

Transcriptomics and metabolomics analyses provide insights into plant resistance of *Alsophila spinulosa*

Wei-Cheng Yang

Guizhou Normal University

Qin-Qin He

Guizhou Chishui *Alsophila* National Nature Reserve Administration

Yu Jiang (✉ jiangyu20090820@163.com)

Guizhou Normal University

Li-Jun Zhang

Guizhou Normal University

Jia-Xing Xiao

Guizhou Normal University

Jiao Yang

Guizhou Normal University

Bing-Jie Che

Guizhou Normal University

Bing-Cheng Zhang

Guizhou Normal University

Hang-Dan Chen

Guizhou Normal University

Jiang Li

Biozeron Shenzhen, Inc

Research Article

Keywords: Metabolomics, Transcriptomics, Cyatheaceae, *Alsophila spinulosa*, *Gymnosphaera metteniana* var. *subglabra*

Posted Date: August 31st, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3234090/v1>

License: (cc) (i) This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background:

Tenthredinidae feeds on *Alsophila*. This study aimed to detect special metabolites that attract pests.

Metabolomics and transcriptomics analyses were used on the *Alsophila spinulosa* and *Gymnosphaera metteniana* var. *subglabra* samples.

Results:

The results of transcriptome changes indicate that differentially expressed genes (degrees) are involved, such as genes expressing kaempferol-3-O-glucoside-7-O-rhamnoside. It has a wide range of expression patterns in plants and may lead to the color and morphology of leaves of the *A. spinulosa* and *G. metteniana* var. *subglabra*.

Conclusions:

These results are helpful to identify effective biological control methods and provide a scientific basis for maintaining a good habitat for *Alsophila* spp. in the nature reserve in later stages.

1. Background

Tree-ferns (Cyatheaceae) are one of the few remaining tree-shaped ferns on Earth. Fossil evidence of tree ferns suggests that they first appeared in the Early Jurassic or Late Triassic, and were widely distributed and flourished in the middle of the Mesozoic. There are currently six genera and approximately 500 species of tree ferns, primarily found in tropical and subtropical regions. Within China, there are two genera consisting of 14 species and 2 varieties, distributed in the southwestern and southern regions of the country [1].

Tree ferns, including *A. spinulosa* and *G. metteniana* var. *subglabra*, have attracted much attention for their many utilization values. Given the ancient and significant nature of tree ferns, investigating their geology, paleoclimatology, and ancient ecology is of great importance. These plants are intermediate between lower plants and higher plants, providing an important reference value for studying the evolution of ancient vegetation and the phylogeny of ferns. Additionally, tree ferns are tall and beautiful, and crown like a giant umbrella with high ornamental value. Another species of tree fern, known as *Cyathea*, is also of great importance due to its high medicinal value for treating ailments such as rheumatism, coughs, and bone strength, among others [2, 3]. Flavonoids D01 isolated from the stem leaves of this species, have been found to possess anti-tumor activity and can efficiently attenuate chemotherapy drugs [4]. Moreover, vitexin, a compound obtained from the stem, has antibacterial properties, while apigenin exhibits anti-tumor, anti-inflammatory, antibacterial, antiviral, and cardiovascular protective effects [5, 6]. Despite their numerous utilitarian values and aesthetic appeal, populations of *A. spinulosa* have been drastically declining over the last few decades due to habitat loss, fragmentation, and human activities.

In China, these ferns are rare, with wild populations and individual specimens restricted to tropical and subtropical montane regions that occupy warm, humid, and shady niches at low latitudes. As a result, the species is now listed in the Chinese Red Book of endangered species [7].

A. spinulosa, *G. metteniana* var. *subglabra* and its varieties mainly show in plant morphology, leaves, petioles, branches and distribution area: tree fern stem up to 6 meters or higher, diameter 10–20 cm, tree-like, thin leaves, deep split, brown and barbed branches, mostly distributed in the valley and forest near water and hidden. *G. metteniana* var. *subglabra* and its varieties are significantly shorter (most about 2 meters), while *G. metteniana* var. *subglabra* have thick leaves, wax layer, and shallow crack, petioles, branches dark, which are mostly distributed in humid dense forest or stone crevices. At present, the most serious thing is that *A. spinulosa* and *G. metteniana* var. *subglabra* are endangered by plant-eating insects, and more than one phytophagous insect have different sizes of insect pests in different seasons, which affect the normal growth and development of plants. There are insect pests of different sizes in different seasons, affecting the normal growth and development of plant [8, 9]. According to the field observation, the host insects showed different harm degree and different distribution on the host plants on the two plants. The statistical analysis found that the insect mouth density was significantly different between the two plants. It is widely believed that there are two theories, that cause of this phenomenon may be caused by adult spawning. Theory 1, the spawning choice of parental female adults must first be conducive to the survival, growth and development of individual offspring. The referred-effectiveness theory (Preference-performance theory), the spawning choice of parental female adult insects is not only related to the growth and development fitness of the offspring, but also restricted by the host defense response, intraspecific competition, parental fitness and natural enemies and other factors, to choose the host plants conducive to their own nutrition. "The plant vigor hypothesis" believes that herbivorous adults choose large and robust plants to lay eggs because of their high biomass and nutritional value, and choose host plants conducive to their own nutrition ". The feeding experience of larvae of herbivore insects can affect the adult spawning, also known as "feeding marks". *Rhoptroceros cyatheae* and *Phthonoloba viridifasciata* were studied to explore the feeding preference. Results analysis shows that, *R. cyatheae*, first-instar worms have no feeding preference for *A. spinulosa* and *G. metteniana* var. *subglabra*, which may be weak and have no choice for feeding objects. However, second-instar worms clearly prefer tree-ferns (further supplemented by data, excluding data chance), and 3–5 years old, and become more obvious as age increases. *P. viridifasciata*, the 1st-4th year old obviously prefers *G. metteniana* var. *subglabra*, and with the gradual increase of age, the preference phenomenon is more obvious. So what are the ingredients of *A. spinulosa* and *G. metteniana* var. *subglabra* that influencing the behavior?

Aowadays, the research focuses on tree fern diversity and community characteristics in typical growth areas, chemical composition analysis of different parts, DNA extraction and ISSR analysis, cultivation and breeding techniques, and prediction of tree fern potential geographical distribution area and area in the context of climate change. Ruan et al [10] developed 15 SSR markers using FIASCO method, four of the SSR markers can be used for the genetic analysis of Hainan tree fern population. Using morphological features and ITS sequence analysis methods by Liu et al [11] to study the endophytic

fungal diversity in Chishui, Guizhou. A total of 84 isolates were obtained from 250 tissue blocks, they belongs to 2 phyla, 3 classes, 10 orders, 18 families and 25 genera. Among them, Ascomycota is the dominant bacteria, which relative abundance is as high as 92.86%. In addition, from the perspective of different organizational diversity, the Shannon, Simpson diversity index and Pielou uniformity of tree ferns are higher than the leaves. Based on morphological characteristics and DNA sequence analysis techniques, the diversity, community composition and distribution of endophytes on tree ferns were discussed by Zang et al. According to the characteristic values of the fungal colonization rate, abundance, diversity, similarity, and fungal species in the petiole litter, the distribution characteristics of fungi in the surface-falling tree ferns were significantly changed. Excitingly, the genome of *A. spinulosa* has been released, thereby establishing a vital foundation for future investigations [12]. Furthermore, the chloroplast genome of *A. spinulosa* has been made available for study [13]. Most notably, the full-length transcriptome of this species has been comprehensively reported, enabling scientists to gain a deeper understanding of its biological processes and functions [14]. With these significant advancements in genomic research, it is now possible to explore the unique characteristics of *A. spinulosa* at a molecular level.

At present, more and more reports study the relationship between plants and insects at home and abroad. Plants produce some volatile secondary substances in the metabolic process, which affect insect behavior and attract increasingly more attention as an important part of plant defense. Plant resistance (R) genes are critical components of plant defense mechanisms against various biotic stresses, including insect pests [15]. The identification of these genes and their functions in different plant species has contributed greatly to our understanding of plant-insect interactions. In sorghum, two R genes have been implicated in defense against greenbug aphids, highlighting their importance in protecting plants from insect damage [16]. In recent years, there has been increasing recognition of the value of joint analysis of metabolome and transcriptome data in biological research [17, 18]. This approach offers greater sensitivity and specificity in identifying gene expression levels, gene functions, and metabolome data. As a result, it has the potential to provide a more accurate and comprehensive understanding of the complexity of biological systems. By combining transcriptomics and metabolomics data, researchers can obtain a more comprehensive understanding of the molecular mechanisms underlying plant defense and insect resistance. Overall, the joint analysis of transcriptome and metabolome data represents an important tool in the study of plant-insect interactions and has the potential to contribute significantly to the development of more effective strategies for pest management in agriculture.

This research aimed to elucidate the molecular characteristics of *A. spinulosa* through the integration of transcriptomics and metabolomics data derived from leaf tissue. Additionally, this study represents the first investigation into the R genes of this species. By employing a multi-omics approach, we were able to comprehensively analyze the molecular mechanisms underlying the physiological processes of *A. spinulosa*. Specifically, we identified key transcripts and metabolites associated with various biological pathways, shedding light on the plant's response to stress and adaptation to its environment. Moreover, our investigation into the R genes of *A. spinulosa* provides a significant contribution to the understanding of the plant's immune response. By characterizing these genes, we have gained valuable insights into the

mechanisms that enable the plant to recognize and defend against pathogenic invaders. Overall, this study not only advances our knowledge of the molecular biology of *A. spinulosa*, but also provides a foundation for further research in plant genomics and functional genomics.

The experimental team first used untargeted metabolomics technology, based on PCA score, to select 111 and other different metabolites. By comparing *A. spinulosa* and *G. metteniana* var. *subglabra*, all differential metabolites were divided into 97 metabolic pathways, which can be classified into four categories, metabolism, genetic information processing, environmental information processing and human diseases, among which the largest category was metabolism (58.31%). Different secondary metabolites were selected, they were aligned to the KEGG database, and the 20 most significant differential metabolic pathways were selected to draw the enrichment classification map. The results showed that there were 54 significantly different metabolites between *A. spinulosa* and *G. metteniana* var. *subglabra*. Among them, flavonoids, phenolic acids, lipid acids, organic acids, tannins and other classes contain more compounds, with 16, 15, 8, 4, 4, and 3 kinds, respectively (Yang et al., unpublished data). Metabolomics is considered the closest approach to a biological phenotype, and our team attempted to integrate and analyze the metabolomics and transcriptomic results. Through transcriptome sequencing technology (RNA-seq) to identify the regulatory mechanism of the key enzyme genes between *A. spinulosa* and *G. metteniana* var. *subglabra*. The NCBI database was used as the reference genome database to excavate the enzyme genes synthesized from the secondary metabolites of tree ferns, and the different genes between them were selected and analyzed. It lays a foundation for the in-depth study of the medicinal function identification, phenylpropane pathway and flavonoid synthesis pathway.

2. Materials and methods

2.1. Experimental materials and tissue collection

The leaves of *A. spinulosa* (S-L-1) and *G. metteniana* var. *subglabra* (H-L-1) were collected and deposited at the herbarium of GZNU. Test materials (Chishui sign) were obtained from the nature reserve. Three blades of *A. spinulosa* and *G. metteniana* var. *subglabra* were washed with 75% alcohol. All samples were frozen in liquid nitrogen and stored at -80 °C for sequencing analysis. After sample preparation, transcriptomics and metabolomics analysis was performed. Three biological replicates were performed for each pericarp sample.

2.2. Library preparation for transcriptomics sequencing

Total RNA was used as input material for the RNA sample preparations. Briefly, mRNA was purified from total RNA by using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in First Strand Synthesis Reaction Buffer (5X). First strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase, then use RNaseH to degrade the RNA. Second strand cDNA synthesis was subsequently performed using DNA Polymerase I and dNTP. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of 3' ends of DNA fragments, Adaptor with hairpin loop structure were ligated to prepare

for hybridization. In order to select cDNA fragments of preferentially 370~420 bp in length, the library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, USA). Then PCR amplification, the PCR product was purified by AMPure XP beads, and the library was finally obtained. The cDNA library was sequenced by 2 X 151 bp paired-end sequencing on an Illumina NovaSeq 6000 sequencing platform.

2.3. Quantification of gene expression level and differential expression analysis

Raw reads were trimmed by fastp (v 0.18.0) [19] to remove adapter contamination and reads with high uncertainty ($N > 10\%$) or low base quality were discarded. With the reference genome of *A. spinulosa* [20], the quantification of gene expression level and differential expression analysis methods were as published in previous studies [21].

2.4. GO and KEGG enrichment analysis of differentially expressed genes

The Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis were using clusterProfiler [22].

2.5. Identification and analysis of differential metabolites

The Q1, Q3, retention time, declustering potential (DP) and collision energy (CE) were used for metabolite identification. The SCIEX OSV1.4 software was used to open the downtime mass spectrum file, and the chromatographic peaks were integrated and corrected. The chromatographic peaks were screened according to the minimum peak height of 500, signal-to-noise ratio of 5, smoothing number of 1 and other information. The peak area of each chromatographic peak represents the relative content of the corresponding substance. Finally, the integral data of all chromatographic peaks are derived to obtain the qualitative and quantitative results of metabolites [23]. These metabolites were annotated using the KEGG database[24], HMDB database (<http://www.hmdb.ca/>) and Lipidmaps database (<http://www.lipidmaps.org/>). The condition of differential metabolites was variable importance in the projection (VIP) value of the PLS-DA model ≥ 1 and independent sample t-test's p-value ≤ 0.05 .

2.6. Correlation between genes and metabolites

The correlation between genes and metabolites was using 'cor.test' under R software (v 3.5.1). And we focused on the results of the absolute value of Pearson correlation coefficient ≥ 0.9 with p-value ≤ 0.01 .

2.7. Identification of R genes

The RGAugury pipeline [25] was used to screen the entire gene set for resistance gene analog (RGA) prediction. The default *P*-value cutoff for initial RGA filtering was set to $1e-5$ for BLASTP. Protein-protein interaction (PPI) network was analyzed using the Search Tool for the Retrieval of Interacting Gene (STRING) database [26], which included direct and indirect associations of proteins. The hub gene was calculated using cytoHubba [27].

2.8. Validation of gene expression using qRT-PCR

We performed reverse transcription assays of total RNAs using the PrimeScript first-strand cDNA synthesis kit (Takara, Dalian, China) and designed gene-specific primers with Primer6. qRT-PCR experiment was carried out using SYBRGreen Fast qPCR Master Mix (Sangon Biotech, China) on an ABI StepOne Plus Real-Time System (ABI, USA), following the manufacturer's instructions. XsACTIN was used as the internal reference gene to normalize the qRT-PCR expression data. The $2^{-\Delta\Delta Ct}$ method [28] was employed to calculate the relative mRNA abundance in each sample for every gene. We also estimated the Pearson correlation coefficient between the qRT-PCR and RNA-seq gene expression profiles in R version 3.6.3.

3. Results

3.1. Metabolome profiling of *Alsophila*

After collecting fresh leaves from *G. metteniana* var. *subglabra* and *A. spinulosa* (Fig. 1, A), metabolomics was used to analyze the metabolites. To improve the accuracy of analysis, each group designed three replicates. We identified 373 and 399 metabolites in positive and negative modes, respectively (Table S1). In *A. spinulosa*, the top 10 metabolites with highest metabolite abundance were "2-Caffeoyl-L-tartaric acid", "Isocitric Acid", "Kaempferol-3-O-glucoside-7-O-rhamnoside", "Kaempferol-3-O-neohesperidoside", "Skimmin", "Luteolin-6-C-glucoside", "Demethyl coniferin", "γ-Linolenic Acid", "Methylmalonic acid", and "Aromadendrin-7-O-glucoside". In *G. metteniana* var. *subglabra*, the top 10 metabolites with highest metabolite abundance were "Skimmin", "Isocitric Acid", "6-Methylmercaptapurine", "γ-Linolenic Acid", "α-Linolenic Acid", "Spermine", "Histidinol", "N-Benzylmethylene isomethylamine", "L-Glutamine" and "L-Lysine" (Fig. 1, B). The quantity of differential accumulated metabolites (DAMs) was statistically analyzed (Table S2). Clustering heat maps of DAMs showed that significant differences in 111 metabolites were found. Of these, 82 DAMs were up-regulated and 29 metabolites were dramatically down-regulated (Fig. 1, C). To verify the function of the DAMs, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment was conducted. The main metabolic pathways involved include "valine, leucine and isoleucine degradation", "pyruvate metabolism, 2-Oxocarboxylic acid metabolism", "valine, leucine and isoleucine biosynthesis", "citrate cycle (TCA cycle)", "biosynthesis of amino acids", "biosynthesis of secondary metabolites", "aminoacyl-tRNA biosynthesis", "glucosinolate biosynthesis", "glyoxylate and dicarboxylate metabolism", "ABC transporters", and so on (Fig. 1, D). Notably, the pathway with the most DAF enrichment pathways was the biosynthesis of secondary metabolites, which had 21 DAFs.

3.2. Transcriptomics analysis of *Alsophila*

Six cDNA libraries from fresh leaves of *G. metteniana* var. *subglabra* (H) and *A. spinulosa* (S) were constructed. This generated 231.57 Mb of clean data, of which 38.60 Mb was obtained for each sample (Q30 ≥ 92.6%). The sequencing raw data has been deposited into the Short Reads Archive (SRA) database under the accession number SRR24032773, SRR24032774, SRR24032775, SRR24033065,

SRR24033066, and SRR24033067. Sequence comparison between the reads of each variety and the reference genome indicated a similarity level of 67.54-89.30%. The mRNA abundance of each gene in each sample was profiled as fragment per kilo bases per million reads (FPKM). A total of 29,060 genes had an FPKM expression value above 1.0 in at least one sample (Fig. S1). The hierarchical cluster dendrogram showed the expression pattern of biological repeats be clustered together (Fig. S2). Furthermore, a total of 14,639 DEGs were identified in the H_vs_S comparison groups, with 7,764 upregulated genes and 6,875 downregulated genes (Fig. 2, A). Regarding 5 of the most-differentially upregulated genes, three genes had defined functions: indole-3-pyruvate monooxygenase YUCCA1 (*YUCCA1*), pleiotropic drug resistance protein 1 (*PDR1*), and cholesterol 22-hydroxylase CYP90B27 (*CYP90B27*). Correspondingly, all the 5 most-differentially down-regulated genes had defined functions: 14 kDa zinc-binding protein (*ZBP14*), expansin-A10 (*EXPA10*), protein MEN-8 (*MEN-8*), methyl jasmonate esterase 1 (*MJE1*), and mannose-specific lectin (*dfa*). In addition, results of KEGG analysis indicated that 38 pathways were enriched (Fig. 2, B, Table S3). Of which, DAMs were also significantly enriched in “pantothenate and CoA biosynthesis”, “stilbenoid, diarylheptanoid and gingerol biosynthesis”, “valine, leucine and isoleucine biosynthesis”. GO enrichment analysis revealed that the top 20 most significant GO categories were "regulation of transcription, DNA-templated", "protein phosphorylation", "carbohydrate metabolic process", "defense response", "transmembrane transport", "membrane", "transmembrane transporter activity", "copper ion binding", "iron ion binding", "ADP binding", "sequence-specific DNA binding", "ATPase activity", "monooxygenase activity", "protein kinase activity", "microtubule binding", "carbohydrate binding", "protein serine/threonine kinase activity", "DNA binding", "DNA-binding transcription factor activity", "oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular" (Fig. 2, C).

3.3. Identification of genes involved in disease resistance

Most of the proteins encoded by the characterized RGAs, including nucleotide-binding site (NBS)-containing proteins, receptor-like protein kinases (RLKs), and receptor-like proteins (RLPs), contain conserved domains, such as NBS, leucine-rich repeat (LRR), and Toll/interleukin-1 receptor (TIR) [29]. Using a genome-wide scanning pipeline [25], a total of 1,290 RGAs were identified in the *A. spinulosa* genome (Table S4). 62.2% (803) genes of the RGAs belonged to the RLK category, and there were 110 NBS-related RGAs, of which 25 were of the TIR type. 839 RGAs were found to be expressed in leaf tissue (Fig. S3). Of which, there were 606 RGAs overlapped with DEGs, including 426 RLK-encoding RGAs (Table 2). Ten of the most-differentially genes were as follows: disease resistance protein L6 (*L6*), receptor-like protein kinase At1g49730 (*At1g49730*), disease resistance protein RUN1 (*RUN1*), leucine-rich repeat receptor-like serine/threonine-protein kinase BAM1 (*BAM1*), leucine-rich repeat receptor-like protein kinase At5g63930 (*At5g63930*), salt tolerance receptor-like cytoplasmic kinase 1 (*OsL_16820*), G-type lectin S-receptor-like serine/threonine-protein kinase SD2-5 (*SD25*), wall-associated receptor kinase 2 (*WAK2*), disease resistance protein RPS2 (*RPS2*) and L-type lectin-domain containing receptor kinase VIII.1 (*LECRK81*). Notably, of the differentially expressed 606 RGAs, 551 genes could be further supported by known resistance genes reference from the latest PRGdb [30]. In KEGG enrichment analysis, these genes enriched in “Plant-pathogen interaction” and “MAPK signaling pathway - plant” pathway. In the GO

enrichment analysis, these genes were assigned to 16 GO terms, such as "protein kinase activity," "ATP binding," "protein phosphorylation," "defense response," "protein binding," and more (Table S5). Gene network plays an important role in various organisms and systems, which effectively help us reveal the essential rules of a large number of biological processes and reactions in organisms. Using these 551 genes' protein sequences as queries, 250 interactions were identified by STRING database (Table S6). The first four key genes that calculated by sytoHubba with MCC algorithm were *BAM1*, inactive leucine-rich repeat receptor-like protein kinase At5g48380 (*BIR1*), receptor-like protein kinase BRI1-like 3 (*BRL3*), and leucine-rich repeat receptor-like serine/threonine/tyrosine-protein kinase SOBIR1 (*SOBIR1*) (Fig. 3, A). The maximum-likelihood phylogenetic tree showed that the closest relative genes of *BAM1* were LRR receptor-like serine/threonine-protein kinase FLS2 (*FLS2*) and receptor-like kinase TMK (TMK) (Fig. 3, B). PPI network also revealed that *BAM1* had interaction with *FLS2* and *TMK* (Table S6). We inferred these genes may play essential role in the molecular mechanism of insect resistance in *Alsophila*.

3.4. Correlation between DEGs and DAMs associated with R genes

The expression levels of 14,097 DEGs were significantly correlated with the abundance of 111 DAMs. Further analysis showed that 585 significantly differentially expressed RGAs were significantly positively correlated with 111 DAMs, while 574 significantly differentially expressed RGAs were significantly negatively correlated with 111 DAMS (Table S7, S8). Subsequently, interaction networks were constructed for significantly positively correlated RGAs between DAMs, and significantly negatively RGAs between DAMs, respectively. According to the results of cytoHubba, in the significantly positively correlated network, the top 10 core genes are probably inactive leucine-rich repeat receptor-like protein kinase At5g48380 (*BIR1*), receptor-like protein kinase BRI1-like 3 (*BRL3*), *BAM1*, phytoalkaline kinase receptor 1 (*PSKR1*), leucine-rich repeat protein kinase family protein (*AT1G27190*), leucine-rich receptor-like protein kinase family protein (*At5g46330*), *SOBIR1*, lysin-containing receptor-like kinase 1 (*LYK1*, also named as *CERK1*), leucine-rich repeat transmembrane protein kinase family protein (*AT1G68400*), and Leucine-rich repeat protein kinase family protein (*AT1G67510*). The first 10 core metabolites are N-Benzylmethylene isomethylamine, 1-O-p-Coumaroylquinic acid, Neochlorogenic acid (5-O-Caffeoylquinic acid), Ferulic acid-4-O-glucoside, Scopoletin-7-O-glucuronide, 1-O-Feruloylquinic acid, 3-O-Feruloylquinic acid, Luteolin-4'-O-glucoside, Kaempferol-7-O-glucoside, and Kaempferol-4'-O-glucoside. In the significantly negatively correlated network, the top 10 genes in the core are *BIR1*, *BRL3*, *BAM1*, *At5g46330*, *AT1G27190*, *SOBIR1*, Leucine-rich receptor-like protein kinase family protein (*PSY1R*), STRUBBELIG-receptor family 8 (*SRF8*), LRR receptor-like serine/threonine-protein kinase HSL2 (*HSL2*), and *PSKR1*. The first 10 core metabolites are Demethyl coniferin, Luteolin-8-C-glucoside (Orientin), Aromadendrin-7-O-glucoside, 4-O-(6'-O-Glucosylcaffeoyl)-3,4-dihydroxybenzoic acid, LysoPC 20:5, Procyanidin B2, Kaempferol-3-O-glucoside-7-O-rhamnoside, Kaempferol-3-O-neohesperidoside, Quercetin-7-O-rutinoside, and Cinnamtannin B1 (Fig. 4A, Fig. 4B).

3.5. Verification of RNA-Seq gene expression

Real-time quantitative PCR (qRT-PCR) was used to validate the results and expression profiles of the genes identified via Illumina sequencing analysis. We selected 10 DEGs, all of which were RGAs, including wall-associated receptor kinase-like 1, leucine-rich repeat receptor-like serine/threonine-protein kinase BAM2, *FLS2*, and others (Table S9, S10). All 10 RGAs displayed the same expression pattern, with the Pearson correlation coefficient of fold change between the PCR experiment and RNAseq being 0.78 (Fig. S4). These results verify the reliability and accuracy of our transcriptome analysis.

4. Discussion

The production of certain volatiles (green leaf volatiles, terpenes) released by plants, whether pre-formed or induced, directly contribute to defense and play an important role in the indirect defense strategies adopted by plants [31]. We detected kaempferol-3-O-glucoside-7-O-rhamnoside that significantly increased in *A. spinulosa*. It has a broad expression pattern and is involved in vascular strand development in the leaf, control of leaf shape, size and symmetry, male gametophyte development and ovule specification and function. Anthers of double mutants (*bam1bam2*) appeared abnormal at a very early stage and lack the endothecium, middle, and tapetum layers. We predicted that kaempferol-3-O-glucoside-7-O-rhamnoside might influence leaf development. Besides, Wang [32] investigated the antitumor, antioxidant and anti-inflammatory activities of kae, kae-7-O-glu, kae-3-O-rha and kae-3-O-rut. They demonstrated that kae has better antitumor activity, possibly because kae significantly inhibits AKT phosphorylation and caspase-3, caspase-7, caspase-9 and PARP cleavage while the other kae glucosides do not. Kae also demonstrated better antioxidant and anti-inflammatory activities. Therefore, the accumulation of kaempferol-3-O-glucoside-7-O-rhamnoside in *A. spinulosa* could be applied to clinical therapy.

Declarations

Acknowledgments

This work was supported by National Natural Science Foundation of China (30560024), the Qianlin Kehe (2022), No.27 grants from the Forestry research Project of Guizhou Province; Guizhou Science and Technology Support Plan Project (Guizhou Science Support [2017] No. 2811; Major Science and Technology Project of Guizhou Province (No. 3022-03 [2016])).

References

1. Wang XL, Luo WC. Current status of tree fern in China based on bibliometric analysis. Jiangsu Forestry Science and Technology. 2016;43(5):41-45
2. Liu HL. Current status of tree fern in China based on bibliometric analysis. Jinan University. 2012.
3. Zhang X, Sun JY, Yang Q. Study on the effect of cough and expectorant by tree ferns stem. Liaoning J Tradit Chin Med. 2018;45(7):1514-1515.

4. Tang X. Pharmacological effects of flavonoids DO1. Zhongshan University. 2004.
5. Gong JW, Chen FZ, Li SH. Preliminary study on the antibacterial activity of tree fern leaf and stems. Anhui Agricultural Science. 2007;33:10566+10568.
6. Guo S, Zhang WX, Yu W. Progress in the pharmacological effects of apigenin. Journal of Hubei University of Science and Technology (medical edition). 2016;30:273-276.
7. Zhou Y, Chen G P, Wang T. Isolation and Characterization of Microsatellite Loci in the Tree Fern *Alsophila spinulosa*. Am Fern J. 2008;98(1):42-45.
8. He QQ, Yang WC, Liang S, Zhang TY. Preliminary study on the pest of tree family in Chishui. Environmental Protection Technology. 2019;25(3):29-32+40.
9. Xu DS. Biology and ecology of major tree fern pests. Guizhou Normal University. 2020
10. Ruan X, Wang Z, Wang T, Su Y. Novel nuclear SSR markers in the large frond tree fern *Alsophila gigantea* and its congeneric species *Alsophila spinulosa*. Acta Scientiarum Naturalium Universitatis Sunyatseni. 2017;56(1):115-120.
11. Liu YI, Zhang LN, Liang L. Study on endophytic fungal diversity in the Nature Reserve in Chishui, Guizhou. J Microbiol. 2021;40(10):2673-2684.
12. Huang X Guo YH Tian L Li QZ Yue ZQ. The genome size determination of *Alsophila spinulosa* based on flow cytometry Journal of Hunan Agricultural University(Natural Sciences), 2020;46(2):171-175.
13. Ma Y, Jiao P, Qi Z, Jiang Z, Guan S. Characterization of the complete chloroplast genome of *Alsophila spinulosa*, an endangered species endemic to China. Mitochondrial DNA B. 2020;5(3):2262-2263.
14. Hong Y, Wang Z, Li M, Su Y, Wang T. First Multi-Organ Full-Length Transcriptome of Tree Fern *Alsophila spinulosa* Highlights the Stress-Resistant and Light-Adapted Genes. Front Genet. 2022;12:784546.
15. Kaloshian I. Gene-for-gene disease resistance: bridging insect pest and pathogen defense. J Chem Ecol. 2004;30:2419-2438.
16. Zhang H, Huang J, Huang Y. Identification and characterization of plant resistance genes (R genes) in sorghum and their involvement in plant defense against aphids. Plant Growth Regul. 2022;96(3):443-461.
17. Bai C, Wu C, Ma L, Fu A, Zheng Y, Han J, Li C, Yuan S, Zheng S, Gao L, Zhang X, Wang Q, Meng D, Zuo J. Transcriptomics and metabolomics analyses provide insights into postharvest ripening and senescence of tomato fruit under low temperature. Hortic Plant J. 2023;9(1):109-121.
18. Lu JK, Xu Y, Meng ZL, Cao M, Liu S, Kato-Noguchi H, Yu WW, Jin B, Wang L. Integration of morphological, physiological and multi-omics analysis reveals the optimal planting density improving leaf yield and active compound accumulation in Ginkgo biloba. Ind Crops Prod. 2021;172:114055.

19. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17):i884-i890.
20. Huang X, Wang W, Gong T, Wickell DA, Kuo LY, Zhang X, Wen JI, Kim H, Lu F, Zhao H, et al. The flying spider-monkey tree fern genome provides insights into fern evolution and arborescence. *Nat Plants*. 2022;8(5):500-512.
21. Sun C, Li J, Dong J, Niu Y, Hu J, Lian J, Tian YY, Shi Q, Ye X. Chromosome-level genome assembly for the largemouth bass *Micropterus salmoides* provides insights into adaptation to fresh and brackish water. *Mol Ecol Resour*. 2020;21(1):301-315.
22. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The innov*. 2021;2(3):100141.
23. Bai, C., C. Wu, L. Ma, A. Fu, Y. Zheng, J. Han, C. Li, S. Yuan, S. Zheng, L. Gao, X. Zhang, Q. Wang, D. Meng and J. Zuo . "Transcriptomics and Metabolomics Analyses Provide Insights into Postharvest Ripening and Senescence of Tomato Fruit Under Low Temperature." *Horticultural Plant Journal*. 2021; 9(1):13.
24. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27-30.
25. Li P, Quan X, Jia G, Xiao J, Cloutier S, You M. RGAugury: a pipeline for genome-wide prediction of resistance gene analogs (RGAs) in plants. *BMC genomics*. 2016,17(1):1-10.
26. Szklarczyk D, Gable A L, Lyon D, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2018;47:D607-D613.
27. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol*. 2014; 8(4):1-7.
28. Livak K J, Schmittgen T D. Analysis of relative gene expression data using real-time quantitative PCR and the 2– $\Delta\Delta$ CT method. *meth*. 2001;25(4):402-408.
29. van Ooijen G, van den Burg HA, Cornelissen BJC, Takken FLW. Structure and function of resistance proteins in solanaceous plants. *Annu Rev Phytopathol*. 2007;45:43-72.
30. Osuna-Cruz CM., Gallart AP, Donato AD, Sundesha V, Andolfo G, CiglianoRA, Sanseverino W, Ercolano MR. PRGdb 3.0: a comprehensive platform for prediction and analysis of plant disease resistance genes. *Nucleic Acids Res*. 2017;46:D1197-D1201.
31. Gatehouse JA. Plant resistance towards insect herbivores: a dynamic interaction. *New Phytol*. 2002; 156(2):145-169.
32. Wang J, Fang X, Ge L, Cao F, Zhao L, Wang Z, Xiao W. Antitumor, antioxidant and anti-inflammatory activities of kaempferol and its corresponding glycosides and the enzymatic preparation of kaempferol. *PLos One*. 2018;13(5):e0197563.

Figures

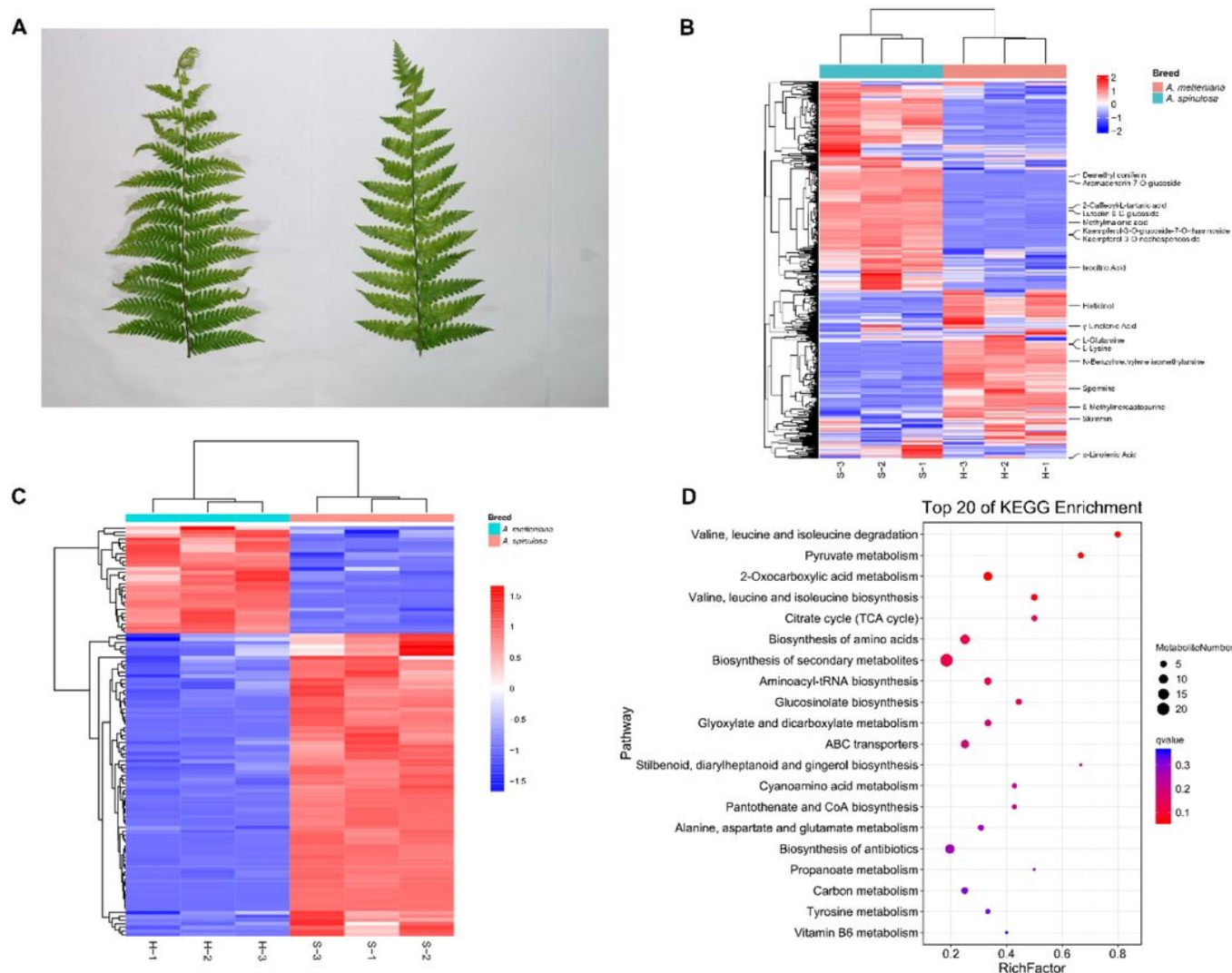


Figure 1

Metabolome profiling of *Alsophila*.

(A) The leaves of *A. spinulosa* and *G. metteniana* var. *subglabra*; (B) The heatmap of metabolites, each column represents a sample, and each row represents a metabolite; (C) The heatmap of DAMs, each column represents a sample, and each row represents a DAM. (D) Scatter plot for KEGG enrichment results. The rich factor is the ratio of DAMs numbers annotated in this pathway term to all metabolite numbers annotated in this pathway term. The greater the Rich factor, the greater the degree of pathway enrichment. A qvalue is the corrected p value ranging from 0 to 1, and a lower value indicates greater pathway enrichment.

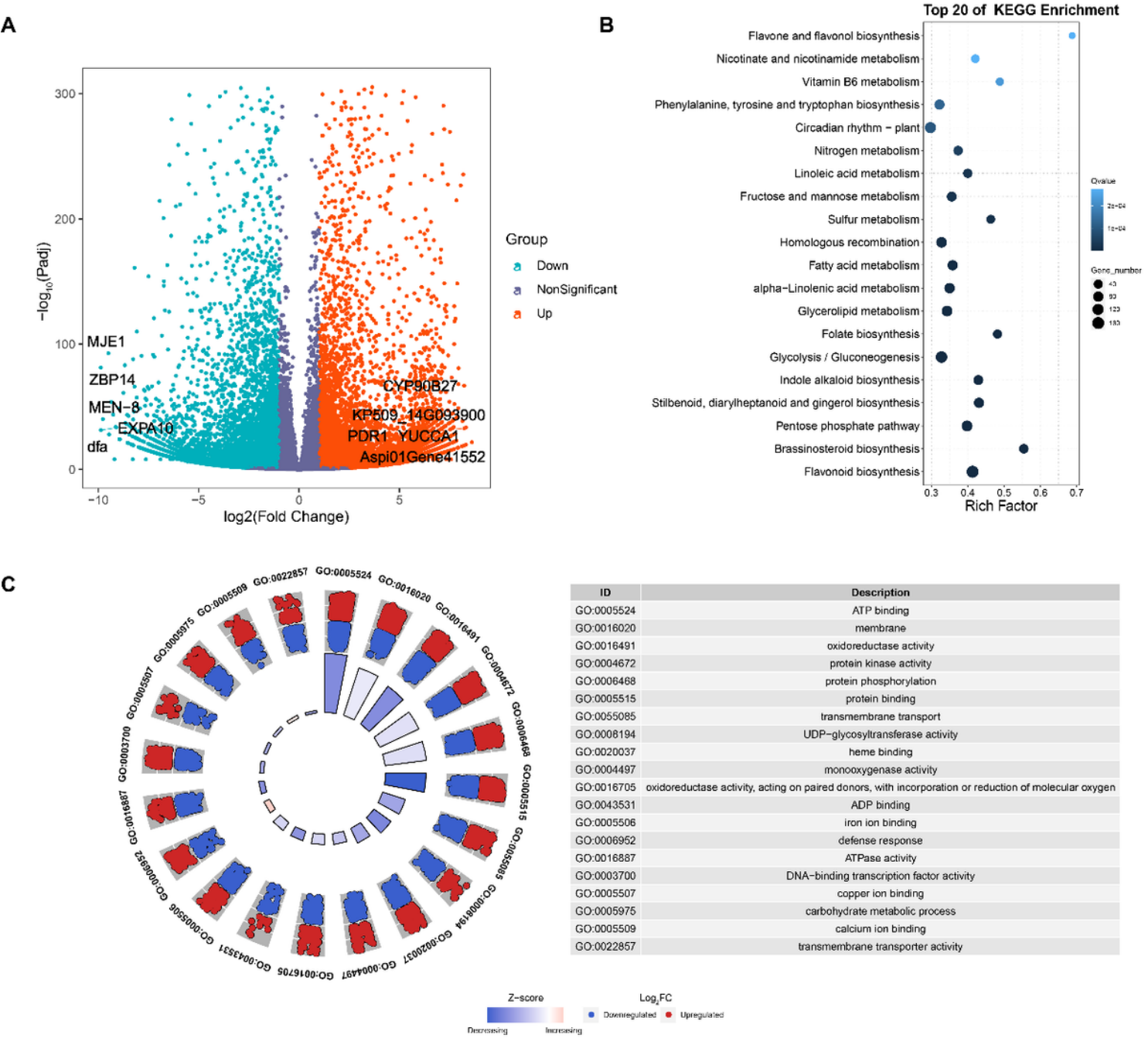


Figure 2

Transcriptomics analysis of *Alsophila*.

(A) Volcano plots illustrate the magnitude and significance of DEGs.

(B) Scatter plot for KEGG enrichment results. The Rich factor is the ratio of differentially expressed gene numbers annotated in this pathway term to all gene numbers annotated in this pathway term. The greater the Rich factor, the greater the degree of pathway enrichment. A Q value is the corrected p value ranging from 0 to 1, and a lower value indicates greater pathway enrichment. (C) GO enrichment circle plots for the top 20 most significant GO categories. The outer circle shows the relative fold change for each

significant RNA feature compared to air contributing to the GO term with blue dots showing down-regulated RNA features and red dots showing upregulated features. The z-score color scale represents the number of up-regulated genes minus the number of down-regulated genes for a given GO term divided by the square root of the total count. The associated tables present the GO term ID and function for the enriched term.

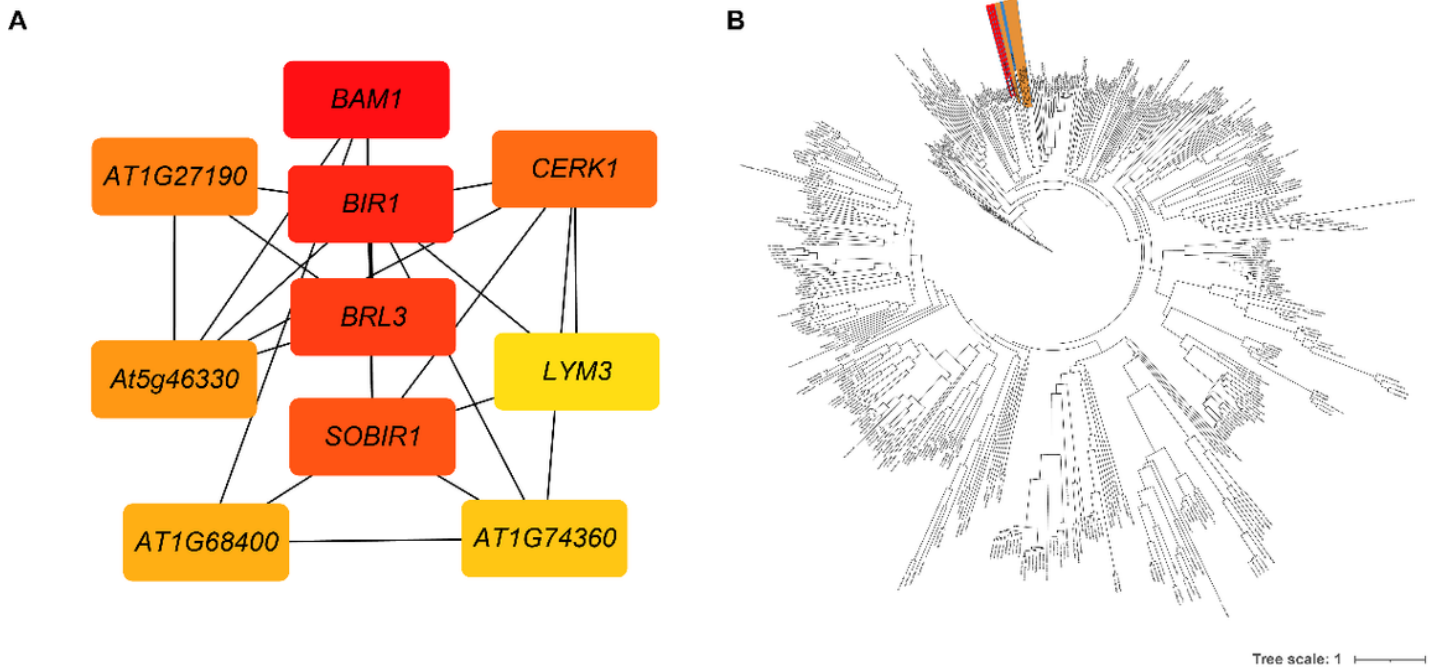


Figure 3

The interactions and evolutionary relationships of differentially expressed RGAs.

(A) The interaction network of top 10 nodes ranked by degree algorithm. The more forward ranking is represented by a redder color. (B) The maximum-likelihood phylogenetic tree of 551 RGAs. The *BAM* is highlighted in red, the *FLS2* in yellow, and the *TMK* in blue.

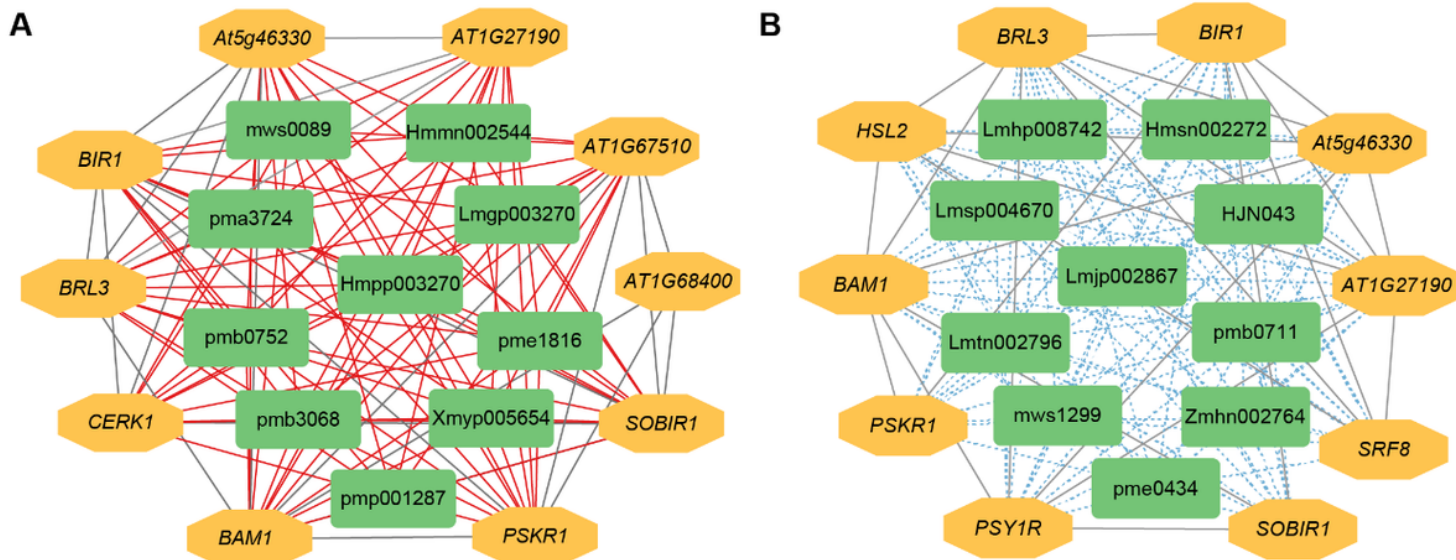


Figure 4

The interaction network between DEGs and DAMs associated with R genes.

(A) Positive regulatory network; (B) Negative regulatory network. A red line indicates a positive correlation, whereas a blue line indicates a negative correlation between compound content and gene expression.

pmp001287: N-Benzylmethylene, pmb3068: 1-O-p-Coumaroylquinic, pme1816: Neochlorogenic, Hmmn002544: Ferulic, Lmcp003270: Scopoletin-7-O-glucuronide, pma3724: 1-O-Feruloylquinic, pmb0752: 3-O-Feruloylquinic, Hmcp003270: Luteolin-4'-O-glucoside, mws0089: Kaempferol-7-O-glucoside*, Xmcp005654: Kaempferol-4'-O-glucoside, Hmsn002272: Demethyl, mws1299: Luteolin-8-C-glucoside, Lmtn002796: Aromadendrin-7-O-glucoside, Zmhn002764: 4-O-(6'-O-Glucosylcaffeoyl)-3,4-dihydroxybenzoic, Lmhp008742: LysoPC, pme0434: Procyanidin, Lmsp004670: Kaempferol-3-O-glucoside-7-O-rhamnoside, Lmcp002867: Kaempferol-3-O-neohesperidoside*, pmb0711: Quercetin-7-O-rutinoside, HJN043: Cinnamtannin

Supplementary Files

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

- [002SuppFigs.docx](#)
- [002SupplementaryTables.xls](#)