farnesyl pyrophosphate synthetase; FPP, farnesyl pyrophosphate; SS, squalene epoxidase; β-AS, beta-amyrin synthase; LUS, lupeol synthase; CYP450, cytochrome P450; UGT, uridine diphosphate-dependent glycosyltransferases.

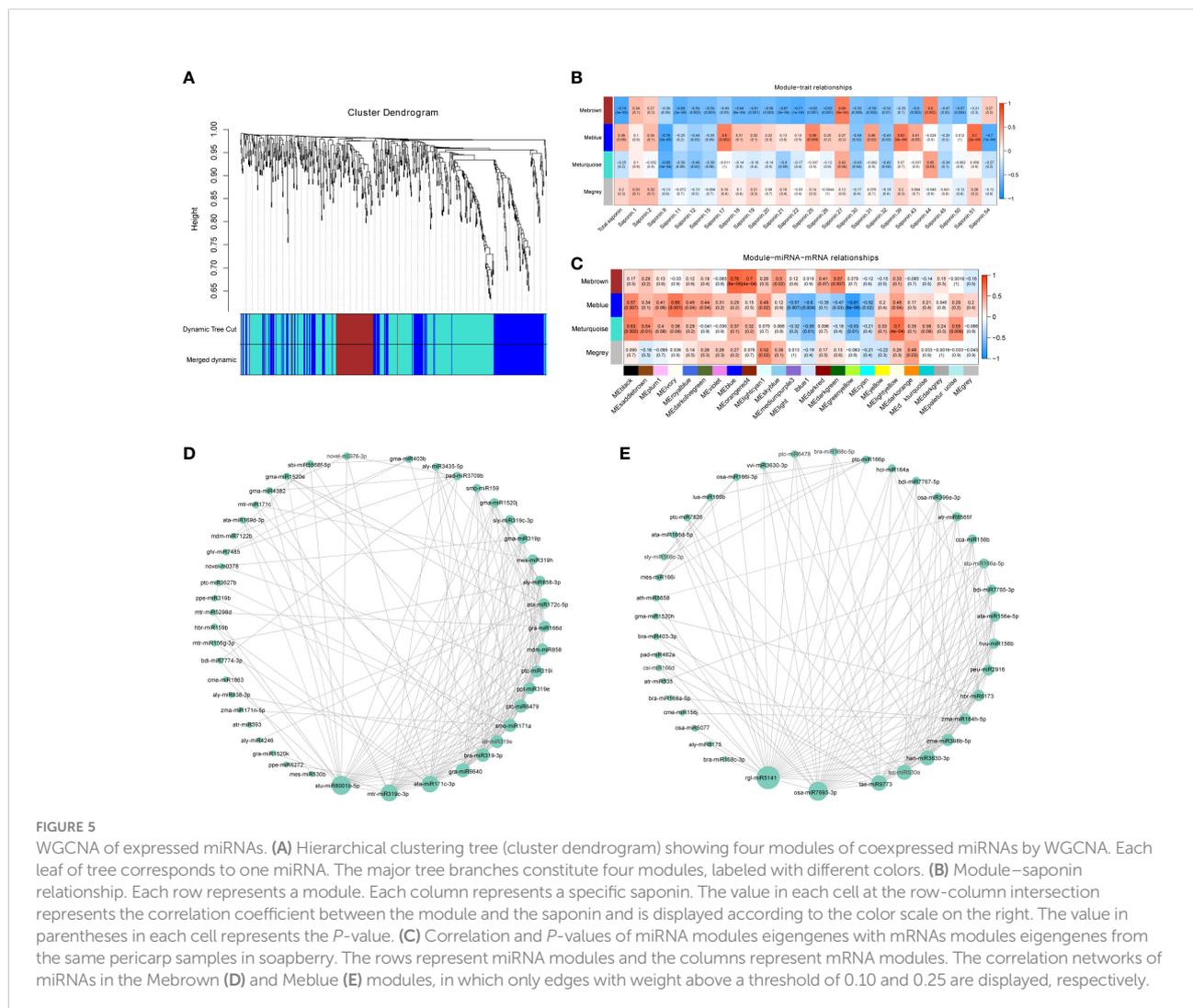
Coexpression network analyses

Next, we used the miRNA gene expression profile data as the original data for WGCNA. After elimination of deletions and outlier values, 616 original miRNAs remained. These miRNAs were clustered into four tree branches, and each of which represented a module (labeled with different colors, with the grey module containing the uncorrelated genes) (Figure 5A and Table S13). The number of miRNAs (module size) ranged from 80 (Mebrown) to 325 (Meturquoise) miRNAs for each module, with an average size of 205 miRNAs (Table S13). Module-trait relationship analysis showed that the Mebrown module with 80 miRNAs was highly negatively correlated with total saponin and 15 other saponins, including the major saponins: saponin 11, saponin 19, saponin 20, and saponin 22 (Figure 5B, Table S1). The Meblue module contained 204 miRNAs and showed a significant association with saponin 9, saponin 17, saponin 25, saponin 39, saponin 51, and saponin 54 (Figure 5B). In addition,

as miRNAs can negatively regulate mRNA abundance, miRNA MEs were correlated with mRNA MEs from the same pericarp samples in soapberry. According to the results of our previous mRNA module-saponin correlation analysis, the MEblue and MEgreenyellow modules were significantly correlated with most saponins. As shown in Figure 5C, the mRNA module (MEgreenyellow) was highly negatively correlated to one miRNA module (Mebrown). Thus, two miRNA modules (Mebrown and Meblue) were finally extracted for further study.

KEGG enrichment analysis of target genes of miRNAs in the identified modules

There were 1432 and 3128 predicted target genes in the mRNA transcriptomes of the same pericarp samples in the miRNA Mebrown and Meblue modules, respectively. The results of



KEGG enrichment analysis showed that transcription machinery and glycosyltransferases were significantly enriched in the Mebrown module (Figure S8A). In addition, KEGG pathways related to sesquiterpenoid and triterpenoid biosynthesis, terpenoid backbone biosynthesis, and cytochrome P450 were also involved. In the Meblue module, the target genes were significantly associated with messenger RNA biogenesis, transcription machinery, and ABC transporters (Figure S8B). In addition, nine saponin biosynthetic structural genes, including 3-hydroxy-3-methylglutaryl-CoA reductase (*HMGR*), mevalonate kinase (*MVK*), 1-deoxy-D-xylulose-5-phosphate synthetase (*DXS*), 2-C-methyl-D-erythritol-2,4-cyclic phosphate synthetase (*MDS*), isopentenyl pyrophosphate isomerase (*IDI*), *FPS*, and β -AS, were also associated with terpenoid backbone biosynthesis and sesquiterpenoid and triterpenoid biosynthesis (Figure S8B). Another six saponin biosynthesis candidate genes, including CYP450s and UGTs, were also found among the target genes of miRNAs in the Meblue module.

Identification of hub miRNAs

In the Mebrown module, 45 miRNAs showed high correlations (edge weight ≥ 0.10), and *stu-miR8001b-5p*, *mtr-miR319c-3p*, *ata-miR171c-3p*, *gra-miR8640*, and *bra-miR319-3p* were identified as the top five hub miRNAs (Figure 5D). A total of 59 genes, including 1 structural gene and 7 TFs, were predicted to be targets of these hub miRNAs (Table S14). Among them, *bra-miR319-3p* was predicted to target a mevalonate kinase (*SmMVK2*) in the mevalonic acid pathway and a TFs of the WRKY family (*SmWRKY101*).

The network diagram of the Meblue module was composed of 39 coexpressed miRNAs (edge weight ≥ 0.25), of which *rgl-miR5141*, *osa-miR7693-3p*, *tae-miR9773*, *han-miR3630-3p*, and *tcc-miR530a* had the highest degrees of connectivity (Figure 5E). In total, 257 genes, including 4 structural genes and 17 TFs, were predicted to be targets of these five miRNAs (Table S14). Of these, *han-miR3630-3p* was predicted to target one *DXS*, three CYP450s, and one *bHLH*, seven WRKYs, and two MYBs. In addition, *tcc-miR530a* targeted *SmbHLH29*, a bHLH family TF.

Most of the hub miRNAs in the Mebrown and Meblue modules identified in the WGCNA were also identified as DEMs among the various developmental stages. These results suggested that WGCNA and differential expression analysis identified not only similar biological processes associated with saponin biosynthesis but also the same miRNAs.

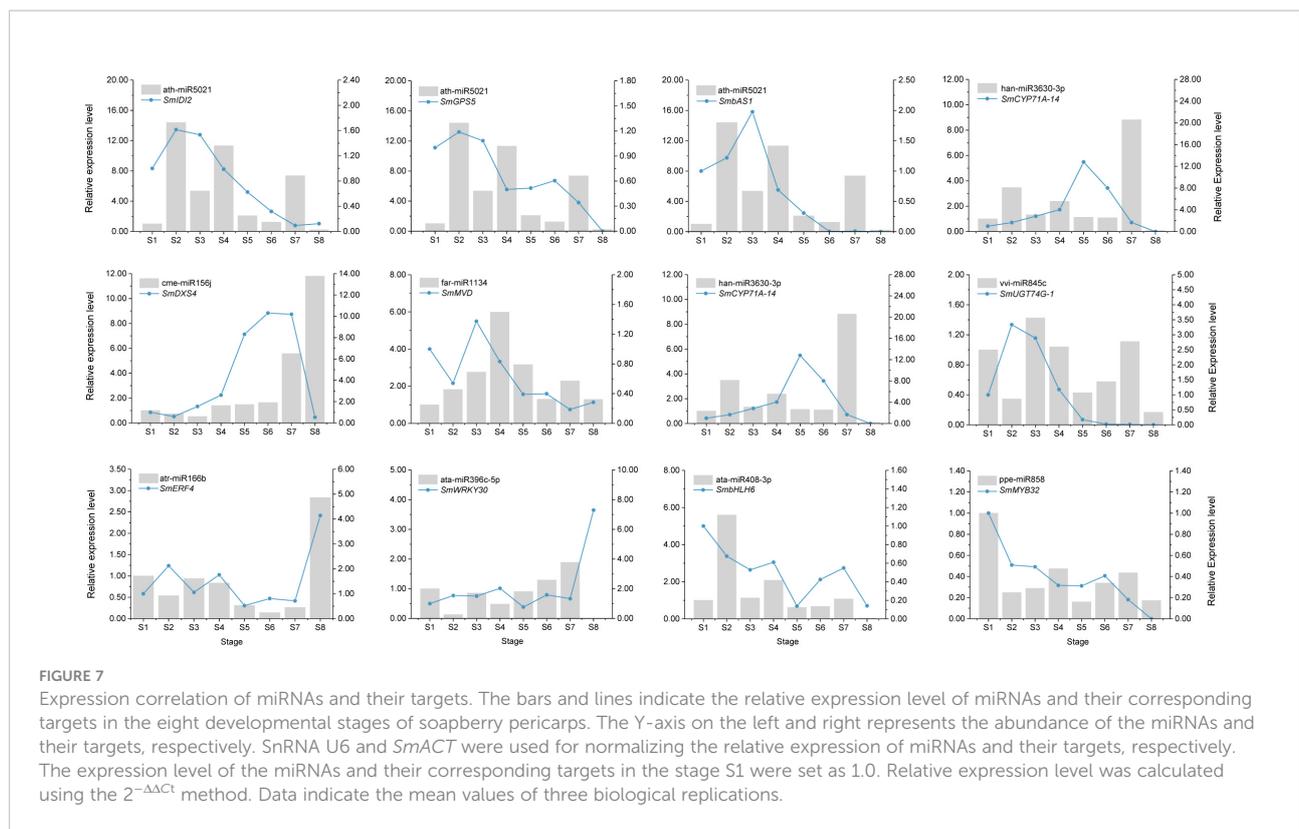
Integration of negative correlations between miRNAs and mRNAs with target prediction in the identified modules

The correlations between mRNA and miRNA expression levels were analyzed according to the transcriptomic data from

the eight developmental stages of soapberry pericarp ($r < -0.4$, $P < 0.05$). In total, 74 and 94 significant negative miRNA–target gene pairs were identified in modules Mebrown and Meblue, respectively (Table S15). In module Mebrown, a total of seven conserved miRNAs were predicted to target 74 genes (Figure 6A). In addition, *aly-miR158b-3p* had the largest number of target genes, and followed by *atr-miR8577*, *aly-miR838-3p*, and *far-miR1134*. Two TFs that may be related to saponins biosynthesis, *SmERF39* and *SmMYB52*, were the targets of *atr-miR8577* and *ppt-miR414*, respectively. In addition, 86 genes were negatively correlated with 34 miRNAs in the Meblue module (Figure 6B). Of these, the TFs from the TCP, Trihelix, C3H, SBP, and Nin-like families, were targets of *ata-miR156e-5p*, *ata-miR5168-3p*, *ath-miR5021*, *cme-miR156g*, *lus-miR166b*, and *mtr-miR5291c*. In addition to TFs, miRNA has also been found to regulate other regulators (e.g., transcriptional regulators (TRs), protein kinases (PKs), transporters), enzymes, proteins, and other genes with unknown functions in the identified modules. These genes may also play important roles in the biosynthesis of triterpenoid saponin in soapberry, and further studies are therefore required.

qRT-PCR analysis of the miRNAs and their targets

From the high-throughput sequencing results, we randomly selected nine miRNAs and their corresponding 12 target genes (8 structural genes and 4 TFs) for qRT-PCR. The results indicated that the relative expression levels of these miRNAs at eight developmental stages were essentially consistent with the sRNA-Seq data (Figure S9A). In addition, the expression patterns of 12 target genes also showed strong correlations between the qRT-PCR and RNA-Seq data (Figure S9B). These results further suggested that our high-throughput sequencing data were reliable for evaluating miRNA–mRNA regulatory network in soapberry. Furthermore, the differential expression trends of these nine miRNAs were measured by qRT-PCR. The expression levels of two miRNAs, *ath-miR5021* and *ata-miR408-3*, peaked at stage S2, which two other miRNAs (*atr-miR166b* and *cme-miR156j*) showed the highest expression at stage S8. These results suggest that different miRNAs may regulate saponin biosynthesis at different stages. Simultaneously, the expression patterns of these miRNAs were negatively correlated with their corresponding target genes at most pericarp developmental stages (Figure 7). For example, the level of *cme-miR156j* expression was higher at stage S8, followed by stages S7 and S6. The corresponding target gene, *SmDXS4* showed the reverse trend. The expression levels of *ata-miR396c-5p* and *SmWRKY30* also showed negative correlations at most pericarp developmental stages. However, in the case of these 12 miRNA–mRNA pairs, the expression of the target gene was partially positively correlated with the expression of the miRNA at different stages (Figure 7). The relative changes in the



miRNAs) were identified. The reads of known miRNAs ranged from a few to hundreds of thousands, similar to previous studies in *Solanum lycopersicum* and *O. sativa* (Cheah et al., 2015; Candar-Cakir et al., 2016). In addition, 56.99% of known miRNA families exhibited low abundance (RPM < 100) in soapberry pericarp, and only a few miRNA families were highly expressed, as in the reports of other plant species, including *C. sinensis* (Guo et al., 2017), *Arachis hypogaea* (Zhao et al., 2010), and *Gossypium hirsutum* (Xie et al., 2015). The miRNA expression patterns are highly conserved in different plants, with most members of the miR156, miR159, miR166, and miR398 families being highly expressed in most plants, including *O. sativa*, *Triticum aestivum*, and *P. notoginseng* (Devi et al., 2013; Wang et al., 2014; Wei et al., 2015). We found that miR159, miR166, miR2916, miR482, miR396, miR162, miR894, and miR398 were the eight most abundantly expressed miRNA families in soapberry pericarp, similar to the results reported in other plants. Some miRNA families are monocot- or dicot- specific (Mandhan et al., 2012), as further demonstrated by the presence of miR403 (dicot-specific miRNA) in soapberry. In the present study, with pericarp development, the RPM value of miR156 increased from 175 (stage S1) to 35499 (stage S8), while the RPM value of miR172 decreased from 92 (stage S1) to 5 (stage S8) (Table S7), consistent with the previous results showing the opposite regulation of these two miRNAs (Mandhan et al., 2012).

Presence of many novel miRNAs in soapberry

Although early research showed that most plant miRNAs have been conserved throughout evolution, increasing numbers of novel miRNAs with tissue-specific, stage-specific, and low expression have been revealed by sRNA sequencing (Zhao et al., 2012). A total of 815 novel miRNAs were predicted in soapberry by bioinformatics analyses in the present study, and most of these miRNAs begin with 5'-uridine, which is consistent with typical miRNA sequences (Ye et al., 2020). As expected, most of the novel miRNAs had a low expression, in agreement with the previous results in *Arabidopsis* and *S. lycopersicum* (Fahlgren et al., 2007; Cao et al., 2014). Therefore, there are still many unannotated sequences in these 24 libraries, of which 35.85%–53.54% of the reads were unannotated. We speculate that many low-abundance non-coding RNAs remain to be discovered.

Identified differential expression of miRNA

A variety of miRNAs are involved in the biosynthesis of plant secondary metabolites. In addition, some miRNAs show significant increases or decreases in expression in major organs or at major developmental stages (Hao et al., 2012; Chen et al.,

2020). We identified 1372 DEMs at different pericarp developmental stages. Several stage-specific miRNAs were also detected in this study. We divided these 1372 DEMs into 12 clusters by Mfuzz clustering analysis, and most were preferentially expressed at one or two stages, suggesting that stage-specific expression of the DEMs reflects their specific biological functions (Figure S4). The expression of known miRNAs, such as miR156, miR854, miR172, and miR166, changed significantly during the pericarp development, and these miRNAs were predicted to target genes involved in regulation of the saponin biosynthesis pathway. These miRNAs were also identified in *P. notoginseng*, *P. ginseng*, and *C. borivilianum* (Mathiyalagan et al., 2013; Wei et al., 2015; Kajal and Singh, 2017).

miRNA targets structural genes and TFs related to saponin biosynthesis in soapberry

Identification of the target genes of the identified miRNAs will help elucidate their roles in various metabolic pathways (Samad et al., 2019). Many conserved miRNAs have the same or homologous targets in different plants, indicating their conserved function in specific biological processes (Liu et al., 2017). Singh et al. (2016) predicted that miR156 targets *DXS*, which is involved in the 2-C-methyl-D-erythritol 4-phosphate pathway and governs terpenoid biosynthesis in *Mentha* spp. We also reached a similar conclusion in soapberry. In our study, three miRNAs (cme-miR156j, ppe-miR156b, and stu-miR156f-3p) were predicted to target *SmDXS4*. qRT-PCR analyses further showed that cme-miR156j negatively regulated the expression of *SmDXS4* at most pericarp developmental stages (Figure 7). In addition, regulation of terpenoid biosynthesis by miR5021 has been observed in previous studies. For example, the regulation of *HMGR* and *IDI* by miR5021 was reported in *Xanthium strumarium* (Fan et al., 2015), and miR5021 in *Catharanthus roseus* homologous to *Arabidopsis* was predicted to target *HDS*, which is involved in terpenoid biosynthesis (Pani and Mahapatra, 2013). In the present study, ath-miR5021 was predicted to target *SmIDI1*, *SmIDI2*, *SmGPS5*, *SmbAS1*, *SmCYP71D-3*, and *SmUGT74G-2* in soapberry. Further, we also verified that ath-miR5021 was inversely correlated with *SmIDI2*, *SmGPS5*, *SmbAS1*, and *SmCYP71D-3* expression by qRT-PCR at most developmental stages (Figure 7). Taken together with the results of previous studies, we inferred that cme-miR156j and ath-miR5021 are involved in the saponin biosynthesis in soapberry pericarp.

TFs are proteins that bind to specific DNA sequences, usually motifs within the promoters of target genes, to control their transcription (Ye et al., 2020). Several studies have suggested that miRNAs can target TFs that control plant secondary metabolite biosynthesis, including anthocyanin

(Gou et al., 2011) and sesquiterpenoid (Yu et al., 2015). In most cases, highly conserved miRNAs regulate homologous targets encoding TFs at the same target sites in every plant species in which they exist (Gao et al., 2018). In the present study, at least five types of TFs (bHLH, WRKY, bZIP, ERF, and MYB) were predicted to be miRNA targets associated with saponin biosynthesis, involving 158 miRNA families (miR858, miR482, miR7817, miR5021, miR166, miR3951, and miR159, etc.). Among these TFs, several bHLHs have been shown to play crucial roles in the biosynthesis of triterpenoid saponin. For example, two bHLH TFs, TSAR1 and TSAR2, were shown to participate in the biosynthesis of nonhemolytic soyasaponins and hemolytic soyasaponins in *M. truncatula*, respectively (Mertens et al., 2016). GubHLH3 was shown to activate the transcription of the soyasaponin biosynthetic genes *CYP93E3* and *CYP72A566* in *Glycyrrhiza uralensis*, and overexpression of *GubHLH3* in transgenic *G. uralensis* hairy roots enhanced the levels of β -AS, *CYP93E3*, and *CYP72A566* transcripts (Tamura et al., 2018). Here, 54 bHLH genes were found to be targeted by 43 known miRNAs (including miR319, miR159, miR156, and miR408) and 11 novel miRNAs, indicating their roles in triterpenoid saponin biosynthesis in soapberry. This was similar to previous observations that miR156 and miR159 play important roles in regulating secondary metabolite biosynthesis by targeting bHLH TFs (Gou et al., 2011; Song et al., 2011). Furthermore, other TFs important for saponin biosynthesis, such as WRKY, bZIP, ERF, and MYB, were also shown to be miRNA targets (Xu et al., 2021a). Our study suggested that MYBs may be the targets of miR159, miR166, miR482, miR319, and miR858 in soapberry. Previous studies have also indicated that this is also the case in *Mentha spicata* (Reddy et al., 2017), *Arabidopsis thaliana* (Sharma et al., 2016), and *Malus domestica* (apple) (Xia et al., 2012). We also predicted that atr-miR166b targets *SmERF4*, ata-miR396c-5p targets *SmWRKY30*, ata-miR408-3p targets *SmbHLH6*, and ppe-miR858 targets *SmMYB32*. The results of qRT-PCR analyses confirmed that atr-miR166b, ata-miR396c-5p, ata-miR408-3p, and ppe-miR858 negatively regulate the expression of *SmERF4*, *SmWRKY30*, *SmbHLH6*, and *SmMYB32*, respectively (Figure 7). Taken together, these observations suggest that the predicted miRNAs are likely to target a variety of mRNAs encoding TFs. miRNAs have established a complex and diverse network to regulate saponin biosynthesis in soapberry pericarp, but further studies are needed to fully clarify the specific mechanisms of regulating this response.

Regulatory networks participated in saponin biosynthesis

A great deal of evidence suggests that a group of miRNAs (clusters and families) may be involved in the regulation of a group of common targets (Ponsuksili et al., 2013) and are related

to phenotypes. Here, WGCNA was performed to cluster miRNA products and revealed two modules (Mebrown and Meblue) significantly correlated with saponins. The Meblue module was also significantly negatively correlated with a mRNA module (MEgreenyellow). KEGG enrichment analysis showed that target genes from the Mebrown and Meblue modules were associated with transcription machinery and ABC transporters, respectively. Meanwhile, the pathways of sesquiterpenoid and triterpenoid biosynthesis and terpenoid backbone biosynthesis were also associated with the target genes. The miRNAs in these two selected modules may play important regulatory roles in saponin biosynthesis and transport. Furthermore, hub miRNAs are highly interconnected nodes in the network and may play regulatory roles in the common coexpressed network (Nunez et al., 2013). In our study, two hub miRNAs in MEbrown, mtr-miR319c-3p and bra-miR319-3p, were predicted to target *SmWRKY101* and *SmMVK2*, respectively. A hub miRNA in MEblue, han-miR3630-3p, was predicted to target a *DXS*, three *CYP450s*, one *bHLH*, seven *WRKYs*, and two *MYBs*. In addition, these three hub miRNAs were also identified as the DEMs during pericarp development. The results of qRT-PCR analyses showed an inverse correlation between han-miR3630-3p and *SmCYP71A-14* levels at most pericarp developmental stages. Based on the described functions of these miRNAs in this study, these hub miRNAs may be important for saponin biosynthesis in soapberry.

In addition, we found many miRNAs that were negatively correlated with their target genes in the modules Mebrown and Meblue. In the Mebrown module, genes that may participate in saponin biosynthesis, such as TFs, TRs, PKs, and transporters, were significantly negatively correlated with, and some of them were predicted to be targets of, atr-miR8577, aly-miR838-3p, aly-miR158b-3p, and far-miR1134. Many genes of the Meblue module were predicted to be associated with saponin biosynthesis, including six TFs, one TR, one PK, and five transporters. These genes were significantly negatively correlated with—and were predicted to be targets of—miRNAs, such as ath-miR5021, han-miR3630-3p, and cme-miR156j. Similarly, other miRNAs, including gra-miR8689, osa-miR414, and ptc-miR166p, were significantly negatively correlated with many enzymes, proteins, and other genes of unknown function, which may also play important roles in the biosynthesis of triterpenoid saponin in soapberry. However, although some miRNAs in the network have biological functions described in the literature, some do not have documented functions. It is possible that these unknown miRNAs may participate in the same pathways together with known miRNAs to regulate specific functions or pathways related to saponin biosynthesis.

Although the typical effect of miRNAs is to downregulate mRNA expression, there are also cases of positive miRNA regulation (Guo et al., 2010). In the present study, some miRNAs were not negatively correlated with their target genes,

and qRT-PCR also showed a lack of negative correlations of the relative expression of eight miRNA with that of their target gene pairs at certain stages. However, the positive correlations may also reflect secondary miRNA targets (Nunez et al., 2013) or adaptive target miRNA responses (Mamdani et al., 2015). Therefore, some interacting miRNA–mRNA pairs may not be negatively correlated, and this requires specific analysis based on specific situations.

Conclusions

This is the first study to investigate the miRNA expression profiles of soapberry pericarp. We identified 3036 miRNAs, including 2221 known miRNAs and 815 novel miRNAs, in soapberry pericarp, as well as their target genes. KEGG enrichment analysis suggested that some of the differentially expressed miRNAs, including ath-miR5021_R4-18L20, han-miR3630-3p_R5-20L22, and ppe-miR858_R3-18L21, may be involved in terpenoid backbone biosynthesis and sesquiterpenoid and triterpenoid biosynthesis pathways in soapberry pericarp. We constructed two coexpression miRNA modules/networks that were highly correlated with variations in the saponin contents or highly negatively correlated with the mRNA modules. Within the two networks, we identified 10 key candidate miRNAs that were weighted as module hub miRNAs. Furthermore, multiple miRNA–mRNA regulatory models were proposed, especially ath-miR5021–*SmIDI2/SmGPS5/SmbAS1/SmCYP71D-3/SmUGT74G-2*, cme-miR156j–*SmDXS4/SmbHLH18*, han-miR3630-3p–*SmCYP71A-14/SmbHLH54/SmMYB135/SmWRKY32*, atr-miR8577–*SmERF39*, ppe-miR858–*SmMYB5/SmMYB32*, and ppt-miR414–*SmMYB52*. Some of these miRNAs and their targets were validated by qRT-PCR. This study provided new evidence for miRNA regulation in triterpenoid saponin biosynthesis in soapberry and identified potential targets for designing experiments to regulate saponin content.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary Material](#).

Author contributions

YX, LJ and ZC conceived the research and designed the experiments. YX and JL wrote the manuscript. XJ, XW and LW analyzed the data. YH, YHG, SG and XHW performed the experiments. GZ, TZ and YG modified the language and revised

the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (No. 32071793) and the Special Foundation for National Science and Technology Basic Research Program of China (No. 2019FY100803).

Acknowledgments

We sincerely thank Shuijing Luo for his tending of the plants. We are also grateful to Kunjing Wu and Kui Liu for his guidance in the experiment. We gratefully acknowledge the assistance of Shuanglong Shi and Jing zhong in this study.

Conflict of interest

XJ and XHW were employed by Liming Jia. Author XJ was employed by Hangzhou KaiTai Biotechnology Co., Ltd. Author

XHW was employed by Yuanhua Forestry Biological Technology Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.1037784/full#supplementary-material>

References

- Augustin, J. M., Kuzina, V., Andersen, S. B., and Bak, S. (2011). Molecular activities, biosynthesis and evolution of triterpenoid saponins. *Phytochemistry* 72, 435–457. doi: 10.1016/j.phytochem.2011.01.015
- Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297. doi: 10.1016/S0092-8674(04)00045-5
- Beier, S., Thiel, T., Münch, T., Scholz, U., and Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33, 2583–2585. doi: 10.1093/bioinformatics/btx198
- Bologna, N. G., and Voinnet, O. (2014). The diversity, biogenesis, and activities of endogenous silencing small RNAs in arabidopsis. *Annu. Rev. Plant Biol.* 65, 473–503. doi: 10.1146/annurev-arplant-050213-035728
- Burge, S. W., Daub, J., Eberhardt, R. Y., Tate, J. G., Barquist, L., Nawrocki, E. P., et al. (2013). Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.* 41, 226–232. doi: 10.1093/nar/gks1005
- Candar-Cakir, B., Arican, E., and Zhang, B. (2016). Small RNA and degradome deep sequencing reveals drought- and tissue-specific microRNAs and their important roles in drought-sensitive and drought-tolerant tomato genotypes. *Plant Biotechnol. J.* 14, 1727–1746. doi: 10.1111/pbi.12533
- Cao, X., Wu, Z., Jiang, F., Zhou, R., and Yang, Z. (2014). Identification of chilling stress-responsive tomato microRNAs and their target genes by high-throughput sequencing and degradome analysis. *BMC Genomics* 15, 1130. doi: 10.1186/1471-2164-15-1130
- Cheah, B. H., Nadarajah, K., Divate, M. D., and Wickneswari, R. (2015). Identification of four functionally important microRNA families with contrasting differential expression profiles between drought-tolerant and susceptible rice leaf at vegetative stage. *BMC Genomics* 16, 692. doi: 10.1186/s12864-015-1851-3
- Chen, C., Xie, F., Hua, Q., Zur, N. T., Zhang, L., Zhang, Z., et al. (2020). Integrated sRNAome and RNA-seq analysis reveals miRNA effects on betalain biosynthesis in pitaya. *BMC Plant Biol.* 20, 437. doi: 10.1186/s12870-020-02622-x
- Cuperus, J. T., Fahlgren, N., and Carrington, J. C. (2011). Evolution and functional diversification of MIRNA genes. *Plant Cell* 23, 431–442. doi: 10.1105/tpc.110.082784
- Devi, S. J. S. R., Madhav, M. S., Kumar, G. R., Goel, A. K., Umakanth, B., Jahnavi, B., et al. (2013). Identification of abiotic stress miRNA transcription factor binding motifs (TFBMs) in rice. *Gene* 531, 15–22. doi: 10.1016/j.gene.2013.08.060
- Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D. S. (2003). MicroRNA targets in drosophila. *Genome Biol.* 1, R1. doi: 10.1186/gb-2003-5-1-r1
- Fahlgren, N., Howell, M. D., Kasschau, K. D., Chapman, E. J., Sullivan, C. M., and Cumbie, J. S. (2007). High-throughput sequencing of arabidopsis microRNAs: evidence for frequent birth and death of miRNA genes. *PLoS One* 2, e219. doi: 10.1371/journal.pone.0000219
- Fan, R., Li, Y., Li, C., and Zhang, Y. (2015). Differential microRNA analysis of glandular trichomes and young leaves in *Xanthium strumarium* L. reveals their putative roles in regulating terpenoid biosynthesis. *PLoS One* 10, e0139002. doi: 10.1371/journal.pone.0139002
- Gao, J., Chen, H., Yang, H., Yong, H., Tian, Z., and Li, J. (2018). A brassinosteroid responsive miRNA-target module regulates gibberellin biosynthesis and plant development. *New Phytol.* 220, 488–501. doi: 10.1111/nph.15331
- Gou, J. Y., Felippes, F. F., Liu, C. J., Weigel, D., and Wang, J. W. (2011). Negative regulation of anthocyanin biosynthesis in arabidopsis by a miR156-targeted SPL transcription factor. *Plant Cell* 23, 1512–1522. doi: 10.1105/tpc.111.084525
- Guo, H., Ingolia, N. T., Weissman, J. S., and Bartel, D. P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 466, 835–840. doi: 10.1038/nature09267
- Guo, Y., Zhao, S., Zhu, C., Chang, X., Yue, C., Wang, Z., et al. (2017). Identification of drought-responsive miRNAs and physiological characterization of tea plant (*Camellia sinensis* L.) under drought stress. *BMC Plant Biol.* 17, 211. doi: 10.1186/s12870-017-1172-6
- Gupta, O. P., Karkute, S. G., Banerjee, S., Meena, N. L., and Dahuja, A. (2017). Contemporary understanding of miRNA-based regulation of secondary metabolites biosynthesis in plants. *Front. Plant Sci.* 8, 374. doi: 10.3389/fpls.2017.00374
- Hao, D. C., Yang, L., Xiao, P. G., and Liu, M. (2012). Identification of taxus microRNAs and their targets with high-throughput sequencing and degradome analysis. *Physiol. Plantarum* 146, 388–403. doi: 10.1111/j.1399-3054.2012.01668.x
- Hernández-Plaza, A., Szklarczyk, D., Botas, J., Cantalapiedra, C. P., Giner-Lamia, J., Mende, D. R., et al. (2022). eggNOG 6.0: enabling comparative genomics across 12535 organisms. *Nucleic Acids Res.*, gkac1022. doi: 10.1093/nar/gkac1022

- Hollender, C. A., Kang, C., Darwish, O., Geretz, A., Matthews, B. F., Slovins, J., et al. (2014). Floral transcriptomes in woodland strawberry uncover developing receptacle and anther gene networks. *Plant Physiol.* 165, 1062–1075. doi: 10.1104/pp.114.237529
- Kajal, M., and Singh, K. (2017). Small RNA profiling for identification of miRNAs involved in regulation of saponins biosynthesis in *Chlorophytum borivillanum*. *BMC Plant Biol.* 17, 265. doi: 10.1186/s12870-017-1214-0
- Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42, 68–73. doi: 10.1093/nar/gkt1181
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* 9, 559. doi: 10.1186/1471-2105-9-559
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Liu, M., Yu, H., Zhao, G., Huang, Q., Lu, Y., and Bo, O. (2017). Profiling of drought-responsive microRNA and mRNA in tomato using high-throughput sequencing. *BMC Genomics* 18, 481. doi: 10.1186/s12864-017-3869-1
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq 2. *Genome Biol.* 15, 550. doi: 10.1186/s13059-014-0550-8
- Mamdani, M., Williamson, V., McMichael, G. O., Blevins, T., Aliev, F., Adkins, A., et al. (2015). Integrating mRNA and miRNA weighted gene co-expression networks with eQTLs in the nucleus accumbens of subjects with alcohol dependence. *PLoS One* 10, e0137671. doi: 10.1371/journal.pone.0137671
- Mandhan, V., Kaur, J., and Singh, K. (2012). SmRNAome profiling to identify conserved and novel microRNAs in *Stevia rebaudiana bertonii*. *BMC Plant Biol.* 12, 197–197. doi: 10.1186/1471-2229-12-197
- Mathiyalagan, R., Subramaniyam, S., Natarajan, S., Kim, Y. J., Sun, M. S., Kim, S. Y., et al. (2013). In silico profiling of microRNAs in Korean ginseng (*Panax ginseng Meyer*). *J. Ginseng Res.* 37, 227–247. doi: 10.5142/jgr.2013.37.227
- Mertens, J., Pollier, J., Bossche, R. V., Lopez-Vidriero, I., Franco-Zorrilla, J. M., and Goossens, A. (2016). The bHLH transcription factors TSAR1 and TSAR2 regulate triterpene saponin biosynthesis in *Medicago truncatula*. *Plant Physiol.* 170, 194–210. doi: 10.1104/pp.15.01645
- Nunez, Y. O., Truitt, J. M., Gorini, G., Ponomareva, O. N., Blednov, Y. A., Harris, R. A., et al. (2013). Positively correlated miRNA-mRNA regulatory networks in mouse frontal cortex during early stages of alcohol dependence. *BMC Genomics* 14, 725. doi: 10.1186/1471-2164-14-725
- Pani, A., and Mahapatra, R. K. (2013). Computational identification of microRNAs and their targets in *Catharanthus roseus* expressed sequence tags. *Genom Data* 1, 2–6. doi: 10.1016/j.gdata.2013.06.001
- Ponsuksili, S., Du, Y., Hadlich, F., Siengdee, P., Murani, E., Schwerin, M., et al. (2013). Correlated mRNAs and miRNAs from co-expression and regulatory networks affect porcine muscle and finally meat properties. *BMC Genomics* 14, 533. doi: 10.1186/1471-2164-14-533
- Reddy, V. A., Wang, Q., Dhar, N., Kumar, N., Venkatesh, P. N., Rajan, C., et al. (2017). Spearmint R2R3-MYB transcription factor MsMYB negatively regulates monoterpene production and suppresses the expression of geranyl diphosphate synthase large subunit (MsGPPS_LSU). *Plant Biotechnol. J.* 15, 1105–1119. doi: 10.1111/pbi.12701
- Rehmsmeier, M., Steffen, P., Hochsmann, M., and Giegerich, R. (2004). Fast and effective prediction of microRNA/target duplexes. *RNA* 10, 1507–1517. doi: 10.1261/rna.5248604
- Samad, A. F. A., Rahnamaie-Tajadod, R., Sajad, M., Jani, J., Murad, A. M. A., Noor, N. M., et al. (2019). Regulation of terpenoid biosynthesis by miRNA in *Persicaria minor* induced by *Fusarium oxysporum*. *BMC Genom* 20, 586. doi: 10.1186/s12864-019-5994-5
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1293903
- Sharma, D., Tiwari, M., Pandey, A., Bhatia, C., Sharma, A., and Trivedi, P. K. (2016). MicroRNA858 is a potential regulator of phenylpropanoid pathway and plant development in arabidopsis. *Plant Physiol.* 171, 01831. doi: 10.1104/pp.15.01831
- Singh, N., Srivastava, S., Shasany, A. K., and Sharma, A. (2016). Identification of miRNAs and their targets involved in the secondary metabolic pathways of *Mentha* spp. *Comput. Biol. Chem.* 64, 154–162. doi: 10.1016/j.compbiolchem.2016.06.004
- Song, X., Li, Y., Cao, X., and Qi, Y. (2019). MicroRNAs and their regulatory roles in plant-environment interactions. *Annu. Rev. Plant Biol.* 70, 489–525. doi: 10.1146/annurev-arplant-050718-100334
- Song, Q. X., Liu, Y. F., Hu, X. Y., Zhang, W. K., Ma, B., Chen, S. Y., et al. (2011). Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing. *BMC Plant Biol.* 11, 5. doi: 10.1186/1471-2229-11-5
- Sun, C., Wang, J., Duan, J., Zhao, G., Weng, X., and Jia, L. (2017). Association of fruit and seed traits of *Sapindus mukorossi* germplasm with environmental factors in southern China. *Forests* 8, 491. doi: 10.3390/f8120491
- Tamura, K., Yoshida, K., Hiraoka, Y., Sakaguchi, D., Chikugo, A., Mochida, K., et al. (2018). The basic helix-loop-helix transcription factor GubHLH3 positively regulates soyasaponin biosynthetic genes in *Glycyrrhiza uralensis*. *Plant Cell Physiol.* 59, 783–796. doi: 10.1093/pcp/pcy046
- Upadhyay, A., and Singh, D. K. (2012). Pharmacological effects of *Sapindus mukorossi*. *Rev. Do Instituto Medicina Trop. Sao Paulo* 54, 273–280. doi: 10.1590/S0036-46652012000500007
- Wang, B., Sun, Y., Song, N., Wei, J. P., Wang, X. J., Feng, H., et al. (2014). MicroRNAs involving in cold, wounding and salt stresses in *Triticum aestivum* L. *Plant Physiol. Biochem.* 80, 90–96. doi: 10.1016/j.plaphy.2014.03.020
- Wei, R., Qiu, D., Wilson, I. W., Zhao, H., Lu, S., Miao, J., et al. (2015). Identification of novel and conserved microRNAs in *Panax notoginseng* roots by high-throughput sequencing. *BMC Genomics* 16, 1–10. doi: 10.1186/s12864-015-2010-6
- Xia, R., Zhu, H., An, Y. Q., Beers, E. P., and Liu, Z. (2012). Apple miRNAs and tasiRNAs with novel regulatory networks. *Genome Biol.* 13, R47. doi: 10.1186/gb-2012-13-6-r47
- Xie, R., Jin, Z., Ma, Y., Pan, X., Dong, C., Pang, S., et al. (2017). Combined analysis of mRNA and miRNA identifies dehydration and salinity responsive key molecular players in citrus roots. *Sci. Rep.* 7, 42094. doi: 10.1038/srep42094
- Xie, F., Wang, Q., Sun, R., and Zhang, B. (2015). Deep sequencing reveals important roles of microRNAs in response to drought and salinity stress in cotton. *J. Exp. Bot.* 66, 789–804. doi: 10.1093/jxb/eru437
- Xu, Y., Chen, Z., Jia, L., and Weng, X. (2021a). Advances in understanding of the biosynthetic pathway and regulatory mechanism of triterpenoid saponins in plants (in chinese). *Sci. Sin. Vitae* 51, 525–555. doi: 10.1360/SSV-2020-0230
- Xu, Y., Gao, Y., Chen, Z., Zhao, G., Liu, J., Wang, X., et al. (2021b). Metabolomics analysis of the soapberry (*Sapindus mukorossi* Gaertn.) pericarp during fruit development and ripening based on UHPLC-HRMS. *Sci. Rep.* 11, 11657. doi: 10.1038/s41598-021-91143-0
- Xu, Y., Jia, L., Chen, Z., and Gao, Y. (2018). Advances on triterpenoid saponin of *Sapindus mukorossi* (in chinese). *Chemistry* 81, 1078–1088. doi: 10.14159/j.cnki.0441-3776.2018.12.003
- Xu, Y., Zhao, G., Ji, X., Liu, J., Zhao, T., Gao, Y., et al. (2022). Metabolome and transcriptome analysis reveals the transcriptional regulatory mechanism of triterpenoid saponin biosynthesis in soapberry (*Sapindus mukorossi* Gaertn.). *J. Agric. Food Chem.* 70, 7095–7109. doi: 10.1021/acs.jafc.2c01672
- Ye, J., Zhang, X., Tan, J., Xu, F., Cheng, S., Chen, Z., et al. (2020). Global identification of *Ginkgo biloba* microRNAs and insight into their role in metabolism regulatory network of terpene trilactones by high-throughput sequencing and degradome analysis. *Ind. Crops Prod* 148, 112289. doi: 10.1016/j.indcrop.2020.112289
- Yin, S. W., Chen, J. C., Sun, S. D., Tang, C. H., Yang, X. Q., Wen, Q. B., et al. (2011). Physicochemical and structural characterization of protein isolate, globulin and albumin from soapnut seeds (*Sapindus mukorossi* Gaertn.). *Food Chem.* 128, 420–426. doi: 10.1016/j.foodchem.2011.03.046
- Yu, Z. X., Wang, L. J., Zhao, B., Shan, C. M., Zhang, Y. H., Chen, D. F., et al. (2015). Progressive regulation of sesquiterpene biosynthesis in arabidopsis and patchouli (*Pogostemon cablin*) by the miR156-targeted SPL transcription factors. *Mol. Plant* 8, 98–110. doi: 10.1016/j.molp.2014.11.002
- Zhao, G., Gao, Y., Gao, S., Xu, Y., and Jia, L. (2019). The phenological growth stages of *Sapindus mukorossi* according to BBCH scale. *Forests* 10, 462. doi: 10.3390/f10060462
- Zhao, Y. J., and Li, C. (2018). Biosynthesis of plant triterpenoid saponins in microbial cell factories. *J. Agr. Food Chem.* 66, 12155–12165. doi: 10.1021/acs.jafc.8b04657
- Zhao, Y. T., Wang, M., Fu, S. X., Yang, W. C., Qi, C. K., and Wang, X. J. (2012). Small RNA profiling in two *Brassica apus* cultivars identifies microRNAs with oil production- and development-correlated expression and new small RNA classes. *Plant Physiol.* 158, 813–823. doi: 10.1104/pp.111.187666
- Zhao, C. Z., Xia, H., Frazier, T. P., Yao, Y. Y., Bi, Y. P., Li, A. Q., et al. (2010). Deep sequencing identifies novel and conserved microRNAs in peanuts (*Arachis hypogaea* L.). *BMC Plant Biol.* 10, 3. doi: 10.1186/1471-2229-10-3
- Zheng, C., Zhao, L., Wang, Y., Shen, J., Zhang, Y., Jia, S., et al. (2015). Integrated RNA-seq and sRNA-seq analysis identifies chilling and freezing responsive key molecular players and pathways in tea plant (*Camellia sinensis*). *PLoS One* 10, e0125031. doi: 10.1371/journal.pone.0125031