



OPEN ACCESS

EDITED BY

Boping Wu,
Zhejiang Agriculture and Forestry University,
China

REVIEWED BY

Meng Jiang,
Zhejiang University, China
Yun Zhao,
Chinese Academy of Sciences (CAS), China

*CORRESPONDENCE

Bingguang Xiao

✉ xiaobgsubmission@126.com

[†]These authors have contributed equally to this work

RECEIVED 21 April 2025

ACCEPTED 30 May 2025

PUBLISHED 24 June 2025

CITATION

Tong Z, Yang K, Chen X, Xu F, Sui X, Huang Y, Zhu S, Shen E, Wu S, Fan L and Xiao B (2025) Multiomic analysis of the synthetic pathways of secondary metabolites in tobacco leaves at different developmental stages. *Front. Plant Sci.* 16:1615756. doi: 10.3389/fpls.2025.1615756

COPYRIGHT

© 2025 Tong, Yang, Chen, Xu, Sui, Huang, Zhu, Shen, Wu, Fan and Xiao. 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.

Multiomic analysis of the synthetic pathways of secondary metabolites in tobacco leaves at different developmental stages

Zhijun Tong^{1†}, Kun Yang^{2†}, Xuejun Chen¹, Fei Xu², Xueyi Sui¹, Yujie Huang², Shenyun Zhu², Enhui Shen², Sanling Wu³, Longjiang Fan² and Bingguang Xiao^{1*}

¹Key Laboratory of Tobacco Biotechnological Breeding, Yunnan Academy of Tobacco Agricultural Sciences, Kunming, China, ²Institute of Crop Sciences & Institute of Bioinformatics, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, China, ³Analysis Center of Agrobiological and Environmental Sciences, Faculty of Agriculture, Life and Environment Sciences, Zhejiang University, Hangzhou, China

Introduction: *Nicotiana tabacum*, widely cultivated for its economic and scientific value, produces a broad range of secondary metabolites that play critical roles in determining leaf quality and flavor. Despite substantial progress, the comprehensive regulatory landscape governing secondary metabolite biosynthesis during *N. tabacum* leaf development remains largely unclear.

Methods: To better understand the molecular regulatory mechanisms underlying the biosynthesis of secondary metabolites, particularly flavonoids, during *N. tabacum* leaf development, we conducted a transcriptomic and non-targeted metabolomic sequencing and analysis at three critical developmental stages: vigorous growth stage (T1), topping stage (T2), and harvest stage (T3).

Results: Based on our transcriptomic and metabolomic data, 25 unigenes exhibiting stage-specific expression patterns that were strongly associated with flavonoid accumulation were identified. We found that during early developmental stages (T1-T2), upregulated expression of chalcone synthase (CHS) and chalcone isomerase (CHI) correlated with enhanced flavonoid backbone biosynthesis. In contrast, during the later stage (T3), increased expression of dihydroflavonol 4-reductase (DFR) and anthocyanidin synthase (ANS) was consistent with elevated anthocyanin accumulation.

Conclusion: This study systematically analyzed the coordinated regulatory network of flavonoid biosynthesis during leaf development in *N. tabacum*, revealing dynamic metabolic shifts across developmental stages. The findings offer novel molecular insights into the mechanisms underlying leaf quality formation and establish a theoretical framework for functional studies of candidate genes, reinforcing the utility of *N. tabacum* as a model species for secondary metabolism research and breeding innovation.

KEYWORDS

tobacco, multiomic analysis, secondary metabolites, developmental stages, transcriptomic

1 Introduction

Tobacco (*Nicotiana tabacum* L.), a significant economic crop in the Solanaceae family, originated in the Americas and is now extensively cultivated in Cuba, China, the United States, and other regions. Beyond its commercial value, tobacco serves as an important model organism for studying the regulatory mechanisms of plant secondary metabolism (Jassbi et al., 2017). The plant produces diverse secondary metabolites, including alkaloids, phenolic compounds, and terpenoids, which play pivotal roles in plant defense (Yan et al., 2017; Elser et al., 2023), stress adaptation (Hu et al., 2023; Gu et al., 2024), and environmental interactions (Wu et al., 2025). While nicotine and related alkaloids have been the primary focus of previous research, metabolites such as flavonoids, polyamines, and diterpenes also contribute significantly to the chemical diversity and biological functionality of tobacco leaves (Wu et al., 2023). These compounds not only participate in plant defense and environmental adaptation, but also play key roles in determining tobacco leaf quality and flavor (Wang et al., 2024).

The biosynthesis of secondary metabolites is regulated at multiple levels, exhibiting distinct dynamic patterns during leaf development (Qin et al., 2020; Lv et al., 2024). Using an integrated metabolomics approach, we can characterize the dynamic changes of metabolites and bioactivities during plant leaf development (Bai et al., 2024). The changes in metabolic profiles during leaf growth and development accurately reflect alterations in plant physiological functions, energy metabolism requirements, and environmental adaptation strategies (Zhang et al., 2018). For instance, young leaves are primarily characterized by active cell division and primary metabolic activities, while mature leaves tend to accumulate secondary metabolites associated with tissue differentiation and environmental defense functions (Zhou and Liu, 2022).

The application of multi-omics technologies has become a widely adopted method for exploring intricate biosynthetic mechanisms. Multi-dimensional data analysis enables in-depth exploration of the intricate regulatory networks governing secondary metabolite biosynthesis. Transcriptomics can precisely reflect gene expression patterns, comprehensively revealing the dynamics of gene expression in metabolic pathways and providing a crucial foundation for identifying key structural genes and transcription factors. For example, transcriptomic analysis of cold-stressed tobacco revealed 16,204 differentially expressed genes and highlighted the role of photosynthesis and flavonoid biosynthesis pathways (Song et al., 2024). Metabolomics technologies, particularly untargeted metabolomics, can directly capture end-product information of gene expression, comprehensively detecting metabolite species and providing evidence for further elucidating the mechanisms underlying phenotypic formation (Zhang et al., 2022). By integrating transcriptomics and metabolomics, researchers can systematically investigate the relationship between transcriptional regulatory networks and metabolite biosynthesis during plant development (Xu et al., 2020; Yang et al., 2023a). Through an integrative analysis of transcriptomic and metabolite profiles (Wang et al., 2023), demonstrated that phosphate application enhanced cold tolerance in alfalfa by modulating the biosynthesis

of key metabolites. Similarly, Moschen et al. identified transcription factors that were pre-regulated under drought conditions in sunflower, enhanced understanding of the molecular mechanisms involved in sunflower under drought conditions (Moschen et al., 2017).

In tobacco research, integrated multi-omics analyses have successfully identified stage-specific metabolic pathways and candidate regulatory genes involved in nicotine biosynthesis, flavonoid modification, and terpenoid diversification, which deepens our understanding of the complex regulatory networks underlying tobacco leaf development and secondary metabolism (Jiao et al., 2020; Yang et al., 2023b; Niu et al., 2025). These findings not only reveal the dynamic coordination between metabolic flux and gene expression but also provide potential molecular targets for variety improvement and metabolic engineering (Qi et al., 2021; Pan et al., 2025). Moreover, a deeper understanding of the developmental regulation of secondary metabolism offers valuable guidance for optimizing leaf harvesting timing, curing methods, and processing techniques. As tobacco continues to serve as a model for studying specialized metabolism, such omics-driven insights will also benefit the broader plant science community by informing research on stress adaptation, metabolite regulation, and biomass quality improvement.

Despite significant progress, the complete regulatory landscape of secondary metabolite biosynthesis during tobacco leaf development remains to be fully elucidated (Chang et al., 2020). The complexity arises from the diversity of metabolites, spatiotemporal specificity of gene expression, and the confounding effects of environmental factors (Hu et al., 2024). Therefore, systematically integrating transcriptomic and metabolomic data across critical developmental stages offers valuable insights into the regulation of secondary metabolism in model plant systems, and provides a solid foundation for reconstructing comprehensive metabolic networks.

2 Materials and methods

2.1 Plants and sample preparation

Samples of a flue-cured tobacco cultivar LY09A were collected from the experimental field of the Yunnan Academy of Tobacco Agricultural Sciences, located in Kunming City, Yunnan Province (25°03'N, 102°39'E). Leaf samples were harvested at three key developmental stages: the vigorous growth stage (T1, 60 days after transplanting), the topping stage (T2, 90 days after transplanting), and the harvest stage (T3, 150 days after transplanting).

2.2 Metabolite extraction and profiling

Metabolite extraction and profiling were performed by Wuhan Metware Biotechnology Co., Ltd. using non-targeted metabolomics approaches. Briefly, 20 mg of powdered tobacco leaf samples were homogenized with 400 µL of 70% methanol containing internal standards (2-chlorophenylalanine) and vortexed for 15 min. After

adding 200 μ L petroleum ether, the mixture was vortexed for 5 min and centrifuged at 12,000 $\times$ g for 10 min at 4°C. The lower aqueous phase was filtered through a 0.22- μ m PTFE membrane and stored at –20°C prior to LC-MS/MS analysis. Chromatographic separation was achieved on a Waters ACQUITY UPLC HSS T3 C18 column (1.8 μ m, 2.1 $\times$ 100 mm) maintained at 40°C. The mobile phases consisted of (A) ultrapure water with 0.1% formic acid and (B) acetonitrile with 0.1% formic acid, delivered at 0.4 mL/min with the following gradient: 5% B (0 min), linear increase to 90% B (11 min), held for 1 min, then returned to 5% B (12.1 min) and re-equilibrated for 1.9 min. The injection volume was 2 μ L. Mass spectrometry was performed on an Agilent 6545 QTOF/MS system equipped with an ESI source operating in both positive (ESI+) and negative (ESI–) ionization modes. Key parameters included: ion source voltage ($\pm$ 2,500 V for ESI+, $\pm$ 1,500 V for ESI–), nebulizer gas (40 psi), drying gas flow (8 L/min at 325°C), and sheath gas flow (11 L/min at 325°C). Data acquisition covered m/z 50–1,000 with MS/MS fragmentation at 135 V.

Raw data were converted to mzXML format using ProteoWizard (Holman et al., 2014), processed via XCMS for peak alignment, retention time correction, and SVR-based normalization. Metabolites were annotated by matching against in-house and public databases (KEGG), with stringent QC criteria (CV < 30% in QC samples). A total of 8,651 metabolites were detected, including 3,063 with MS/MS verification (2,525 in ESI+, 538 in ESI–).

2.3 Whole-transcriptome sequencing and transcriptomic analysis

Samples for RNA sequencing were collected in 2022. Three replicates of each plant were sent to BGI for RNA-seq on the BGISEQ-500 platform and bisulfate-seq. mRNA was purified using poly-T oligo-attached magnetic beads, fragmented at high temperature with divalent cations, and converted to first-strand cDNA with reverse transcriptase and random primers. Second-strand cDNA was synthesized using DNA Polymerase I and RNase H, followed by adapter ligation after adding an ‘A’ base. Products were purified and enriched via PCR amplification. PCR yield was quantified by Qubit, and samples were pooled to form single-stranded DNA circles (ssDNA circles) for the final library. ssDNA circles underwent rolling circle replication (RCR) to generate DNA nanoballs (DNBs), which enhanced fluorescent signals during sequencing (Zhai et al., 2020). DNBs were loaded onto patterned nanoarrays, and 100 bp (or 150 bp) paired-end reads were performed on the BGISEQ-500 platform (Fehlmann et al., 2016) using the Combinational Probe-Anchor Synthesis Sequencing Method for subsequent data analysis.

2.4 Statistical analysis

Raw RNA-seq reads were initially quality-filtered and trimmed using fastp (version 0.21.0) (Chen et al., 2018) with default parameters. Clean reads longer than 60 bp and without ambiguous bases (N) were

retained for downstream analysis. The filtered reads were then aligned to the *Nicotiana tabacum* version ZY300 reference genome (<https://www.ebi.ac.uk/ena/browser/view/PRJEB85578>) using HISAT2 (Kim et al., 2015). Gene expression levels were quantified using StringTie (version 2.1.4) (Pertea et al., 2015) and normalized using two metrics: transcripts per million (TPM) and fragments per kilobase of exon per million mapped fragments (FPKM). Differential expression analysis between different developmental stages was conducted using the DESeq2 (Love et al., 2014), and genes with $|\log_2(\text{fold change})| \geq 1$ and adjusted p-value < 0.05 were considered significantly differentially expressed. Functional annotation of differentially expressed and novel genes was performed using BLAST searches against public databases, including Gene Ontology (GO) (The Gene Ontology Consortium 2021) and the Kyoto Encyclopedia of Genes and Genomes (KEGG), and enrichment analyses were used to identify significantly associated biological processes and metabolic pathways.

3 Results

3.1 Transcriptomic and metabolomic profiling in tobacco leaves

After rigorous data processing and quality control, we obtained high-quality transcriptomic datasets (Supplementary Table S1). RNA sequencing (RNA-seq) and untargeted metabolomic profiling were employed to systematically investigate transcriptomic and metabolic changes in tobacco leaves across three developmental stages: T1, T2, and T3. Principal component analysis (PCA) revealed distinct stage-specific clustering at both the transcriptomic and metabolomic levels. For the RNA-seq data, principal components 1 and 2 (PC1 and PC2) accounted for 23.15% and 16.75% of the total variance, respectively (Figure 1A). In the metabolomic data, PC1 and PC2 explained 40.51% and 16.50% of the variance (Figure 1D), highlighting substantial transcriptional and metabolic reprogramming during leaf development. Pearson correlation analysis further confirmed the reliability and consistency of both datasets. Biological replicates from the same developmental stage exhibited strong positive correlations ($r > 0.9$), while correlations between different stages—particularly between T1 and T3—were significantly lower (Figures 1B, E). Violin plots of log-transformed gene expression values showed consistent distributions across all RNA-seq samples (Figure 1C), indicating uniform sequencing depth and high data quality. To enhance the resolution of metabolomic differences, orthogonal partial least squares discriminant analysis (OPLS-DA) was conducted. This analysis clearly separated the metabolic profiles of the T1, T2, and T3 stages (Figure 1F), further supporting distinct metabolic reprogramming during tobacco leaf development.

Meanwhile, we annotated and classified the differentially accumulated metabolites based on their chemical structures and biological functions detected by non-targeted metabolomics technology. In total, over 2,300 high-confidence metabolites were identified, with heterocyclic compounds (373), organic acid derivatives (298), and aldehyde/ketone derivatives (284)

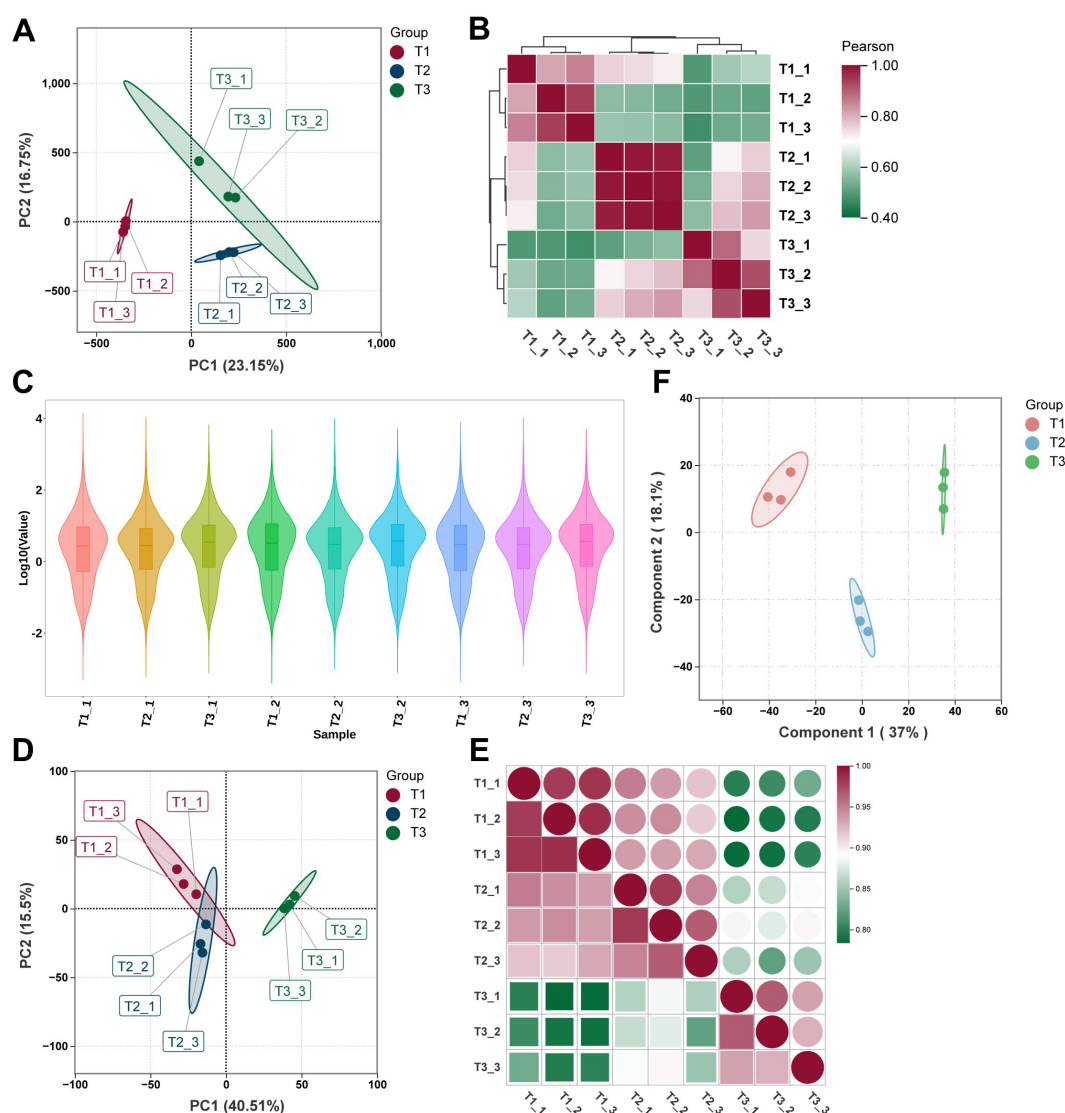


FIGURE 1

Overview of transcriptomic and metabolomic data of tobacco leaves at different developmental stages. (A) Principal component analysis (PCA) of transcriptomic data from T1, T2, and T3 stages. (B) Pearson correlation heatmap of transcriptomic data across all biological replicates. (C) Violin plots showing the distribution of gene expression values across all transcriptomic samples. (D) PCA of untargeted metabolomic data from T1, T2, and T3 stages. (E) Pearson correlation heatmap of metabolomic data across all biological replicates. (F) Orthogonal partial least squares discriminant analysis (OPLS-DA) score plot showing clear separation of metabolic profiles among the three developmental stages.

representing the three major categories (Figure 2A). In addition to primary metabolites, several key secondary metabolites associated with plant growth and defense were also detected, including flavonoids, alkaloids, lignans/coumarins, and terpenoids. Further analysis of their relative abundance revealed distinct accumulation patterns of secondary metabolites across developmental stages: alkaloids increased significantly at the T3 stage, whereas flavonoids and lignans accumulated progressively during leaf maturation (Figure 2B). This phenomenon may be attributed to the substantial allocation of nitrogen resources to nicotine synthesis during tobacco leaf maturation, leading to alkaloid accumulation. In contrast, the rise in flavonoid and lignan levels might be linked to enhanced antioxidant capacity and adaptive responses to environmental stress during leaf development (Supplementary

Table S2) (Figure 2C). Collectively, these results demonstrate the rich metabolic diversity of tobacco leaves during development and highlight the dynamic changes in secondary metabolites, particularly flavonoids.

3.2 Stage-specific functional pathway shifts during tobacco leaf development

To reveal the stage-specific transcriptomic changes and their causes during tobacco leaf development, differential gene expression (DEG) analysis and KEGG pathway enrichment were performed across three developmental stages (T1, T2, and T3). The results showed a large number of DEGs in each pairwise comparison, with

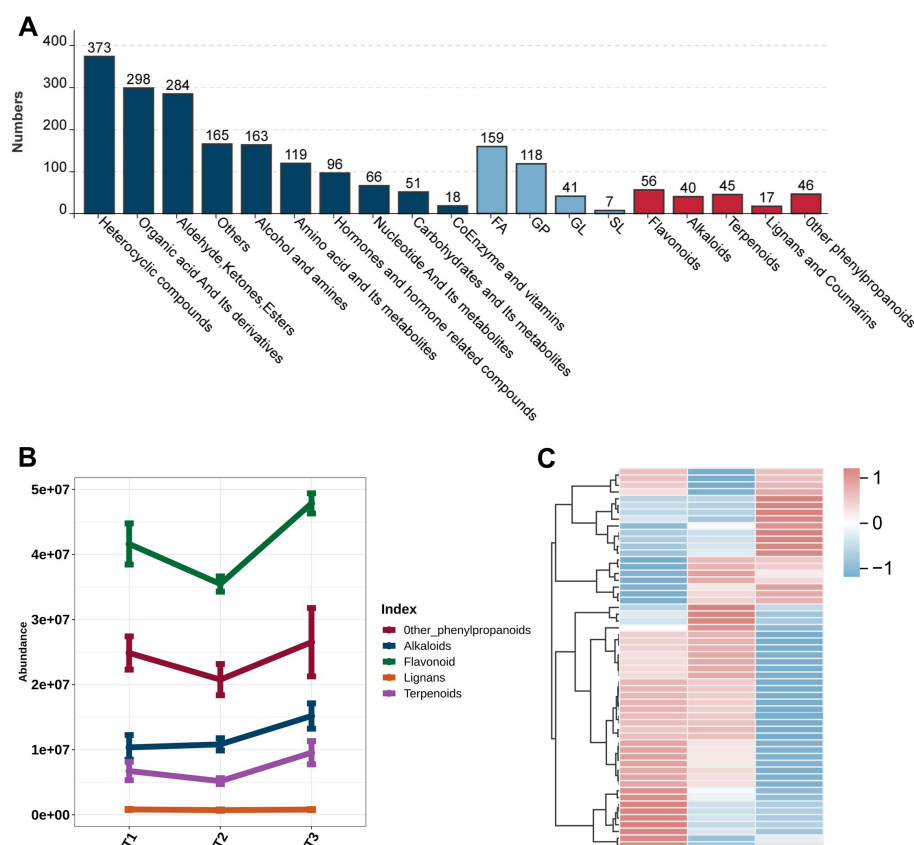


FIGURE 2

Classification of metabolites identified in tobacco leaves and dynamic changes in secondary metabolite contents. (A) Column chart of the classification of metabolites, with the height of different columns representing the number of different compound classes. (B) Line chart of the trend of different compound species over time. (C) Heatmap of flavonoid metabolite accumulation patterns at different developmental stages.

9,962 genes between T1 and T2, 9,863 genes between T1 and T3, and 4,626 genes between T2 and T3. Notably, more drastic transcriptomic changes were observed between T1 and the later stages (T2 and T3) (Supplementary Tables S4–S6) (Supplementary Figures S1A–C). Further enrichment analysis of the KEGG pathway on DEGs revealed that the biological processes involved in different developmental stages showed significant differences. In the T1 vs. T2 comparison, DEGs were significantly enriched in pathways related to energy metabolism and pigment biosynthesis, such as photosynthesis, carbon fixation, flavonoid biosynthesis, and terpenoid backbone biosynthesis (Supplementary Figure S1D). In the T1 vs. T3 comparison, in addition to these pathways, DEGs were also enriched in plant hormone signal transduction, Th17 cell differentiation, and Toll-like receptor signaling, indicating activation of complex hormonal and immune regulatory networks during leaf maturation (Supplementary Figure S1E). The T2 vs. T3 comparison revealed enrichment in detoxification and stress response pathways, including cytochrome P450-mediated metabolism, glutathione metabolism, and xenobiotic metabolism. Lipid-related pathways such as linoleic acid metabolism, steroid hormone biosynthesis, and PPAR signaling were also significantly enriched (Supplementary Figure S1F). These findings suggest a clear developmental trajectory in tobacco leaves, transitioning from

early-stage energy assimilation and pigment production to late-stage stress adaptation and metabolic specialization.

Venn diagram analysis identified 710 differentially expressed genes (DEGs) that were shared across all three pairwise comparisons (Figure 3A), representing a core set of genes dynamically regulated throughout the developmental and maturation stages of tobacco leaves (Supplementary Table S7). Gene Ontology (GO) enrichment analysis of these overlapping DEGs revealed significant enrichment in biological processes related to metabolism and response to stimuli, as well as cellular components such as membranes and organelles (Figure 3B). These results suggest that common molecular pathways are engaged during various developmental transitions, potentially playing critical roles in determining leaf phenotype and function.

3.3 Metabolomic studies revealed differentially accumulated metabolites in tobacco leaves at different developmental stages

In order to study the dynamic changes of metabolites during tobacco leaves development, we compared the metabolomic data of

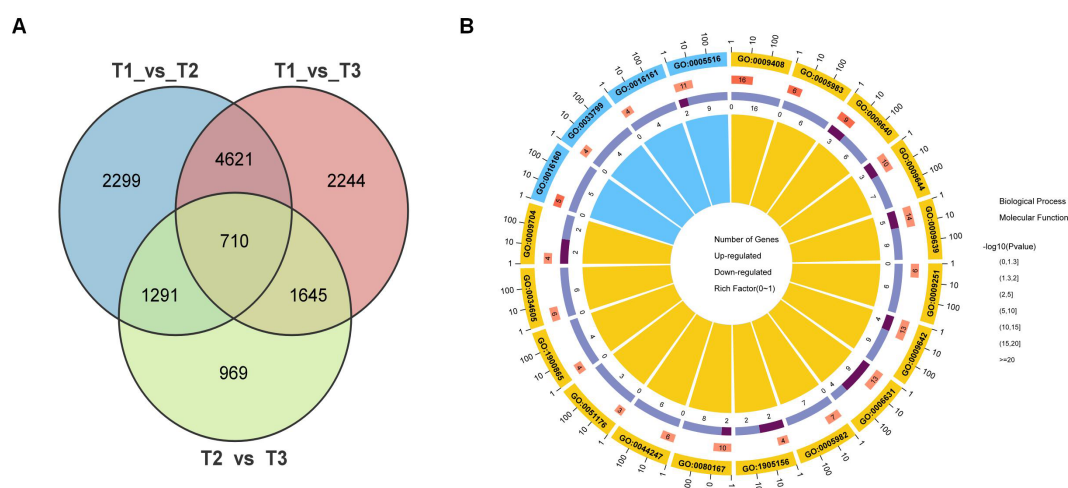


FIGURE 3

Overlap and functional classification of differentially expressed genes (DEGs) during tobacco leaf development. (A) Venn diagram showing the number of DEGs identified in three pairwise comparisons: T1 vs T2, T1 vs T3, and T2 vs T3. A total of 710 DEGs were shared among all three comparisons. (B) GO enrichment analysis of the overlapping DEGs.

T1, T2, and T3 with each other. A total of 28 differentially accumulated metabolites (DAMs) were identified between T1 and T2, and most showed significant up-accumulation ($\log_2\text{FC} > 1$, $p < 0.05$) (Figure 4A). In the comparison between T1 and T3, more DAMs were detected, indicating that the metabolites changed more significantly with leaves development (Figure 4B). It is worth noting that several metabolites showed continuous accumulation from T1 to T3, possibly participating in the metabolic process of leaves maturation. In the comparison of T2 and T3, there were only fewer DAMs, which means that the most significant metabolic changes occurred during the early transition from T1 to T2 (Figure 4C). The Venn diagram illustrates the overlaps of DAMs between pairwise comparisons of developmental stages (T1 vs. T2, T1 vs. T3, and T2 vs. T3) (Figure 4D). Notably, 14 DAMs were shared between T1 vs. T2 and T1 vs. T3, primarily including amino acids and their derivatives (e.g., O-acetyl-L-serine, L-phenylalanine), benzene derivatives (e.g., oxadixyl, benzaldehyde), nucleotides, and various aldehydes and ketones, indicating their consistent involvement in early to late leaf development transitions. Additionally, 9 DAMs were common between T1 vs. T3 and T2 vs. T3, comprising carbohydrates, organic acids, heterocyclic compounds, steroids, and other metabolites, reflecting metabolic shifts associated with the maturation and harvest stages. The absence of shared DAMs between T1 vs. T2 and T2 vs. T3 comparisons, as well as no metabolites common across all three stages, underscores the stage-specific metabolic remodeling in tobacco leaves, supporting the notion of a tightly regulated biosynthetic network modulating secondary metabolism throughout development. These DAMs are mainly enriched in pathways related to secondary metabolism, including alkaloid biosynthesis, phenylpropane metabolism, and flavonoid biosynthesis, and may play potential roles in regulating tobacco leaves quality and adapting to environmental stresses

during development. A detailed list of DAMs and their statistical parameters can be found in [Supplementary Table S8](#).

3.4 The transcription levels of genes related to the flavonoid biosynthesis pathway at different developmental stages

Based on KEGG pathway enrichment and GO functional annotation, a total of 25 unigenes encoding key enzymes involved in the flavonoid biosynthesis pathway were identified in this study. These included five unigenes encoding chalcone synthases (CHS), two for chalcone isomerases (CHI), two for flavanone 3-hydroxylases (F3H), three for dihydroflavonol 4-reductases (DFR), two for flavonol synthases (FLS), and two for anthocyanidin synthases (ANS), among others — all of which play vital roles in catalyzing distinct steps of the flavonoid biosynthetic pathway (Figure 5A). Heatmap analysis revealed distinct expression patterns of these genes across the three developmental stages (T1, T2, and T3) (Figure 5B). For example, CHS genes (e.g., chr7.3897 and chr7.1255) and CHI genes showed significantly higher expression levels during the early stage (T1), suggesting their critical roles in the initial synthesis of flavonoid backbones. In contrast, downstream genes such as DFR (chr4.7209, chr7.6759) and ANS (chr6.2245, chr7.4145), which are involved in anthocyanin biosynthesis, were markedly upregulated at the late developmental stage (T3), indicating a potential accumulation of anthocyanin compounds during leaf maturation (Supplementary Table S9). Overall, these results demonstrate that the transcriptional regulation of flavonoid biosynthesis-related genes is highly stage-specific, with different enzymatic steps activated at specific growth stages to coordinate the metabolic flow of flavonoids. These findings

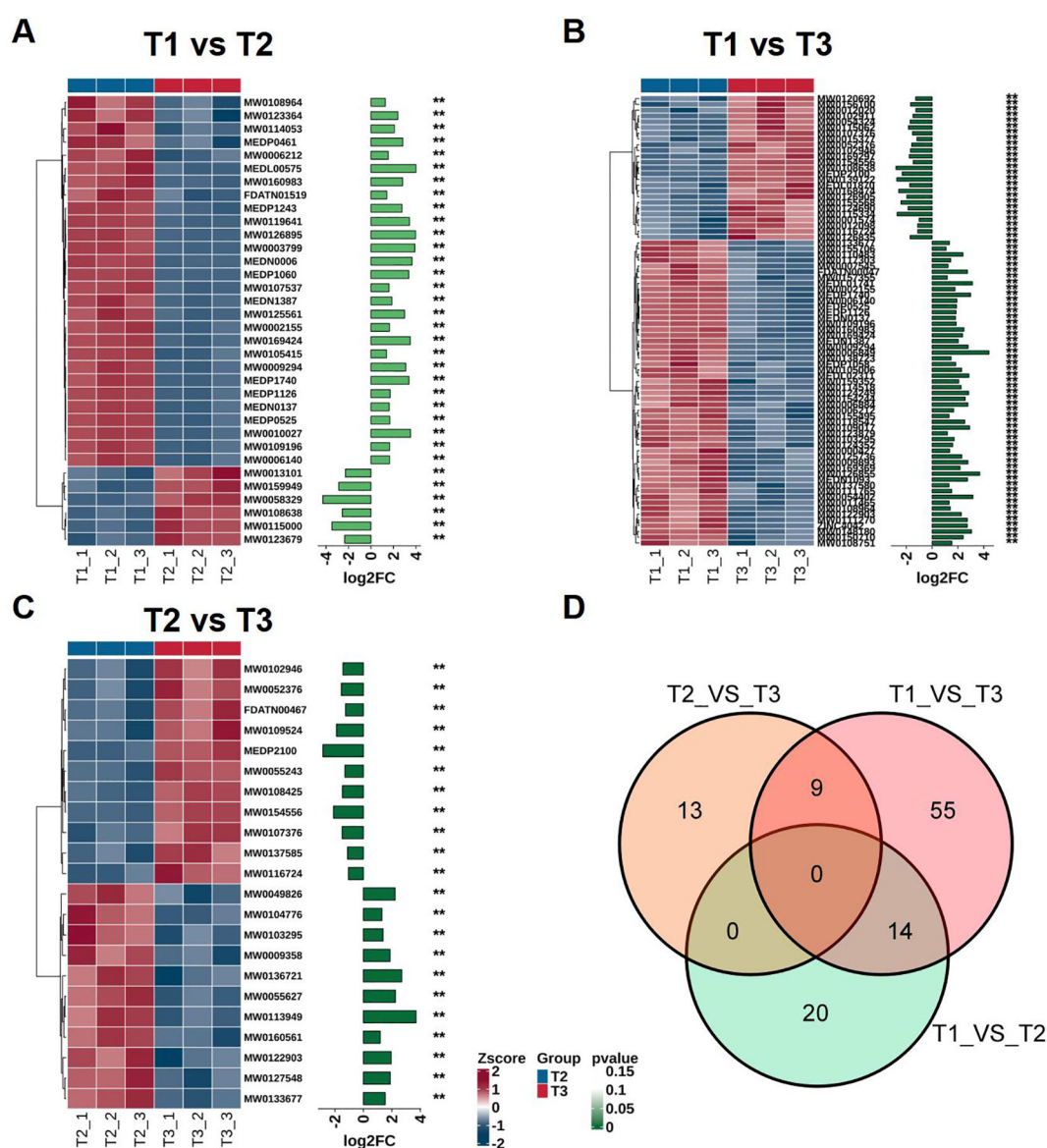


FIGURE 4

Specific changes at different stages of differentially accumulated metabolites (DAMs) in tobacco leaves. (A) T1 and T2 comparison, (B) T1 and T3 comparison, (C) T2 and T3 comparison, (D) Venn diagram illustrating the overlap of DAMs among the three pairwise comparisons. "***" indicates a statistically significant difference at $p < 0.01$ based on Student's *t*-test.

provide important insights into the regulatory mechanisms of flavonoid biosynthesis and the identification of key genes for future research.

4 Discussion

In this study, we identified 25 key candidate genes potentially involved in flavonoid metabolism regulation through transcriptomics and non-targeted metabolomics, laying a crucial theoretical foundation for deciphering the flavonoid metabolic network in tobacco leaves. Notably, although this research primarily focused on flavonoids as important secondary metabolites, the significant

dynamic changes and biosynthetic mechanisms of other metabolic pathways—such as terpenoids, alkaloids, and phenolic acids—during tobacco growth and development remain worthy of investigation. This is particularly relevant given existing studies demonstrating pronounced changes in alkaloids during tobacco root development. These secondary metabolites may exhibit spatiotemporal coordination with flavonoid biosynthesis, providing valuable insights for further research into their potential roles in regulating tobacco quality, stress responses, and developmental processes. Future studies should expand the analytical scope to elucidate the interactions among different metabolic pathways, thereby achieving a comprehensive, systems-level understanding of metabolic regulation during tobacco development.

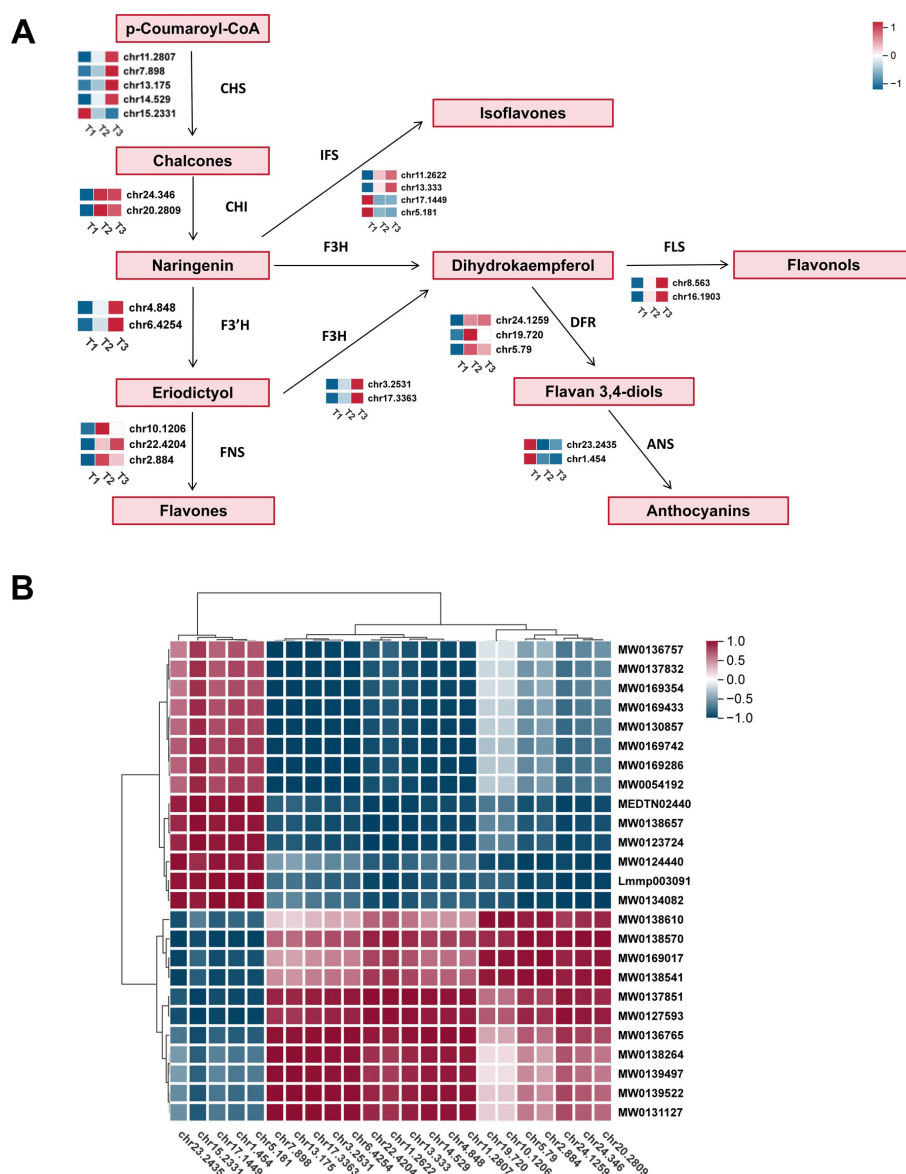


FIGURE 5

The transcriptional map of flavonoid biosynthesis pathway genes in tobacco across different developmental stages. (A) A total of 25 unigenes encoding key enzymes involved in the flavonoid biosynthesis pathway were identified, including CHS, CHI, F3H, F3'H, FNS, IFS, FLS, DFR, and ANS. The heatmaps next to each enzyme represent the relative expression levels of the corresponding unigenes at three developmental stages. (B) Hierarchical clustering heatmap of 27 flavonoid biosynthesis-related genes at three developmental stages.

Moreover, although the integration of metabolomic and transcriptomic data has enabled preliminary construction of a regulatory network for flavonoid biosynthesis, the universality and robustness of these regulatory mechanisms still require further validation. Transcript-level expression changes alone are insufficient to comprehensively and accurately determine the biological functions of the 25 candidate genes. Therefore, subsequent studies should employ experimental validation approaches, such as gene overexpression and CRISPR/Cas9 genome editing, to elucidate the specific roles of these genes in flavonoid synthesis while further characterizing their core expression patterns.

With the continuous advancement of sequencing technologies, it is imperative to conduct population-scale investigations across

diverse tobacco germplasm resources. Strategies such as metabolome-wide association studies (mGWAS), expression quantitative trait loci (eQTL) mapping, and genetic variation analysis can reveal conserved regulatory patterns and core regulatory elements within natural populations. These findings will not only provide theoretical guidance for breeding high-quality tobacco varieties but also yield valuable molecular markers for practical applications.

In summary, this study systematically elucidates the metabolic regulatory mechanisms of flavonoid biosynthesis during tobacco leaf development while highlighting the need for functional validation and population-level studies to fully uncover its biological significance and breeding potential.

In addition to the current findings, there are several valuable directions for future research. For example, the identification of key candidate genes involved in flavonoid biosynthesis provides a valuable starting point for future functional studies and it will be better to perform in-depth functional analyses, including gene editing and overexpression experiments, to uncover the specific regulatory roles of these genes. Furthermore, recognizing the diversity among tobacco varieties—such as flue-cured, cigar, and other types, it is important to explore their metabolomic and transcriptomic differences during similar developmental stages. The integrated their datasets and comparative analyses will help elucidate the molecular basis underlying varietal differences in leaf characteristics and contribute to a broader understanding of metabolic regulation across tobacco types.

5 Conclusion

This study employed an integrated transcriptomic and metabolomic approach to investigate the molecular characteristics of flavonoid biosynthesis and associated differential metabolites across different developmental stages of tobacco leaves. A total of 25 unigenes involved in flavonoid biosynthesis were identified, whose expression patterns showed strong correlation with the accumulation profiles of flavonoid metabolites, including flavones, flavonols, and anthocyanins. The early developmental stages (T1-T2) were characterized by upregulated expression of CHS, CHI, and F3H genes, promoting flavonoid skeleton formation. In contrast, the later stage (T3) exhibited a metabolic shift toward anthocyanin biosynthesis, marked by increased expression of DFR and ANS genes. These findings elucidate the coordinated regulatory network between gene expression and metabolite accumulation in secondary metabolism, providing a theoretical foundation for improving tobacco leaf quality through metabolic engineering and precision breeding strategies.

Data availability statement

The transcriptome data presented in this study have been deposited in the National Genomics Data Center (NGDC) under BioProject accession number PRJCA039108, and the raw RNA-seq data are available in the Genome Sequence Archive (GSA) under accession number CRA024963.

Author contributions

ZT: Resources, Investigation, Writing – original draft, Data curation. KY: Data curation, Methodology, Visualization, Writing –

original draft. XC: Data curation, Writing – review & editing, Investigation. XS: Data curation, Writing – review & editing, Investigation. YH: Writing – original draft, Resources, Validation. FX: Resources, Writing – original draft, Validation. SZ: Resources, Writing – original draft, Validation. ES: Validation, Writing – original draft, Resources. LF: Writing – review & editing, Supervision, Conceptualization. BX: Funding acquisition, Resources, Project administration, Writing – review & editing. SW: Resources, Validation, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was funded by the Yunnan Academy of Tobacco Agricultural Sciences under the following project numbers: 2021530000241013, 2022530000241003, and 2024530000241001.

Conflict of interest

The 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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

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.2025.1615756/full#supplementary-material>

References

- Bai, Z., Tang, J., Li, Y., Li, Z., Gu, S., Deng, L., et al. (2024). Integrated Metabolomics Approach Reveals the Dynamic Variations of Metabolites and Bioactivities in *Paeonia ostii* 'Feng Dan' Leaves during Development. *Int. J. Mol. Sci.* 25, 1059. doi: 10.3390/ijms25021059
- Chang, W., Zhao, H., Yu, S., Yu, J., Cai, K., Sun, W., et al. (2020). Comparative transcriptome and metabolomic profiling reveal the complex mechanisms underlying the developmental dynamics of tobacco leaves. *Genomics* 112, 4009–4022. doi: 10.1016/j.ygeno.2020.07.005
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, 1884–1890. doi: 10.1093/bioinformatics/bty560
- Elser, D., Plieger, D., Villette, C., Moegle, B., Miesch, L., and Gaquerel, E. (2023). Evolutionary metabolomics of specialized metabolism diversification in the genus *Nicotiana* highlights *N*- acylornicotine innovations. *Sci. Adv.* 9, eade8984. doi: 10.1126/sciadv.ade8984
- Fehlmann, T., Reinheimer, S., Geng, C., Su, X., Drmanac, S., Alexeev, A., et al. (2016). cPAS-based sequencing on the BGISEQ-500 to explore small non-coding RNAs. *Clin. Epigenet* 8, 123. doi: 10.1186/s13148-016-0287-1
- Gu, K., Li, X., Su, J., Chen, Y., Yang, C., Li, J., et al. (2024). Physiological and ecological responses of flue-cured tobacco to field chilling stress: insights from metabolomics and proteomics. *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1490633
- Holman, J. D., Tabb, D. L., and Mallick, P. (2014). ProteoWizard: Open-source software for rapid proteomics tools development. *Bioinformatics* 30 (18), 2532–2534. doi: 10.1093/bioinformatics/btu394
- Hu, Z., He, Z., Li, Y., Wang, Q., Yi, P., Yang, J., et al. (2023). Transcriptomic and metabolic regulatory network characterization of drought responses in tobacco. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1067076
- Hu, Y., Ruan, Y., Zhao, X.-L., Jiang, F., Liu, D., Zhu, Q., et al. (2024). PCMD: A multilevel comparison database of intra- and cross-species metabolic profiling in 530 plant species. *Plant Commun.* 5, 101038. doi: 10.1016/j.xplc.2024.101038
- Jassbi, A. R., Zare, S., Asadollahi, M., and Schuman, M. C. (2017). Ecological roles and biological activities of specialized metabolites from the genus *nicotiana*. *Chem. Rev.* 117, 12227–12280. doi: 10.1021/acs.chemrev.7b00001
- Jiao, F., Zhao, L., Wu, X., Song, Z., and Li, Y. (2020). Metabolome and transcriptome analyses of the molecular mechanisms of flower color mutation in tobacco. *BMC Genomics* 21, 611. doi: 10.1186/s12864-020-07028-5
- Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi: 10.1186/s13059-014-0550-8
- Lv, J., Yang, S., Zhou, W., Liu, Z., Tan, J., and Wei, M. (2024). Microbial regulation of plant secondary metabolites: Impact, mechanisms and prospects. *Microbiological Res.* 283, 127688. doi: 10.1016/j.micres.2024.127688
- Moschen, S., Di Rienzo, J. A., Higgins, J., Tohge, T., Watanabe, M., González, S., et al. (2017). Integration of transcriptomic and metabolic data reveals hub transcription factors involved in drought stress response in sunflower (*Helianthus annuus* L.). *Plant Mol. Biol.* 94, 549–564. doi: 10.1007/s11103-017-0625-5
- Niu, L., Hu, W., Wang, F., Shaker, M., and Yang, X. (2025). Integrated transcriptomic and metabolomic analyses elucidate the mechanism by which grafting impacts potassium utilization efficiency in tobacco. *BMC Plant Biol.* 25, 94. doi: 10.1186/s12870-025-06123-7
- Pan, Y., Xiong, G., Peng, Q., Zhang, J., Zhou, F., and Xia, T. (2025). Changes in the metabolome of flue-cured tobacco by dioctyl phthalate and erucamide: mechanism of continuous monocropping. *J. Soil Sci. Plant Nutr.* doi: 10.1007/s42729-025-02342-x
- Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt.3122
- Qi, Y., Li, C., Duan, C., Gu, C., and Zhang, Q. (2021). Integrated metabolomic and transcriptomic analysis reveals the flavonoid regulatory network by eutrema esMYB90. *IJMS* 22, 8751. doi: 10.3390/ijms22168751
- Qin, D., Wang, Q., Li, H., Jiang, X., Fang, K., Wang, Q., et al. (2020). Identification of key metabolites based on non-targeted metabolomics and chemometrics analyses provides insights into bitterness in Kucha [*Camellia kucha* (Chang et Wang) Chang. *Food Res. Int.* 138, 109789. doi: 10.1016/j.foodres.2020.109789
- Song, X., Wang, H., Wang, Y., Zeng, Q., and Zheng, X. (2024). Metabolomics combined with physiology and transcriptomics reveal how *Nicotiana tabacum* leaves respond to cold stress. *Plant Physiol. Biochem.* 208, 108464. doi: 10.1016/j.plaphy.2024.108464
- The Gene Ontology Consortium (2021). The Gene Ontology resource: enriching a Gold mine. *Nucleic Acids Res.* 49 (D1), D325–D334. doi: 10.1093/nar/gkaa1113
- Wang, Z., Luo, Z., Li, Z., Liu, P., He, S., Yu, S., et al. (2024). NtMYB27 acts downstream of NtBES1 to modulate flavonoids accumulation in response to UV-B radiation in tobacco. *Plant J.* 119, 2867–2884. doi: 10.1111/tpj.16958
- Wang, Y., Sun, Z., Wang, Q., Xie, J., and Yu, L. (2023). Transcriptomics and metabolomics revealed that phosphate improves the cold tolerance of alfalfa. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1100601
- Wu, J., Lv, S., Zhao, L., Gao, T., Yu, C., Hu, J., et al. (2023). Advances in the study of the function and mechanism of the action of flavonoids in plants under environmental stresses. *Planta* 257, 108. doi: 10.1007/s00425-023-04136-w
- Wu, S., Tian, L., Guo, S., Lei, H., Zhao, X., Hao, X., et al. (2025). OsLCI, a transaldolase, regulates cell patterning and leaf morphology through modulation of secondary metabolism. *Plant Biotechnol. J.* 70004. doi: 10.1111/pbi.70004
- Xu, J., Chen, Z., Wang, F., Jia, W., and Xu, Z. (2020). Combined transcriptomic and metabolomic analyses uncover rearranged gene expression and metabolite metabolism in tobacco during cold acclimation. *Sci. Rep.* 10, 5242. doi: 10.1038/s41598-020-62111-x
- Yan, N., Liu, Y., Zhang, H., Du, Y., Liu, X., and Zhang, Z. (2017). Solanesol biosynthesis in plants. *Molecules* 22, 510. doi: 10.3390/molecules22040510
- Yang, G., Qin, Y., Jia, Y., Xie, X., Li, D., Jiang, B., et al. (2023a). Transcriptomic and metabolomic data reveal key genes that are involved in the phenylpropanoid pathway and regulate the floral fragrance of *Rhododendron fortunei*. *BMC Plant Biol.* 23, 8. doi: 10.1186/s12870-022-04016-7
- Yang, Q., Zhang, Y., Qu, X., Wu, F., Li, X., Ren, M., et al. (2023b). Genome-wide analysis of UDP-glycosyltransferases family and identification of UGT genes involved in abiotic stress and flavonol biosynthesis in *Nicotiana tabacum*. *BMC Plant Biol.* 23, 204. doi: 10.1186/s12870-023-04208-9
- Zhai, F.-H., Zhan, Q.-Q., Wang, K., Chen, S., and He, R.-H. (2020). A homogeneous DNA nanosphere for fluorescence detection of microRNAs with high-ordered aggregation enhanced emission and enzyme-free cascade amplification. *Sensors Actuators B: Chem.* 320, 128394. doi: 10.1016/j.snb.2020.128394
- Zhang, F., Ma, Z., Qiao, Y., Wang, Z., Chen, W., Zheng, S., et al. (2022). Transcriptome sequencing and metabolomics analyses provide insights into the flavonoid biosynthesis in *Torreya grandis* kernels. *Food Chem.* 374, 131558. doi: 10.1016/j.foodchem.2021.131558
- Zhang, L., Zhang, X., Ji, H., Wang, W., Liu, J., Wang, F., et al. (2018). Metabolic profiling of tobacco leaves at different growth stages or different stalk positions by gas chromatography–mass spectrometry. *Ind. Crops Products* 116, 46–55. doi: 10.1016/j.indcrop.2018.02.041
- Zhou, X., and Liu, Z. (2022). Unlocking plant metabolic diversity: A (pan)-genomic view. *Plant Commun.* 3, 100300. doi: 10.1016/j.xplc.2022.100300