

Genome-scale Transcriptomic Insights Into the Gene-regulatory Network of Seed Abortion in Triploid *Siraitia Grosvenorii*

Rongchang Wei

Guangxi Academy of Agricultural Sciences

Dongping Tu

Guangxi University of Chinese Medicine

Xiyang Huang

Guangxi Institute of Botany, Guangxi Zhuang Autonomous Region and Chinese Academy of Sciences

Zuliang Luo

Chinese academy of Medical Science & Peking Union Medical College

Xiaohua Huang

Guangxi Academy of Agricultural Sciences

Nan Cui

Guangxi Institute of Botany, Guangxi Zhuang Autonomous Region and Chinese Academy of Sciences

Juan Xu

Cash Crops Research Institute, Guangxi Academy of Agricultural Sciences

Faqian Xiong (✉ xfq2002@126.com)

Guangxi Academy of Agricultural Sciences

Haifeng Yan

Guangxi Academy of Agricultural Sciences

Xiaojun Ma

Chinese academy of Medical Science & Peking Union Medical College

Research Article

Keywords: *Siraitia grosvenorii*, seed abortion, transcriptome, gene-regulatory network, plant hormones

Posted Date: October 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-966900/v1>

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

[Read Full License](#)

Abstract

Background

Siraitia grosvenorii (Swingle) C. Jeffrey, also known as Luohanguo or monk fruit, is a famous traditional Chinese medicine ingredient with important medicinal value and broad development prospects. Diploid *S. grosvenorii* has too many seeds. Thus, studying the molecular mechanism of seed abortion in triploid *S. grosvenorii*, identifying the abortion-related genes, and regulating their expression will be a new direction to obtain seedless *S. grosvenorii*. Herein, we examined the submicroscopic structure of triploid *S. grosvenorii* seeds during abortion.

Results

By measuring the content of endogenous hormones, we found that abscisic acid (ABA) and trans-zeatin (ZR) levels were significantly downregulated at days 15 and 20 after flowering. RNA-seq of triploid seeds at different developmental stages was performed to identify the key genes regulating abortion of triploid *S. grosvenorii* seeds. Multiple differentially expressed genes between adjacent stages were identified; seven genes were differentially expressed across all stages. Weight gene co-expression network analysis revealed that the enhancement of monoterpene and terpene metabolic processes might lead to seed abortion by reducing substrate flow to ABA and ZR.

Conclusions

These findings provide insights into the gene-regulatory network of seed abortion in triploid *S. grosvenorii* from different perspectives, thereby facilitating the innovation of the breeding technology of *S. grosvenorii*.

Background

Siraitia grosvenorii (Swingle) C. Jeffrey is a species of the genus *Siraitia* Merr. (Cucurbitaceae). It is native to the southern parts of China and is mainly found in Guangxi Province [1]. *S. grosvenorii* is also known as Luohanguo (LHG) or monk fruit. LHG is not only used as a food ingredient, but also as in traditional Chinese medicine. In China, LHG has been used as a natural cough suppressant and expectorant for over 300 years and is one of the first approved medicinal food homology species. It is also highly favored internationally [1,2].

Previous studies have revealed that several different classes of compounds were isolated from *S. grosvenorii*, including triterpenoids, iridoid, flavonoids glycosides, vitamins, proteins, saccharides, and a volatile oil [3-5]. Because LHG contains sweet glycosides that are naturally low in calories, the extracts from ripe LHG can be used as supplements and sweeteners in sugar-free health foods and beverages [6,7]. In addition, this plant is currently used as an analgesic for the lungs and an emollient for the treatment of sore throat, thirst, and constipation [8]. The crude extracts and purified products of *S.*

grosvenorii have several biological activities, including immunologic, antioxidant, anti-tussive, sputum-reducing, hypoglycemic, hepatoprotective, and antimicrobial [9-13]. The triterpene compound, mogrosides, is considered the main active ingredient, contributing to the sweet taste and the main biologically active ingredient of *S. grosvenorii* [14]. With the development of science and technology, the application of *S. grosvenorii* in the field of medicine can be further explored. Therefore, *S. grosvenorii* is a natural product with high development potential for medicine and commodity production, and it is attracting increasing attention from scientific research community and commercial establishments [3,15].

In the past 10 years, chromosome engineering has been a hot topic in research on the germplasm resources of *S. grosvenorii*, and several polyploid *S. grosvenorii* resources have been obtained. Triploid *S. grosvenorii* is an excellent polyploid germplasm resource. The tetraploid female plant obtained by colchicine induction and the diploid male parent were hybridized, and triploid seeds were obtained after the fruit ripened. Compared with the diploid *S. grosvenorii*, the triploid plants had larger vegetative and reproductive organs. Further, it had seedless fruits or fruits with only few seeds. The content of mogroside V in triploid fruit was 36.28% higher than that in the control. In addition, triploid plants were found exhibiting strong resistance [16,17].

Seed development can be divided into three major phases: embryogenesis, seed maturation, and desiccation. Embryogenic processes include cell division and expansion, morphogenesis, and the beginning of the endosperm and embryo development. A series of morphological, physiological, and biochemical changes are involved in seed development [18-20]. The production of viable seeds is thought to require the co-regulation of multiple plant endogenous hormones. Abscisic acid (ABA) and gibberellin (GA) are the main hormones that regulate seed formation. The process of seed dormancy interruption is determined by the distribution of GA and ABA. In addition, ABA regulates seed maturation, embryo morphogenesis, and desiccation. Auxin is an important molecule that regulates seed development in conjunction with ABA [21-24]. Trans-zeatin (ZR) is a cytokinin that exists in seeds that can induce cell division and proliferation [25]. We speculate that seed abortion may be regulated by fluctuations in the abovementioned four hormones; however, related evidence is absent.

At present, there are few reports on breeding new species of *S. grosvenorii* using genetic-engineering technology. Thus, studying the molecular mechanism of seed abortion and identifying the genes and pathways related to seed abortion in *S. grosvenorii* will provide a theoretical basis for genetic engineering to breed new seedless *S. grosvenorii* species. Here, we analyzed the transcriptome sequencing data of the triploid seeds of LHG at different developmental stages and combined this with the endogenous hormonal changes of seed development to reveal the gene-regulatory network involved in the underlying mechanisms for formation of triploid LHG abortive seeds.

Results

The profile of seed abortion in *S. grosvenorii*

The structural changes in the tissues at each stage were observed using TEM. There was no obvious edema in the cytoplasm and no obvious separation of the plasma wall at stage 5DAF (Figure 1a and b). The nucleus was irregular, the mitochondria were oval, the crest and rough endoplasmic reticulum were moderately dilated, the surface ribosome was locally exfoliated, and the number of intracellular starch granules was abundant. Severe edema of the cytoplasm and slight separation of the plasma wall were observed at stages 10DAF (Figure 1c and d) and 15DAF (Figure 1e and f), respectively. Severe separation of the plasma wall, local damage to the cell membrane, and thickness of the cell wall were observed at 20DAF (Figure 1g and h) and gradually recovery was observed from stages 25DAF (Figure 1i and j) to 30DAF (Figure 1k and l).

We also examined the levels of endogenous plant hormones, ABA, ZR, IAA, and GA3, during seed abortion using ELISA (Figure 2a-d). Interestingly, the levels of ABA and ZR decreased remarkably at stages 15DAF and 20DAF, respectively. The above changes in submicroscopic structure and levels of endogenous plant hormones indicated that 15–20DAF was a key transition period during seed abortion. We will mainly focus on the exploration of the transcriptomic changes at stages 15DAF and 20DAF in the subsequent analyses.

Global analysis of the RNA-Seq data

Genes with significant differential expression ($|\log_2FC| > 1$ and FDR-adjusted p-value < 0.05) were identified between consecutive stages. The number of DEGs in the five comparisons is shown in Figure 3a and b. The total DEGs in 10DAF vs 5DAF were 7,692 (4,145 upregulated and 3,546 downregulated), 15DAF vs 10DAF were 3,252 (1,487 upregulated and 1,765 downregulated), 20DAF vs 15DAF were 984 (364 upregulated and 620 downregulated), 25DAF vs 20DAF were 2,124 (942 upregulated and 1,182 downregulated), and 30DAF vs 25DAF were 885 (258 upregulated and 627 downregulated).

The DEGs were further functionally classified into GO slims (Tables S1–S5; $p < 0.05$). DEGs in 20DAF vs. 15DAF were identified to be involved in meristem structural organization, morphogenesis of a branching structure, adaxial/abaxial axis specification, positive regulation of catalytic activity, oxidoreductase activity, and glucosidase activity (Table S3), whereas those in 25DAF vs. 20DAF were involved in trehalose metabolic process, seed germination, response to sucrose, transmembrane receptor histidine kinase activity, and so forth. (Table S4).

Seven common DEGs shared by all comparisons of adjacent stages were identified (Figure 3b), which revealed fluctuations in expression levels across all stages (Figure 3c). These DEGs were found to be associated with toxin metabolic processes, regulation of gene expression, glucan metabolic process, and external encapsulating structure organization.

Identification of temporal expression trends across *S. grosvenorii* seed transcriptomes

We performed STEM analysis to visualize the expression patterns of mRNAs. As a result, seven expression profiles were found to be statistically significant (Figure 4a). Profiles 0 and 19 tended to be

continuously downregulated and upregulated during seed abortion, respectively, whereas profiles 1 and 18 reached an expression peak and valley at stages 15DAF and 20DAF, respectively.

KEGG pathway enrichment analysis was performed for the four expression trends (19, 0, 1, and 18). The genes in profile 0 with a continuously downregulated trend were mainly enriched in starch and sucrose metabolism, plant hormone signal transduction, linoleic acid metabolism, replication and repair, and other glycan degradation (Figure 4b), whereas those in profile 19, which were continuously upregulated, were related to amino acid metabolism; phenylpropanoid biosynthesis; ubiquinone and other terpenoid-quinone biosynthesis; flavonoid biosynthesis; stilbenoid, diarylheptanoid, and gingerol biosynthesis; diterpenoid biosynthesis; glycolysis/gluconeogenesis; and plant-pathogen interactions (Figure 4c). The genes in profile 1 with the lowest expression level at 15DAF and 20DAF were involved in RNA polymerase, butanoate metabolism, inositol phosphate metabolism, and fatty acid degradation (Figure 4d). Finally, the genes in profile 18 with the highest expression levels at stages 15DAF and 20DAF were involved in arachidonic acid metabolism, amino sugar and nucleotide sugar metabolism, one carbon pool by folate, cyanoamino acid metabolism, glutathione metabolism, isoquinoline alkaloid biosynthesis, phagosome, glycerolipid metabolism, and fructose and mannose metabolism (Figure 4e).

Co-expression network associated with seed abortion

A gene co-expression network was used to cluster 21,152 genes after filtering those with low expression levels, resulting in 16 modules with different colors (Figure 5a). Combined with the module-trait relationship and module significance, we eventually identified the brown2 module as the most phenotypically relevant module, which showed a negative correlation with the level of ABA (the correlation coefficient was -0.71 and the FDR-adjusted p-value was $9e-04$), whereas the lightsteelblue module showed a positive correlation (Figure 5b). The indianred4 and skyblue3 modules were also found to be negatively and positively associated with the level of ZR, respectively, with statistical significance (Figure 5b).

We explored the expression profiles of each module across the different stages (Figure 6a). Interestingly, we found that the genes in the brown2 module were specifically, highly expressed at 15DAF and 20DAF (Figure 6a). By combining the above results, we considered the brown2 module to be negatively associated with the changes in ABA levels at 15DAF. According to the results of GO enrichment analysis, the genes in the brown2 module were enriched in monoterpene metabolic processes and terpene metabolic processes (Figure 6b). Additionally, KEGG pathway analysis revealed that the genes were involved in sesquiterpenoid and triterpenoid biosynthesis and amino sugar and nucleotide sugar metabolism (Figure 6c). These findings indicate that these genes, which are related to terpene metabolism, are highly expressed at stage 15DAF and may play an important role in the decrease of ABA levels, thereby leading to seed abortion.

Finally, the co-expression networks of the brown2, indianred4, lightsteelblue, and skyblue3 modules were filtered by a weight value greater than 0.15 and visualized (Figure 7a-d). The genes with the highest betweenness centrality score in the brown2 module were *evm.TU.tig00004479.11* and

evm.TU.tig00153447.44, which are non-specific lipid-transfer protein 2-like (LTP) and protein root initiation defective 3-like, respectively (Figure 7a). The genes with the highest betweenness centrality score in the indianred4 module included CCR2, PCBER, and CYP75B2, which were cinnamoyl-CoA reductase 2-like, isoflavone reductase-like protein IRL, and cytochrome P450 71A1-like, respectively (Figure7b). The genes with the highest betweenness centrality scores in the lightsteelblue module were beta-galactosidase 10 (Os01g0875500-2, Os01g0875500-3, and Os01g0875500-1) (Figure7c). Finally, the genes with the highest betweenness centrality score in the skyblue3 module were calcium-transporting ATPase 12 plasma membrane-type-like, ACA12-1 and ACA12-2, and the ethylene-responsive transcription factor, ERF105 (ERF5) (Figure 7d).

Discussion

The structural changes in the aborted seeds of *S. grosvenorii* at different phases observed using TEM revealed that the developmental abnormality began on 10DAF and gradually deteriorated (Figure 1). Furthermore, the defects were found to be serious at 20DAF. The results of the endogenous plant hormone determination of these sterile seeds were also consistent with those of TEM (Figure 2). ABA is considered one of the main hormones that regulate seed formation [26], and ZR can induce cell division and proliferation [25]. Both were significantly decreased at stages 15DAF and 20DAF. These findings indicate that 15–20 DAF is the critical period during seed abortion.

Transcriptome data at different developmental stages provide a large amount of data for the dynamic formation of triploid *S. grosvenorii* seeds. The results showed that the number of genes with significant changes in expression was highest during the early stages of seed formation. However, relatively fewer genes changed significantly before and after 20DAF (Figure 3a). Although few genes were differentially expressed at 20DAF vs. 15DAF, these genes were most likely associated with seed abortion as 15–20 DAF was considered to be the critical period during seed abortion. During the embryonic development of a seed, a complete tissue structure must be established, and the cells of the embryo need to become specific and must differentiate into cell types in an integrated manner. This process establishes a simple body structure. This structure is composed of shoot meristem, cotyledons, hypocotyl, root, and root meristem arranged along the apical-basal axis and a concentric arrangement of epidermis, subepidermal basal tissue, and central vascular cylinder arranged concentrically along the radial axis [27]. The basic body of a plant is formed during embryogenesis, and formation of most organs and tissues occurs in the late stage of the embryo [18]. As early as the eight-cell division stage in *Arabidopsis*, the regions that generate the shoot meristem and root meristem begin to differentiate. Meristem is vital in postembryonic organ formation; moreover, the shoot apical meristem can be subdivided into regions with different properties and functions [27]. In addition to meristem structural organization, morphogenesis of a branching structure and adaxial/abaxial axis specification are also important for the formation of the complete tissue structure. Both ABA and ZR regulate embryo morphogenesis. Hence, seed abortion may be caused by abnormal embryonic morphology induced by fluctuations in ABA and ZR levels.

The continuously downregulated genes involved in starch and sucrose metabolism may lead to seed abortion. Sucrose is not only a nutrient, but also a signal that induces storage-related differentiation. It can induce premature mitotic termination and cell enlargement, activate the expression of storage-related genes in maize and *Vicia*, and promote storage activity in cotyledons at the transcriptional level and starch accumulation [28–33]. Under drought stress, the transcription and activity of invertase, which is important for controlling the ratio of sucrose to hexose, are reported to be reduced, thereby interfering with the sucrose use capacity and the ratio of sucrose to hexose, which disrupt maternal tissue development and eventually lead to seed abortion [34]. Sucrose and ABA also have similar effects on gene expression. Sucrose may increase ABA levels or enhance ABA sensitivity [35]. However, this hypothesis has not been proven. Here, the variation trend of plant hormone signal transduction is similar to that of sucrose metabolism, which also reflects the potential mutual assistance between endogenous hormones and sucrose. Notably, many genes that are continuously upregulated during all developmental phases are involved in the biosynthesis of various secondary metabolites. Some classes of secondary metabolites, including phenylpropanoids, triterpenoid glycosides, and flavonoids, have been isolated from *S. grosvenorii* in previous studies [3]. However, upregulated genes associated with secondary metabolite biosynthesis were not involved in the synthesis of mogrosides or other triterpenoid glycosides. Although mogrosides are responsible for the main biological effects of *S. grosvenorii* fruits, the biological effects of the other chemical components were still present. For example, the flavonoid fraction of *S. grosvenorii* leaves showed antioxidative effects in rats [36]. These data can serve as a reference for the directional breeding of *S. grosvenorii* with important medicinal value. In addition, some genes related to plant-pathogen interactions were continuously upregulated. This may explain the previously reported strong resistance of triploid *S. grosvenorii* [16] and provide a theoretical basis for breeding *S. grosvenorii* cultivars with high resistance.

Some genes with the lowest or highest expression levels at 15DAF and 20DAF are involved in the biological processes that are regulated by ABA. These biological processes include inositol phosphate metabolism, fatty acid degradation, and arachidonic acid metabolism (Figure 4). The mRNA levels of D-myo-inositol-3-phosphate synthase and inositol bisphosphate were induced by ABA in *Spirodela polyrhiza* [37]. In *Chlorella vulgaris*, changes in the regulation of fatty acid biosynthesis are triggered by exogenous ABA [38]. In *Mortierella alpina*, ABA stimulates arachidonic acid biosynthesis [13]. In addition, ABA is considered to trigger a wide range of transcriptome changes that help plants respond to environmental stimuli [39]. RNA polymerase is essential for gene transcription. Although the regulation of RNA polymerase by ABA has not been reported, we suspect that these RNA polymerase-related genes, which are expressed at extremely low levels in the critical phases of triploid seed abortion, may be influenced by the endogenous hormone, ABA, to regulate transcription and lead to seed abortion. Furthermore, we speculated that these biological processes, which have been reported or may be regulated by ABA, may lead to abnormal seed development and sterility due to the effects of decreased ABA in these triploid plants.

Further analysis of the co-expression networks provided new insights into the regulatory mechanisms of ABA genes. The genes in the brown2 module, which were negatively associated with the changes in ABA

levels, were enriched in monoterpene metabolic processes and terpene metabolic processes (Figure 6). Both monoterpenes and terpenes contain a series of terpenoids. In higher plants, terpenoids are substrates for several compounds, including ABA, cytokinins, and the phytol side chain of chlorophyll. In addition, ABA and the phytol side chain of chlorophyll, as well as many secondary metabolites, including monoterpene and terpene, are downstream products of isopentenyl diphosphate [40-42]. *CAB40* in *brown2* module encodes the chlorophyll a-b binding protein of LHCII type 1. The expression level of *CAB40* was high and negatively associated with the fluctuation of ABA level at stage 15DAF, suggesting that terpenoids flux toward synthesis of the phytol side chain of chlorophyll. These results suggest that at 15DAF, the enhancement of the metabolic processes of monoterpene and terpene may lead to seed abortion by reducing substrate flow to ABA and cytokinins.

Notably, a variety of genes in the *brown2* module were associated with plant resistance. *CHIT3-2* encodes an acidic chitinase. Acidic chitinase is found in plants responding to pathogen attack and plants treated with salt solutions or salicylic acid [43,44]. Another gene, *LYK2*, may also be involved in plant-microbial interactions. *LYK2* is a homolog of Nod factor receptor 5 (NFR5) of *Lotus japonicus*, which functions in Nod factor recognition during nodulation. The bacterial Nod factor is a GlcNAc lipochitooligosaccharide with a structure similar to that of peptidoglycan [45]. In *Arabidopsis thaliana* and rice, *broad-spectrum Resistance 2 (BSR2)*, a gene encoding an uncharacterized cytochrome P450 protein belonging to the CYP78A family, enhances resistance against *Rhizoctonia solani* [46]. Trehalose-6-phosphatase synthase S4 (TPS4) is a protein involved in trehalose biosynthesis. It was considered to play a positive role in the response to salt stress in *Brassica napus* var. *oleifera* [47]. These resistance-related genes were found to be negatively correlated with ABA, suggesting that they may be affected by ABA or they may be one of the factors causing ABA fluctuations. This hypothesis requires further verification.

Previously, *LTP* was reported to be induced by ABA in the vegetative tissues of rice [48]. However, *LTP*, which had the highest betweenness centrality score in the *brown2* module, was negatively correlated with ABA based on our results, indicating that ABA may have different effects on the regulation of *LTP* expression in different tissues or developmental phases. In previous reports, two cytochrome P450 monooxygenases, *CYP735A1* and *CYP735A2*, were demonstrated to catalyze the biosynthesis of ZR in *Arabidopsis* [49]. Here, in the *indianred4* module, *CYP75B2* expression was negatively correlated with the level of ZR, suggesting a different role of cytochrome P450 monooxygenases in the biosynthesis of ZR.

Conclusions

In conclusion, we found that ABA and ZR fluctuated significantly at 15 and 20 DAF when the abortive seeds displayed structural abnormalities. Subsequent transcriptomic analyses provided us with a large amount of information, including the gene networks that may regulate the biosynthesis of hormones, ABA and ZR, as well as the metabolic pathways or biological processes that may be affected by plant hormones. These findings provide a strong theoretical basis and candidate genes for breeding seedless *S. grosvenorii* species using the genetic-engineering technology.

Methods

Plant materials and microscopic monitoring of seed abortion

S. grosvenorii line seedless 1 used in this study were kindly provided by Mr. Xiangjun Jiang, Guilin yiyuansheng modern biotechnology Co., Ltd, Guilin, China. The triploid seeds of LHG were collected at consecutive intervals of 5 days (5DAF, 10DAF, 15DAF, 20DAF, 25DAF, and 30DAF). Transmission electron microscopy (TEM) analyses were performed to gain insights into the submicroscopic structural changes in the tissues. The fresh tissues were fixed with Servicebio's fixative solution, followed by fixation in 1% osmic acid and 0.1 M phosphate-buffered saline (pH 7.4) for 5 h. Specimens were then dehydrated with alcohol, embedded in SPI-Pon 812 epoxy resin monomer, thinly sectioned, stained with uranyl acetate and lead citrate, and examined under a transmission electron microscope (model HT7700; Hitachi, Japan).

Detection of endogenous plant hormone levels during seed abortion

The levels of endogenous plant hormones, such as ABA, ZR, indole-3-acetic acid (IAA), and gibberellin A3 (GA3), were determined using enzyme-linked immunosorbent assay (ELISA). Each sample was measured in parallel three times, except for the time point 15 days after flowering (DAF) with four replicates.

Briefly, the tissues stored in liquid nitrogen were thawed and homogenized at a temperature of 2–8 °C. Following the manufacturer's instructions of ELISA kit provided by SHUANGYING Biological Ltd. (Shanghai, China), standards and samples were added to the wells of microtiter plates and incubated for 30 min at 37 °C. A total volume of 50 µL of horseradish peroxidase (HRP)-labeled detection antibody was then added and mixed gently, followed by incubation for 30 min at 37 °C. The 3,3',5,5'-tetramethylbenzidine dihydrochloride (TMB) was used as a substrate, and the reaction between the hormones (in standard and samples) and HRP was stopped with a reaction inhibitor within 15 min. The absorbance of each hormone was measured at a wavelength of 450 nm. After obtaining the hormone concentration in the sample, the hormone content in the sample was calculated.

RNA extraction and sequencing

Three biological replicates were collected for each stage. Total RNA at each stage was extracted according to the manufacturer's instructions for the Trizol reagent kit (Invitrogen, Carlsbad, CA, USA). Eukaryotic mRNA with a polyA tail was enriched from total RNA using oligo (dT) coupled to magnetic beads and fragmented. The first strand of cDNA was synthesized with M-MuLV reverse transcriptase, using fragmented mRNA as a template and random oligonucleotides as primers. The mRNA template chain was subsequently degraded by RNase H, and the second strand of cDNA was synthesized using dNTPs and DNA polymerase I. The double-stranded cDNA was purified with the QiaQuick PCR extraction kit (Qiagen, Venlo, Netherlands) and subjected to end repair, dA tailing, and adaptor ligation. cDNA fragments of approximately 200 bp were obtained using AMPure XP beads for PCR amplification, and the PCR product was purified using the AMPure XP system.

The quality and quantity of the library were analyzed using a NanoPhotometer® spectrophotometer (IMPLEN, CA, USA), Qubit® RNA Assay Kit with a Qubit® 2.0 Fluorometer (Life Technologies, CA, USA), and Agilent 2100 bioanalyzer and RNase free agarose gel electrophoresis. Based on the final libraries, a paired-end sequencing strategy (PE150) was carried out on the BGISEQ-500 platform (BGI Technology, Shenzhen, China).

Global and differential gene expression analysis

After removing adapters, low-quality reads, and ambiguous reads from the raw data, the clean reads were aligned to the rRNA reference sequences (no mismatches allowed) using bowtie2 (version 2.2.8) [50]. The rRNA-depleted reads were mapped to the reference genome of *S. grosvenorii* using HISAT 2. 2.4 [51] with “-rna-strandness RF” and other parameters set as a default. Assembly of the aligned reads was performed using StringTie software according to the reference gff file [52, 53].

The number of reads mapped to each transcript was counted for each sample and then normalized as fragments per kilobase of transcripts per million fragments mapped (FPKM). The Pearson’s correlation coefficients of RNA expression between samples were calculated and principal component analysis was performed using R version 3.6.3. Differential expression analysis was performed using DEseq2 [54], and the genes with $|\log_2FC| > 1$ and false discovery rate (FDR) < 0.05 were determined as differentially expressed genes (DEGs) between pairs of adjacent stages.

Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses of the DEGs were performed. All DEGs were mapped to GO terms in the GO database (<http://www.geneontology.org/>) and KEGG database for the pathways [55,56]. Gene numbers were calculated for every term/pathway, and significantly enriched GO terms/KEGG pathways in DEGs compared to the genome background were defined using a hypergeometric test. The calculated p-values were FDR corrected, and GO terms and KEGG pathways with $FDR \leq 0.05$ were defined as significantly enriched GO terms and KEGG pathway respectively.

Temporal analysis

Time-series analysis was performed using a short time-series expression miner (STEM) [57] (p.191). Each mRNA was assigned to the model profile based on the expression pattern based on the correlation coefficient. The number of mRNAs assigned to each model profile and expected to be assigned to a profile were calculated by randomly permuting the original time point values, renormalizing the mRNA expression values, assigning mRNAs to their most closely matching model profiles, and repeating this process for a large number of permutations. The statistical significance of the number of mRNAs assigned to each profile compared with the expected number was then computed.

Seed abortion associated gene co-expression network analysis

To explore the relationship between gene co-expression modules and plant hormone levels, weighted gene co-expression network analysis (WGCNA) was conducted using the WGCNA package in R (3.2.2.)

[58] (p. 559). The genes with reads per kilobase per million (RPKM) > 0.3 were used for WGCNA [59], and modules were obtained using the automatic network construction function, blockwiseModules, with default settings, except that the power was 18, minModuleSize was 50, and mergeCutHeight was 0.7. A module-trait relationship analysis was also performed using the module eigengene and the levels of ABA and ZR.

For genes in each module, GO and KEGG pathway enrichment analyses were conducted to analyze the biological functions of modules as described above. In addition, the network connections among the most connected genes (topological overlap above the threshold of 0.15) for the modules with statistical significance were further visualized using Cytoscape [60].

Abbreviations

ABA: abscisic acid; ZR: trans-zeatin; LHG: Luohanguo; GA: gibberellin; TEM: Transmission electron microscopy; DAF: days after flowering; IAA: indole-3-acetic acid; GA3: gibberellin A3; ELISA: enzyme-linked immunosorbent assay; FDR: false discovery rate; FPKM: fragments per kilobase of transcripts per million fragments mapped; DEGs: differentially expressed genes; GO: Gene Ontology; STEM: short time-series expression miner; KEGG: Kyoto Encyclopedia of Genes and Genomes; LTP: lipid-transfer protein 2-like; IRL: isoflavone reductase-like protei; CYP: cytochrome P450; ERF: ethylene-responsive transcription factor; NFR5: Nod factor receptor 5; BSR2: broad-spectrum Resistance 2; TPS4: Trehalose-6-phosphatase synthase S4; WGCNA: weighted gene co-expression network analysis.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Raw data for this project have been submitted to the Sequence Read Archive (SRA) database of the National Center for Biotechnology Information (Accession number: PRJNA773651).

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by the National Natural Sciences Foundation of China (81860678, 81973413, and 31660084), the National Key Research and Development Program of China (2017YFC1700706), Science and Technology Major Project of Guangxi (Gui Ke AA19254025), the Guangxi Natural Science Foundation Program of China (2019GXNSFAA245100), Guilin scientific research and technology development plan project (20190208-3), Guangxi plant functional substances research and utilization Key Laboratory Project (ZRJJ2020-8).

Authors' contributions

R.W. and D.T. organized and wrote this manuscript. R.W., D.T. and Xy. H. prepared the plant materials and detected the endogenous plant hormone. Z.L., N.C., J.X. and Xh. H. prepared the RNA and helped analysis the sequenced data. F.X., H.Y. and X.M. edited the manuscript. All authors reviewed the manuscript.

Acknowledgements

We thank Xiangjun Jiang (Guilin yiyuansheng modern biotechnology Co., Ltd) for providing the seedlings of *S. grosvenorii* line seedless 1.

References

1. Lu A, Zhang Z. The genus *Siraitia Merr.* in China. *Guihaia*. 1984; 4: 27–33.
2. Zhu Z. A brief analysis of the asexual propagation and planting techniques of *Siraitia grosvenorii*. *Inner Mongolia Forestry*. 2016; 02: 26–27.
3. Zhang H, Li X. Research progress on chemical compositions of Fructus Momordicae. *J. Anhui Agri. Sci.* 2011; 39: 4555–4556, 4559.
4. Qing Z, Zhao H, Tang Q, Mo C, Huang P, Cheng P, Yang P, Yang X, Liu X, Zheng Y, et al. Systematic identification of flavonols, flavonol glycosides, triterpene and siraitic acid glycosides from *Siraitia grosvenorii* using high-performance liquid chromatography/quadrupole-time-of-flight mass spectrometry combined with a screening strategy. *J. Pharm. Biomed. Anal.* 2017; 138: 240–248.
5. Chu D, Yaseen A, Wang L, Chen B, Wang M, Hu, W, Li F. Two New Cucurbitane Glycosides from the Fruits of *Siraitia grosvenorii* Swingle. *Chem. Pharm. Bull.* 2019; 67: 721–724.
6. Zeng Y, Pei W.; Zhu Y, Chen P, Sun Y. Research progress on the flavor, physiological functions and application of 3 kinds of natural sweeteners. *Journal of Food Safety and Quality*. 2019; 10: 38–45.
7. Soejarto DD, Addo EM, Kinghorn AD. Highly sweet compounds of plant origin: From ethnobotanical observations to wide utilization. *J. Ethnopharmacol.* 2019; 243: 112056.
8. Li D, Zhang H. Studies and uses of Chinese medicine Luohanguo-A special local product of Guangxi. *Guihaia*. 2000; 20: 270–276.
9. Liu H, Wang C, Qi X, Zou J, Sun Z. Antiglycation and antioxidant activities of mogroside extract from *Siraitia grosvenorii* (Swingle) fruits. *J. Food Sci. Technol.* 2018; 55: 1880–1888.

10. Chen G, Liu C, Meng G, Zhang C, Chen F, Tang S, Hong H, Zhang C. Neuroprotective effect of mogrol against Abeta1-42 -induced memory impairment neuroinflammation and apoptosis in mice. *J. Pharm Pharmacol.* 2019; 71: 869–877.
11. Abdel-Hamid M, Romeih E, Huang Z, Enomoto T, Huang L, Li L. Bioactive properties of probiotic set-yogurt supplemented with *Siraitia grosvenorii* fruit extract. *Food Chem.* 2020; 303: 125400.
12. Nie J, Yan K, Sui L, Zhang H, Zhang H, Yang X, Lu S, Lu K, Liang X. Mogroside V improves porcine oocyte *in vitro* maturation and subsequent embryonic development. *Theriogenology.* 2020; 141: 35–40.
13. Zhang Y, Zhou G, Peng Y, Wang M, Li X. Anti-hyperglycemic and anti-hyperlipidemic effects of a special fraction of Luohanguo extract on obese T2DM rats. *J. Ethnopharmacol.* 2020; 247: 112273.
14. Li C, Lin LM, Sui F, Wang Z, Huo H, Dai L, Jiang T. Chemistry and pharmacology of *Siraitia grosvenorii*: a review. *Chin. J. Nat. Med.* 2014; 12: 89–102.
15. Wang Q, Qin H, Wang W, Qiu S. The pharmacological research progress of *Siraitia grosvenorii*. *J. Guangxi Tradit. Chin. Med. Univ.* 2010; 13: 75–76.
16. Jiang X, Jiang X, Qin J, Huang X, Liu F. Preliminary study on selection of seedless *Siraitia grosvenorii*. *Guihaia.* 2009; 29: 506-509.
17. Kang G, Li J, Gao J, Fu S, Zhang N, Wu L, Xiang Z. Advance in the Germplasm Resources Creation of *Siraitia grosvenorii*. *Anhui Agri. Sci. Bull.* 2017; 23: 127-128.
18. Park S, Harada JJ. Arabidopsis embryogenesis. *Methods Mol. Biol.* 2008; 427: 3–16.
19. Peris CI, Rademacher EH, Weijers D. Chapter one-green beginnings—Pattern formation in the early plant embryo. *Curr. Top. Dev. Biol.* 2010; 91: 1–27.
20. Smit ME, Weijers D. The role of auxin signaling in early embryo pattern formation. *Curr. Opin. Plant Biol.* 2015; 28: 99–105.
21. Dong T, Park Y, Hwang I. Abscisic acid: biosynthesis, inactivation, homeostasis and signalling. *Essays Biochem.* 2015; 58: 29–48.
22. Bassel GW. To Grow or not to grow? *Trends Plant Sci.* 2016; 21: 498–505.
23. Hauser F, Li Z, Waadt R, Schroeder JI. SnapShot: Abscisic Acid Signaling. *Cell.* 2017; 171: 1708–1708 e1700.
24. Binenbaum J, Weinstain R, Shani E. Gibberellin Localization and Transport in Plants. *Trends Plant Sci.* 2018; 23: 410–421.
25. Alvarez JM, Bueno N, Cuesta C, Feito I, Ordas RJ. Hormonal and gene dynamics in de novo shoot meristem formation during adventitious caulogenesis in cotyledons of *Pinus pinea*. *Plant Cell Rep.* 2020; 39: 527–541.
26. Matilla AJ. Auxin: Hormonal Signal Required for Seed Development and Dormancy. *Plants.* 2020; 9: 705.
27. Laux T, Wurschum T, Breuninger H. Genetic regulation of embryonic pattern formation. *Plant Cell.* 2004; 16 Suppl1: 190–202.

28. Ambrose MJ, Wang TL, Cook SK, Hedley CL. An analysis of seed development in *Pisum sativum* L. IV. Cotyledon cell population in vitro and in vivo. *J. Exp. Bot.* 1987; 38: 1909–1920.
29. Corke FMK, Hedley CL, Wang TL. An analysis of seed development in *Pisum sativum* XI. Cellular development and the position of storage protein in immature embryos grown in vivo and in vitro. *Protoplasma.* 1990; 155: 127–135.
30. Heim U, Weber H, Baumlein H, Wobus U. A sucrose-synthase gene of *Vicia faba* L.: expression pattern in developing seeds in relation to starch synthesis and metabolic regulation. *Planta* 1993; 191: 394–401.
31. Giroux MJ, Boyer C, Feix G, Hannah LC. Coordinated transcriptional regulation of storage protein genes in maize endosperm. *Plant Physiol.* 1994; 106: 713–722.
32. Weber H, Buchner P, Borisjuk L, Wobus U. Sucrose metabolism during cotyledon development of *Vicia faba* L. is controlled by the concerted action of both sucrose-phosphate synthase and sucrose synthase: Expression patterns, metabolic regulation and implications on seed development. *Plant J.* 1996; 9: 841–850.
33. Weber H, Heim U, Golombek S, Borisjuk L, Manteuffel R, Wobus U. Expression of a yeast-derived invertase in developing cotyledons of *Vicia narbonensis* alters the carbohydrate state and affects storage functions. *Plant J.* 1998; 16: 163–172.
34. Andersen MN, Asch F, Wu Y, Jensen CH, Naested H, Mogensen VO, Koch KE. Soluble invertase expression is an early target of drought stress during the critical, abortion-sensitive phase of young ovary development in maize. *Plant Physiol.* 2002; 130: 591–604.
35. Smeekens S. Sugar-Induced Signal Transduction in Plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 2000; 51: 49–81.
36. Chen M. The experimental study of protective effects of anti-oxidation damage on some tissues of the flavones in *Siraitia grosvenorii* leaf on exhaustive swimming in rat. Master's degree thesis of Guangxi Normal University, Guilin, Guangxi, April, 2008.
37. Flores S, Smart CC. Abscisic acid-induced changes in inositol metabolism in *Spirodela polyrrhiza*. *Planta.* 2000; 211: 823–832.
38. Norlina R, Norashikin MN, Loh SH, Aziz A, Cha TS. Exogenous Abscisic Acid Supplementation at Early Stationary Growth Phase Triggers Changes in the Regulation of Fatty Acid Biosynthesis in *Chlorella vulgaris* UMT-M1. *Appl. Biochem. Biotechnol.* 2020; 191: 1653–1669.
39. Chong L, Guo P, Zhu Y. Mediator Complex: A Pivotal Regulator of ABA Signaling Pathway and Abiotic Stress Response in Plants. *Int. J. Mol. Sci.* 2020; 21(20):1-11.
40. Boba A, Kostyn K, Kozak B, Wojtasik W, Preisner M, Prescha A, Gola E.M, Lysh D, Dudek B, Szopa J, et al. *Fusarium oxysporum* infection activates the plastidial branch of the terpenoid biosynthesis pathway in flax, leading to increased ABA synthesis. *Planta.* 2020; 251: 50.
41. Tetali SD. Terpenes and isoprenoids: a wealth of compounds for global use. *Planta.* 2019; 249: 1–8.
42. Burlat V, Oudin A, Courtois M, Rideau M, St-Pierre B. Co-expression of three MEP pathway genes and geraniol 10-hydroxylase in internal phloem parenchyma of *Catharanthus roseus* implicates

- multicellular translocation of intermediates during the biosynthesis of monoterpene indole alkaloids and isoprenoid-derived primary metabolites. *Plant J.* 2004; 38: 131-41.
43. Metraux JP, Burkhart W, Moyer M, Dincher S, Middlesteadt W, Williams S, Payne G, Carnes M, Ryals J. Isolation of a complementary DNA encoding a chitinase with structural homology to a bifunctional lysozyme/chitinase. *Proc. Natl. Acad. Sci. U S A.* 1989; 86: 896–900.
 44. Samac DA, Shah DM. Developmental and Pathogen-Induced Activation of the Arabidopsis Acidic Chitinase Promoter. *Plant Cell.* 1991; 3: 1063–1072.
 45. Zhang X, Wu X, Findley S, Wan J, Libault M, Nguyen HT, Cannon SB, Stacey G. Molecular evolution of lysin motif-type receptor-like kinases in plants. *Plant Physiol.* 2007; 144: 623–636.
 46. Maeda S, Dubouzet JG, Kondou Y, Jikumaru Y, Seo S, Oda K, Matsui M, Hirochika H, Mori M. The rice CYP78A gene BSR2 confers resistance to *Rhizoctonia solani* and affects seed size and growth in Arabidopsis and rice. *Sci. Rep.* 2019; 9: 587.
 47. Marconi G, Pace R, Traini A, Raggi L, Lutts S, Chiusano M, Guiducci M, Falcinelli M, Benincasa P, Albertini E. Use of MSAP markers to analyse the effects of salt stress on DNA methylation in rapeseed (*Brassica napus* var. *oleifera*). *PLoS One.* 2013; 8: e75597.
 48. García-Garrido JM, Menossi M, Puigdoménech P, Martínez-Izquierdo JA, Delseny M. Characterization of a gene encoding an abscisic acid-inducible type-2 lipid transfer protein from rice. *FEBS Lett.* 1998; 428: 193–199.
 49. Takei K, Yamaya T, Sakakibara H. *Arabidopsis* CYP735A1 and CYP735A2 encode cytokinin hydroxylases that catalyze the biosynthesis of trans-Zeatin. *J. Biol. Chem.* 2004; 279: 41866–41872.
 50. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 2012; 9: 357–359.
 51. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 2015; 12: 357–360.
 52. Perteau M, Kim D, Perteau GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* 2016; 11: 1650–1667.
 53. Perteau M, Perteau G.M, Antonescu C.M, Chang T.C, Mendell J.T, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 2015; 33: 290–295.
 54. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014; 15(12): 550.
 55. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000; 25(1): 25–29.
 56. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000; 28: 27–30.

57. Ernst J, Bar-Joseph Z. STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics*. 2006; 7: 191.
58. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008; 9(1): 559.
59. Bai Y, Dougherty L, Cheng L, Zhong G.Y, Xu K. Uncovering co-expression gene network modules regulating fruit acidity in diverse apples. *BMC Genomics*. 2015; 16: 612.
60. Shannon P. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003; 13: 2498–2504.

Figures

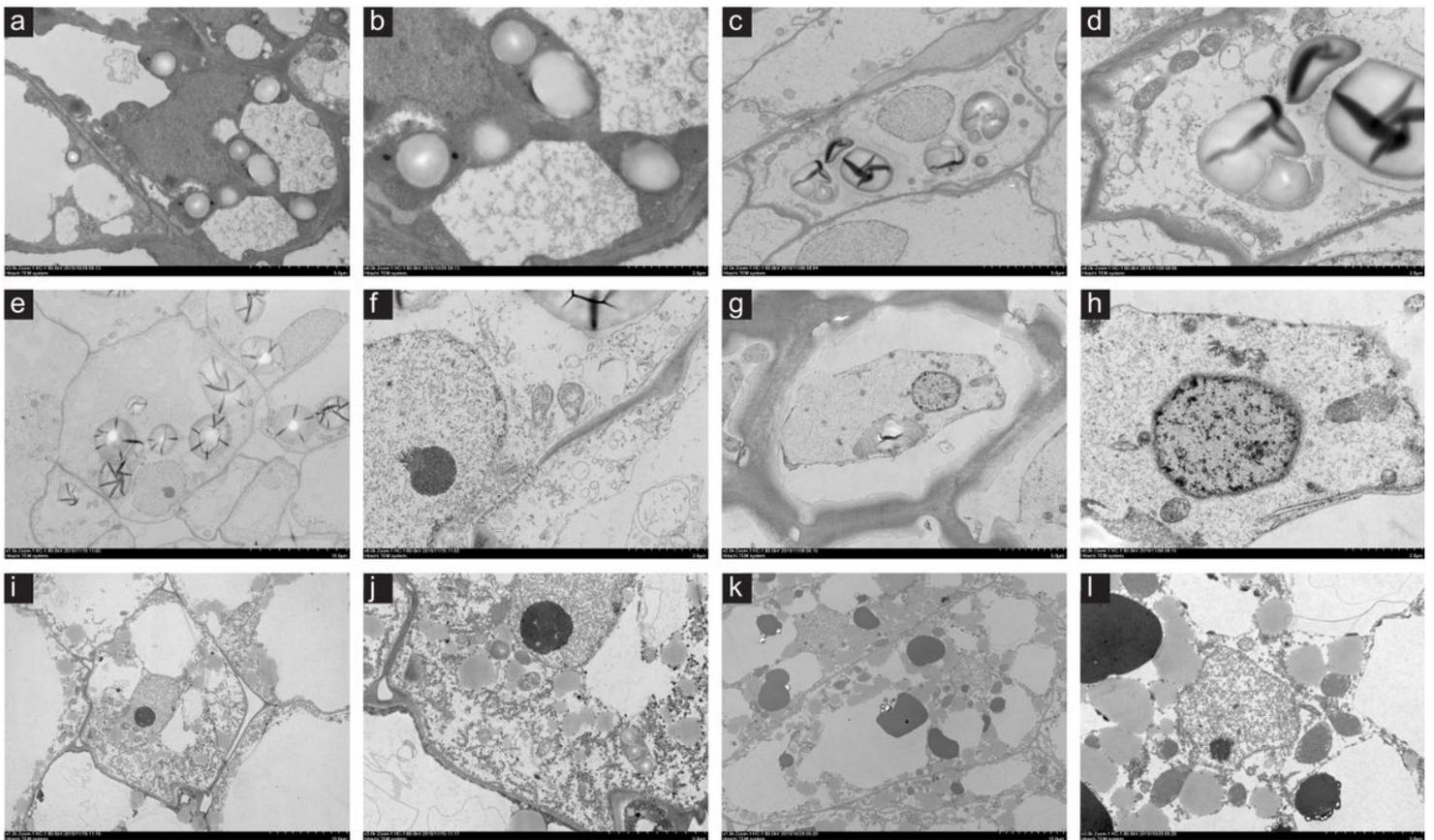


Figure 1

Transmission electron microscopy (TEM) images of the tissues with structural changes during seed abortion. The original (a, c, e, g, i, and k) and zoomed-in (b, d, f, h, j, and l) TEM images are shown for different stages, namely 5DAF (a-b), 10DAF (c-d), 15DAF (e-f), 20DAF (g-h), 25DAF (i-j), and 30DAF (k-l).

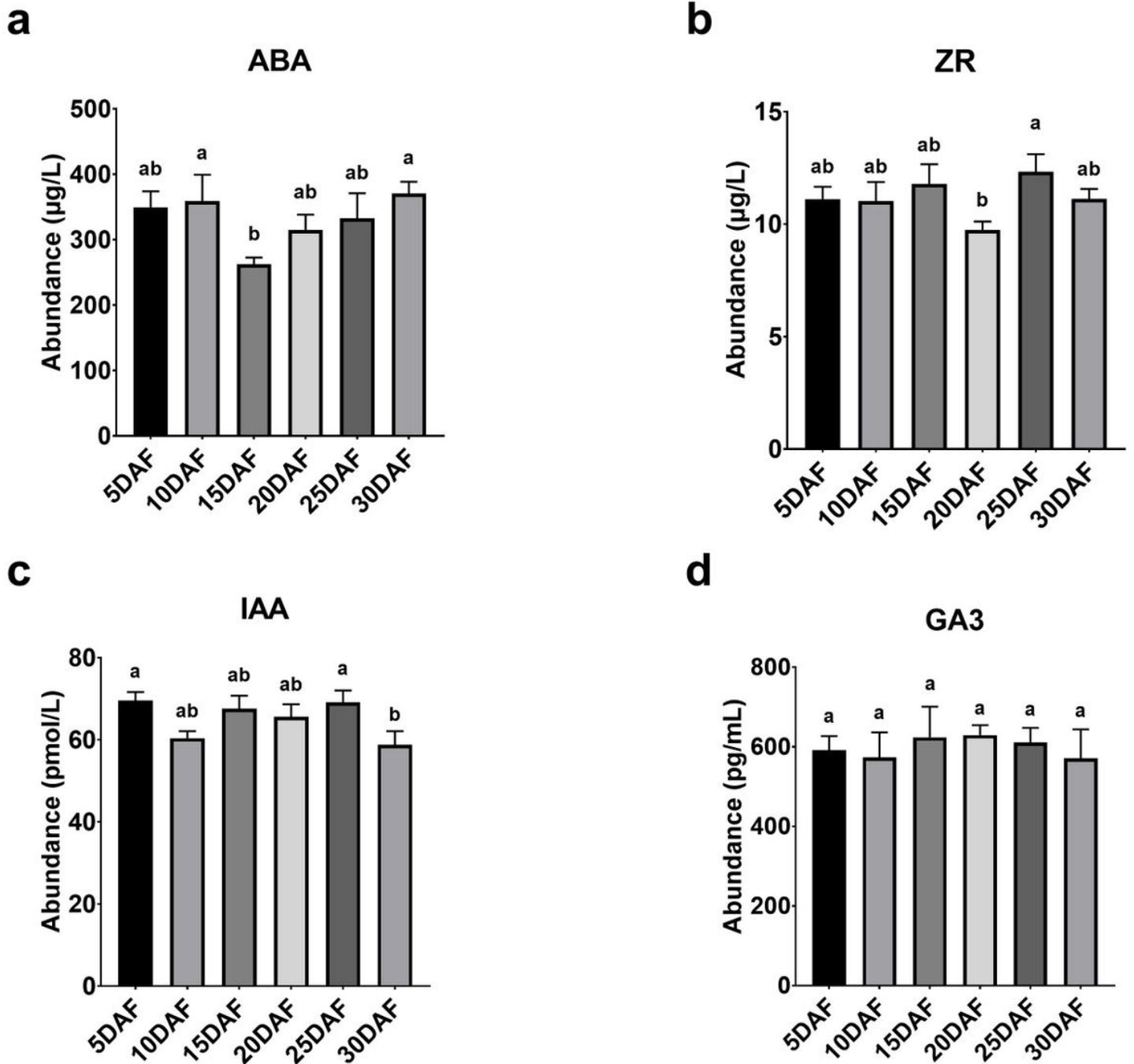


Figure 2

The levels of endogenous plant hormone during seed abortion: ABA (a), ZR (b), IAA (c), and GA3 (d). ABA, abscisic acid; ZR, trans-zeatin-riboside; IAA, indole-3-acetic acid; GA3, gibberellin A3. Data are presented as error bar plots showing the mean \pm SE of three/four replicates.

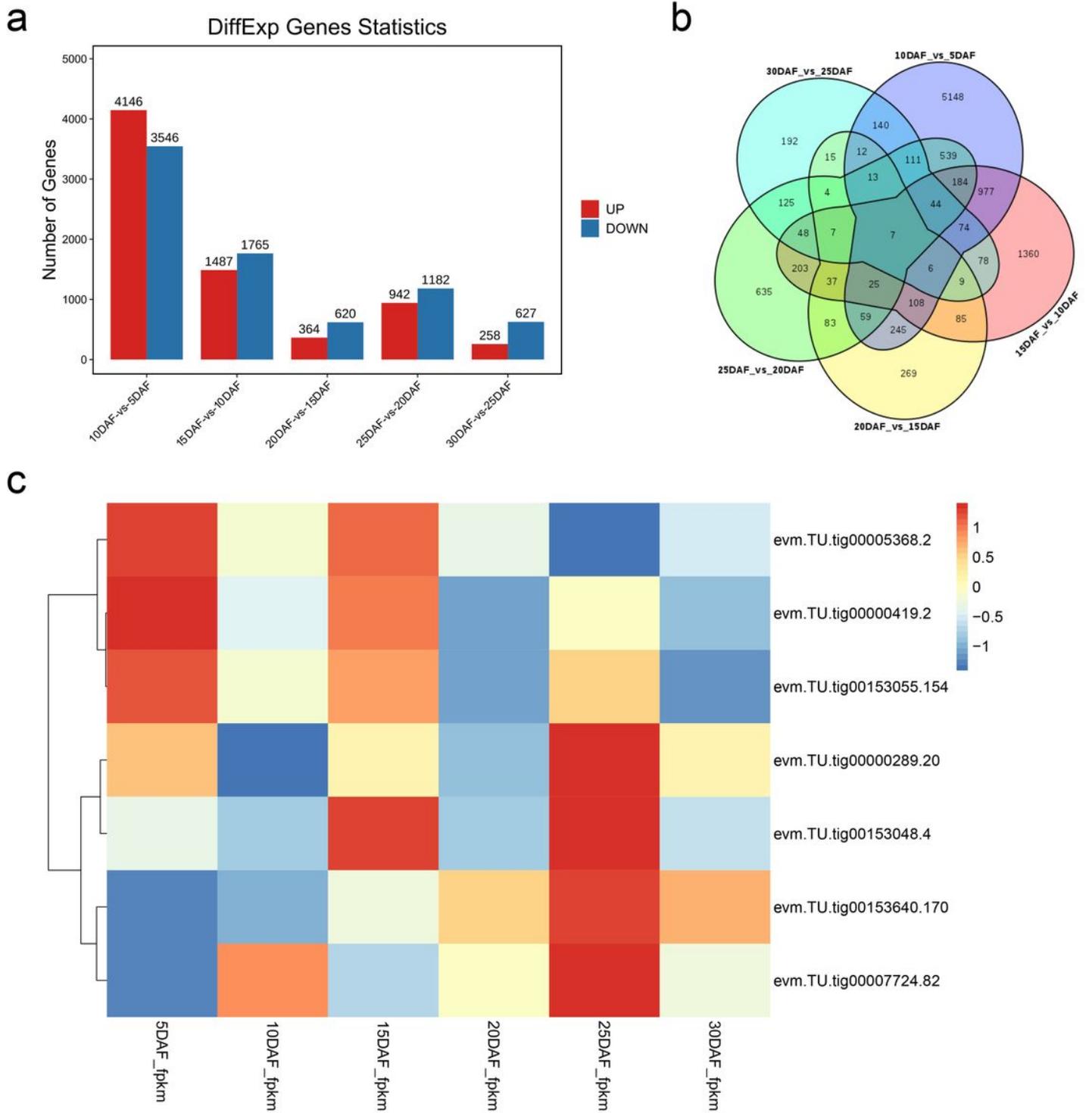


Figure 3

Differentially expressed genes in five comparisons between consecutive stages. (a) Barplot representing the number of upregulated (red) and downregulated (blue) genes between consecutive stages. (b) Venn diagram of differentially expressed genes. (c) Heatmap showing 7 differentially expressed genes among all comparisons in b.

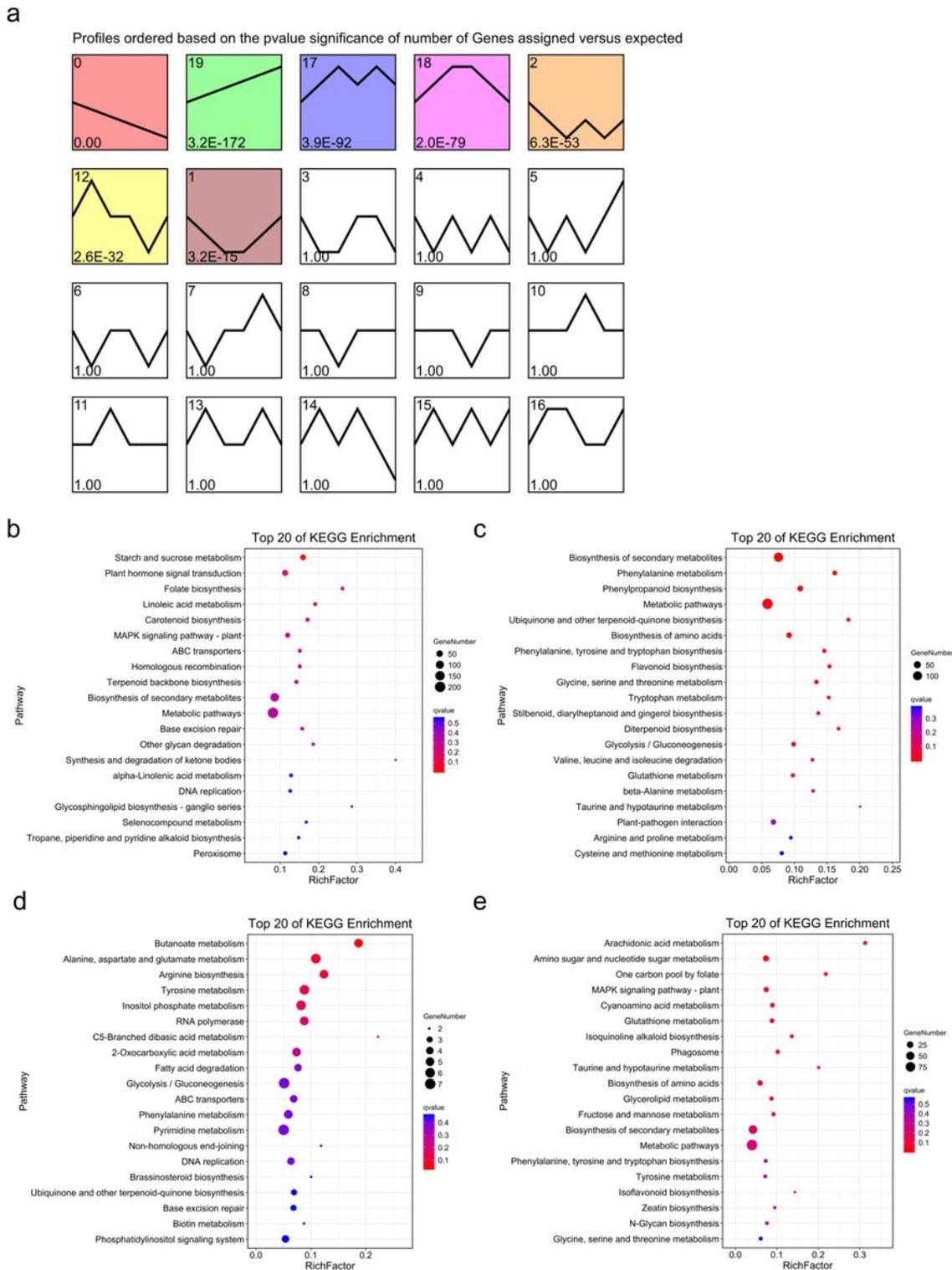


Figure 4

Temporal expression trends during seed abortion. (a) STEM analysis of the gene expression profiles. Each box indicates a model profile, and the colored profiles shown are significant. The numbers in the box provide the order of the profile (upper left) and the p-value indicates significance (lower left). (b-e) Top 20 pathways from KEGG enrichment analysis for the genes in profiles 18 (b), 0 (c), 19 (d), and 1 (e) in (a).

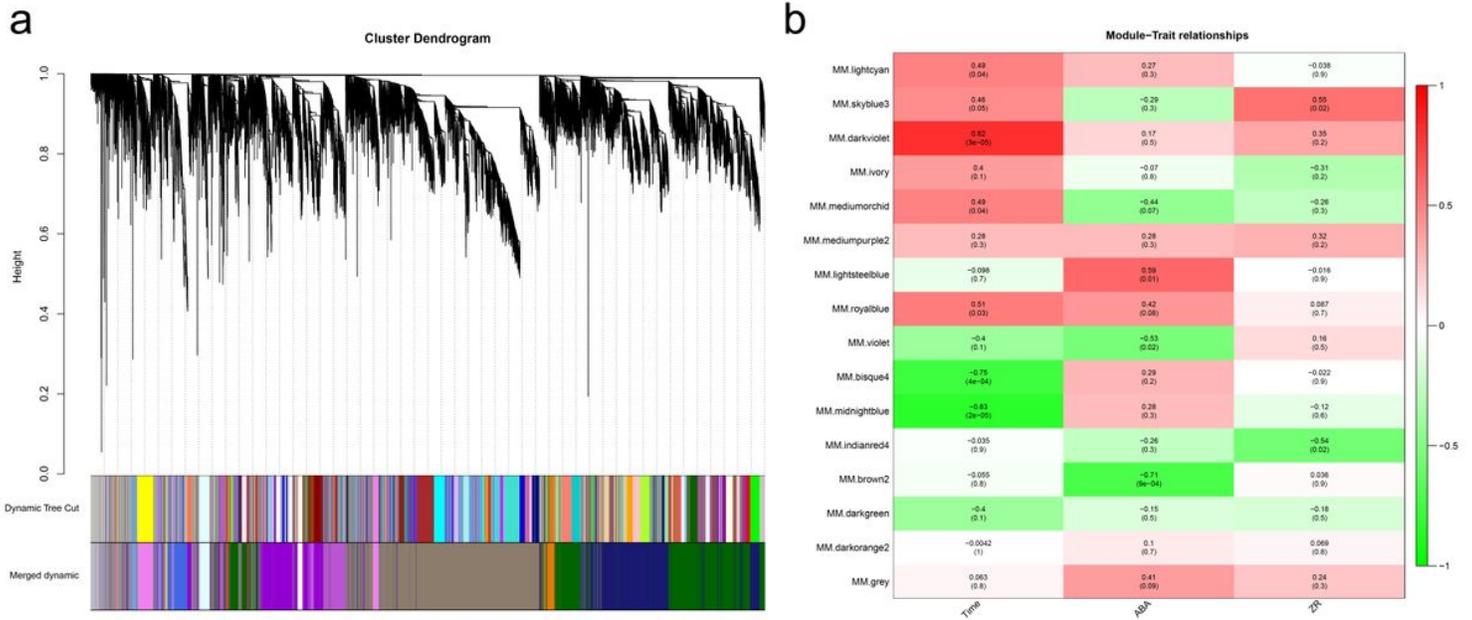


Figure 5

WGCNA and characteristics of the modules with different colors. (a) Hierarchical cluster tree displaying the co-expression modules. (b) The association between the module eigengene E and the traits of development stages and levels of ABA and ZR, respectively. The correlation coefficient is presented in each row, and the FDR-adjusted P-value is shown in parentheses.

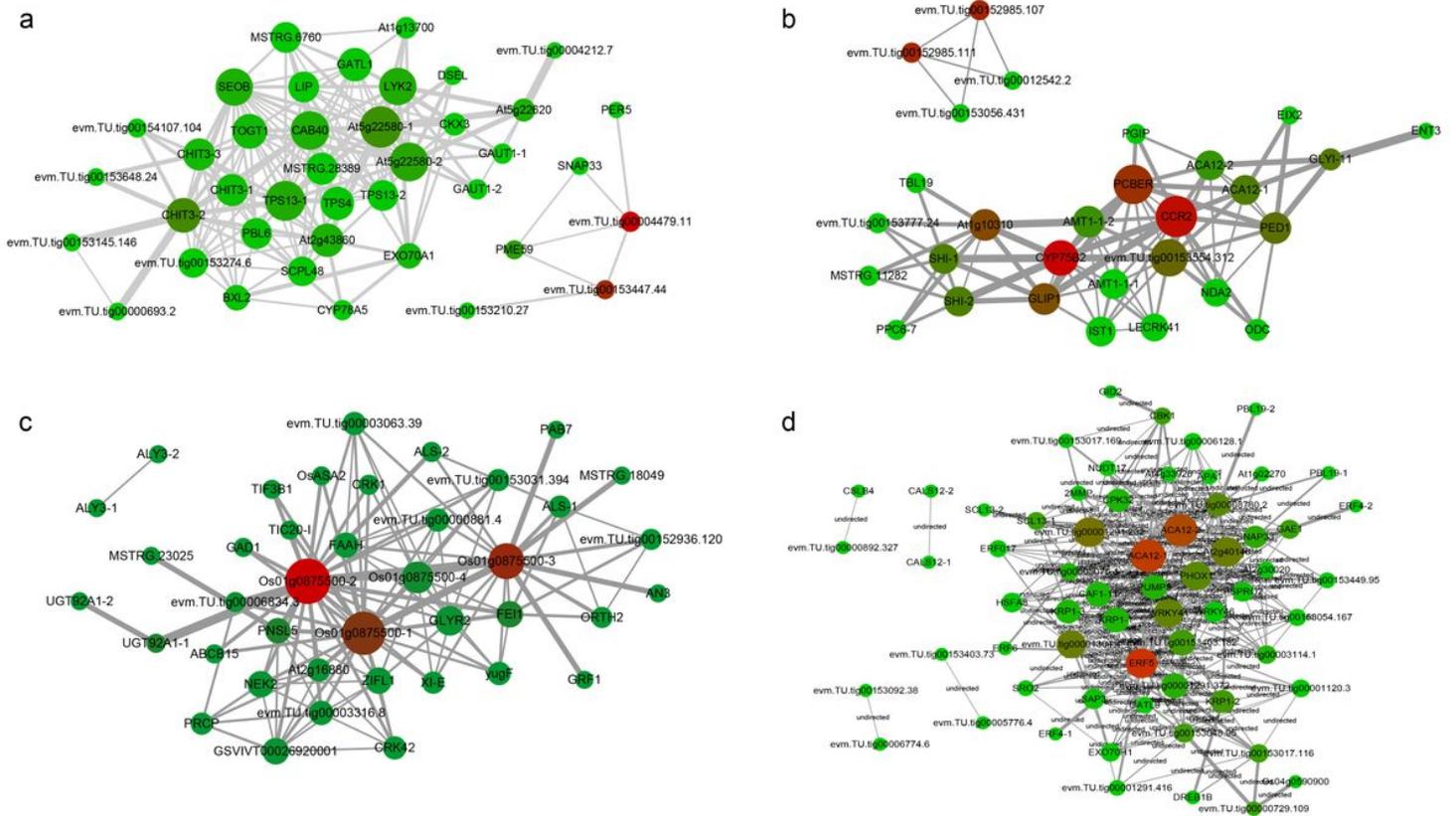


Figure 6

(a) Expression profiles of the WGCNA modules. (b) Top 20 pathways from GO enrichment analysis for the genes in the brown2 module. (c) Top 20 pathways from KEGG enrichment analysis for the genes in the brown2 module.

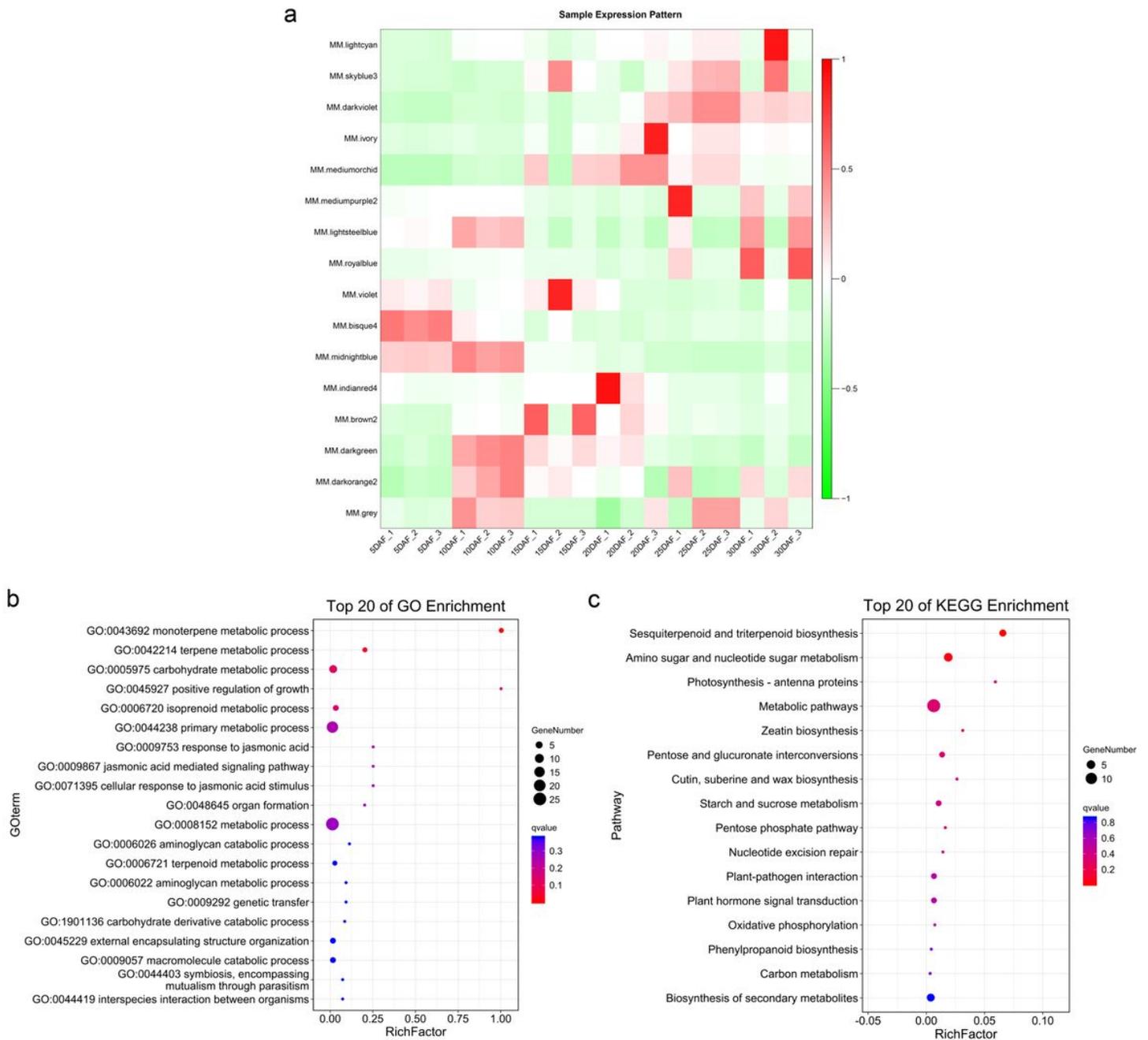


Figure 7

Visualization of the network connections among the most connected genes in the brown2 (a), indianred4 (b), lightsteelblue (c), and skyblue3 (d) modules. The size and red intensity of the nodes are correlated with degree and betweenness centrality, respectively.

Supplementary Files

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

- [TableS1.xls](#)
- [TableS2.xls](#)
- [TableS3.xls](#)
- [TableS4.xls](#)
- [TableS5.xls](#)