

Developmental multi-omics reveals provenance-dependent regulatory and metabolic trajectories underlying geo-authenticity in *Dendrobium officinale*

Chong Feng

Sun Yat-sen University

Xinlian Chen

Guangzhou University of Chinese Medicine

Jun Liu

Sun Yat-sen University

Yu'e Chen

Guangzhou Caizhilin Pharmaceutical Co., Ltd

Enyao Ma

Guangzhou Caizhilin Pharmaceutical Co., Ltd

Depo Yang

Sun Yat-sen University

Zhimin Zhao

zhaozhm2@mail.sysu.edu.cn

Sun Yat-sen University

Research Article

Keywords: *Dendrobium officinale*, geo-authenticity, multi-omics, metabolomics, transcriptomics, WGCNA

Posted Date: August 11th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10370524/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Dendrobium officinale, a medicinal and ornamental species of the genus *Dendrobium*, exhibits marked regional variation in growth characteristics, secondary metabolite accumulation, and overall quality across its natural distribution range in China, yet the developmental and regulatory basis of this variation remains unresolved.

Results

In this study, four provenance-defined accessions from Guangdong (GD), Zhejiang (ZJ), Yunnan (YN), and Guizhou (GZ) were characterized through morphological, anatomical, physiological, metabolomic, and transcriptomic analyses to explore the spatiotemporal basis of geo-authenticity. Metabolomic profiling revealed accession-dependent trajectories of quality-related metabolites, which were summarized using a harvest suitability score (HSS) integrating the relative balance between EI- and SI-associated chemical dimensions across the sampled developmental window. Transcriptomic analysis revealed that inter-accession divergence was established through phase-dependent regulatory reprogramming, initiated at early developmental transitions and progressively structured around signalling, defence, and phenylpropanoid metabolism at later stages. Multi-omics integration identified co-expression modules associated with metabolite-derived quality indices, stem traits, and phytohormones. Hub gene analysis, TF–candidate gene relationships, and promoter cis-element profiling further suggested that quality-related modules are embedded in hormone-responsive and defense-associated regulatory networks.

Conclusions

Together, these findings suggest that geo-authenticity in *D. officinale* is realised through coordinated changes in developmental timing, transcriptional regulation, and metabolite accumulation, and provide an association-based framework for provenance evaluation, developmental monitoring of medicinal quality, and quality-oriented utilization of *D. officinale*.

Background

Dendrobium Sw. (Orchidaceae) is one of the most species-rich genera and is well known for its ornamental and economic importance. The medicinal use history of *Dendrobium* can be traced to the Qin and Han dynasties in China ^[1]. *Dendrobium officinale*, *D. nobile*, *D. huoshanense*, *D. chrysotoxum*, and *D. fimbriatum* are recorded in the current edition of the Chinese Pharmacopoeia ^[2]. In China, *D. officinale* is available in various forms, including fresh stems (used traditionally) and Fengdou (a kind of traditionally processed dried product), as well as modern preparations such as capsules and freeze-dried powders. To date, more than 180 compounds have been isolated from *D. officinale*, including polysaccharides,

bibenzyls, flavonoids, and polyphenolic compounds, many of which have been reported to possess immunomodulatory, hypoglycemic, antioxidant, anti-aging, and gastrointestinal and cardiovascular protective activities ^[3].

Geo-authenticity is closely related to the concept of daodi property in traditional Chinese medicine, which refers to the great quality of Chinese medicinal materials produced in a specific region ^[4, 5]. Investigating daodi property of medicinal plants is important not only for the quality assurance but also for providing a basis for their conservation and sustainable utilization. The concept can be traced back to *Shennong's Classic of Materia Medica*, which states “The land yields, revealing the authenticity and freshness” — implying that the land of origin, authenticity, and freshness of medicinal materials are fundamental to their quality. Environmental factors such as light, temperature, water and soil conditions play important roles in plant growth and in the biosynthesis of secondary metabolites ^[6, 7]. For example, it has been reported that moderate drought and relatively low soil nutrient availability favor the accumulation of secondary metabolites in *Ziziphi Spinosae Semen* ^[8]. Meanwhile, genetic background is widely regarded as an important contributor to daodi property, as the great quality of daodi medicinal materials arises from genotype-dependent metabolite biosynthesis shaped by long-term adaptation ^[9]. The distinctive characteristics of *Salvia miltiorrhiza* from Zhongjiang, Sichuan, have been attributed to heritable genomic variation—specifically, the formation and expansion of phenolic acid biosynthetic gene clusters and the optimization of the regulatory network of the key transcription factor SmWRKY40, which together contribute to its unique quality ^[10]. Revealing differences in gene expression among medicinal plants from different production regions is an important aspect of functional gene research ^[11]. However, the biological basis of daodi property has not yet been fully resolved at the molecular level.

In China, *D. officinale* is distributed across Zhejiang, Guangdong, Yunnan, Guizhou and several other regions. Ancient Chinese materia medica text document multiple production areas for this species and note differences in quality evaluation among materials from different production regions ^[12]. Increasing evidence indicated that *D. officinale* from different production regions differs significantly in agronomic traits, metabolite composition, and stress resistance. For example, significant variation in plant height, stem length, and stem diameter has been reported among regional populations, and samples from different origins can be discriminated based on chemical fingerprints and flavonoid composition ^[13, 14, 15]. Differences in cold tolerance among regional populations have also been observed under low-temperature stress ^[16]. Although previous studies have demonstrated substantial provenance-dependent variation in *D. officinale*, the developmental and regulatory basis underlying such divergence remain poorly understood, particularly under a common cultivation framework.

In this study, four provenance-defined accessions of *D. officinale* from Guangdong (GD), Zhejiang (ZJ), Yunnan (YN), and Guizhou (GZ) were investigated under a common-garden framework. Integrated morphological, anatomical, physiological, metabolomic, and transcriptomic analysis were performed to characterize provenance-dependent spatiotemporal developmental trajectories and to elucidate the molecular basis of geo-authenticity. Here, geo-authenticity is operationalized as reproducible provenance-associated differences in quality-related metabolite profiles and derived quality indices under common

cultivation conditions. Our findings provide a candidate model for understanding provenance-dependent quality formation and offer a basis for quality evaluation and targeted utilization of *D. officinale*.

Results

Spatiotemporal divergence in morphophysiological and phytohormone profiles across provenance-defined accessions

The anatomical, physiological, metabolomic, and transcriptomic analyses were conducted on four provenance-defined accessions over a 240-day developmental time course to characterize developmental variation and provenance-dependent divergence in *D. officinale* (Fig. 1A). Representative morphological profiles at 150 DAT are shown in Fig. 1B; full-stage profiles of stems across all time points are provided in Figure S1. Stem length increased progressively with DAT in all accessions, whereas stem diameter followed accession-specific trajectories (Fig. 1C). In GD, both stem length and diameter increased during early stage, showed little net change after 60 DAT. In ZJ, rapid early stem elongation was accompanied by an initial decrease in stem diameter from 0 to 30 DAT, after which diameter remained relatively stable. GD maintained a greater stem diameter than ZJ from 30 DAT, yet both accessions reached comparable mean stem lengths by 240 DAT (GD: 97.6 mm; ZJ: 108.3 mm). YN exhibited the most modest elongation and was the shortest accession at 240 DAT (75.8 mm), while its stem diameter changed little but remained the largest among the four accessions at most stage. In GZ, stem elongation accelerated during later stages, stem diameter increased during early development before declining as elongation became more pronounced. These accession-specific differences in the balance between axial elongation and radial expansion reflect distinct developmental strategies among the four accessions. All four phytohormones showed significant accession-dependent temporal dynamics (Fig. 1D and Figure S2). The concentration of IAA in GD peaked at 150 DAT (163.6 ng/g FW) and remained elevated at 240 DAT (150.0 ng/g FW), ZJ reached its maximum at 210 DAT (235.4 ng/g FW) and remaining the highest level in four accessions through 240 DAT. The content of IBA rose markedly by 30 DAT in all accessions, with ZJ reached the highest concentration at 210 DAT (2.73 ng/g FW, $p < 0.0001$ vs GD, YN, and GZ). GA_3 was highest at 0 DAT in GD, YN, and GZ and declined progressively thereafter, with ZJ exhibiting a secondary peak at 60 DAT prior to its decline. ZR declined from 0 to 30–90 DAT across all accessions before recovering; GD peaked at 150 DAT (14.97 ng/g FW), while ZJ continued to rise and reached the highest concentration at 240 DAT (16.98 ng/g FW, $p < 0.0001$ vs GD, YN, and G). Across four phytohormones, inter-accession differences were most pronounced during the 150–240 DAT window, indicating that provenance-associated hormonal divergence intensifies during the later stages of stem development. Stem cross-sections at 30 and 240 DAT showed structural maturation in all four accessions (Fig. 1E). At 30 DAT, vascular bundles were distributed throughout the ground parenchyma with relatively thin fiber sheaths and cell walls. By 240 DAT, fiber sheaths around vascular bundles had thickened considerably, and intercellular spaces in the parenchymatous tissue were reduced, consistent with cell expansion and progressive tissue compaction during maturation. SEM of fractured cross-sections confirmed substantially increased tissue compaction

at 240 DAT (Supplementary Figure S3), with vascular bundle boundaries more sharply delineated. Collectively, the morphometric and hormonal data indicates that the four accessions pursued divergent developmental trajectories despite shared cultivation conditions.

(A) Sampling timeline and data acquisition across days after transplanting (DAT). **(B)** Representative morphological profiles of the four accessions at 150 DAT. **(C)** Stem length and stem diameter across sampling time points in each accession ($n \geq 6$ plants per accession per time point). Time-point effects were evaluated within each accession by one-way ANOVA followed by Tukey's multiple-comparisons test ($\alpha = 0.05$). **(D)** Endogenous concentrations of indole-3-acetic acid (IAA) and zeatin riboside (ZR) in stems of four accessions ($n = 3$). Data were analyzed by two-way ANOVA with accession and DAT as factors; simple main effects of accession were evaluated at each DAT by Tukey's multiple-comparisons test (adjusted $P < 0.05$). **(E)** Representative haematoxylin and eosin (H&E)-stained stem cross-sections and scanning electron microscopy (SEM) images of stem fracture surfaces at 30 and 240 DAT for each accession; scale bars as indicated. VB= vascular bundles; CW=cell wall.

Temporal regulation of metabolite-defined quality across stem development

Stem samples of four accessions were collected at seven developmental stages (84 samples in total) and analyzed by LC-MS/MS-based untargeted metabolomics in positive and negative ion modes. A total of 1,104 metabolite features were annotated (Table S2), of which 690 and 23 metabolites were assigned to the efficacy-associated (EI) and safety-associated (SI) categories, respectively. Principal component analysis (PCA) of metabolomic profiles revealed considerable overlap among samples across both accessions and developmental stages, with PC1 (8.8%) and PC2 (4.7%) capturing limited proportions of total variance (Fig. 2A). The absence of clear clustering by accession or time point indicates that metabolomic variation is distributed across many orthogonal dimensions rather than dominated by a small number of principal axes, consistent with the high chemical diversity of the 1,104-feature dataset. Per-metabolite variance partitioning confirmed that the developmental stage (mean partial $R^2 = 11.0\%$, permutation $p < 0.01$) accounted for more individual metabolite variation than accession identity (5.5%, permutation $p < 0.01$), suggesting that developmental regulation operates broadly across the metabolome, while provenance-associated effects are concentrated in specific compound classes (Supplementary Figure S6). To characterize how metabolite-defined quality evolves during stem development, the efficacy index (EI), the safety index (SI), and their balance expressed as the harvest suitability score (HSS) were quantified. Across accessions, EI and SI exhibited divergent temporal trajectories, resulting in stage-dependent variation in HSS (Fig. 2B). EI did not differ significantly among accessions at any time point. SI showed significant accession-level variation at 90 and 240 DAT; overall, ZJ had lower SI than both GD ($p = 0.044$) and GZ ($p = 0.030$) HSS trajectories diverged markedly among accessions: HSS was not significantly affected by accession ($P = 0.082$), DAT ($P = 0.064$), or their interaction ($P = 0.146$). These trajectories therefore reflect dynamic, accession-associated variation in metabolite-defined quality states. The chemical composition of the EI pool remained broadly conserved across accessions and developmental stages (Fig. 2C). Fatty acids and derivatives dominated EI (mean 60–66% of total),

followed by carbohydrates and glycosides (10–16%), benzenoid phenolics (3.5–5.5%), and terpenoids (2.8–6.0%); all other categories each contributed less than 4%. SI pool comprised compounds from two chemical super classes: alkaloids and derivatives (22 compounds) and organic nitrogen compounds, the latter consisting exclusively of linoleoyl ethanolamide. The relative contributions of these two classes fluctuated moderately across accessions and developmental stages but showed no clear directional trend: alkaloids and derivatives (22 compounds; mean 46.5% of total SI signal, range 27.3–62.0%) and organic nitrogen compounds (range 38.0–72.7%). Relative z-scores were calculated for 114 EI-classified metabolites across 11 chemical categories (Fig. 2D). Of these, 84 were identified as significant in at least one pairwise-against-rest OPLS-DA comparison. GD exhibited a stronger fatty acid and lipid enrichment signature ($\Delta z = +0.13$), with azelaic acid as the most strongly elevated compound ($\text{rel}Z = +0.60$). ZJ was characterized by terpenoid accumulation ($\Delta z = +0.24$); with citral showing the largest individual deviation ($\text{rel}Z = +0.60$). In contrast, ZJ showed depletion across most phenylpropanoid categories. YN showed the broadest phenylpropanoid enrichment, with above-average z-scores across flavonoids ($\Delta z = +0.19$), phenolic acids and coumarins (+0.13), benzenoid phenolics (+0.14), and lignans (+0.21); the most strongly enriched single compound was hydroxytyrosol ($\text{rel}Z = +0.61$). GZ displayed no clear category-level dominance (14 Up, 14 Down), with only marginal elevation in heterocyclic actives ($\Delta z = +0.04$).

(A) PCA score plot of metabolomic profiles across all 84 biological samples **(B)** Temporal trajectories of EI_z , SI_z , and HSS within each accession across stages. Data are presented as mean ($n = 3$). A significant accession $\times$ DAT interaction was detected only for SI. Tukey-adjusted simple-effects comparisons showed that SI was lower in ZJ than in GD ($P = 0.034$) and GZ ($P = 0.011$) at 90 DAT, and lower in ZJ than in YN ($P = 0.022$) and GZ ($P < 0.001$) at 240 DAT. No significant accession, DAT, or interaction effects were detected for EI or HSS. **(C)** Superclass-level compositional breakdown of EI-contributing metabolites across DAT for each accession (stacked bars, y-axis). **(D)** Accession-specific EI metabolite enrichment profiles across chemical categories. Relative z-scores are shown for 114 EI-classified metabolites grouped by chemical category.

Developmental stage and provenance identity jointly shape transcriptomic divergence

After quality filtering and alignment to the *D. officinale* reference genome (ASM160598v2), 22,566 expressed genes were detected. Of these, 15,858 genes passing a minimum expression threshold following variance-stabilising transformation (DESeq2 VST) were used for PCA (Fig. 3A); the full set of 22,566 genes was retained for differential expression analysis. PCA resolved samples along PC1(16.4%) and PC2(10.2%) according to both provenance and developmental phase, demonstrating that transcriptomic variation was structured by both accession identity and developmental progression (Fig. 3A). Analysis of differentially expressed genes (DEGs) revealed that the numbers of up- and down-regulated DEGs varied among consecutive phase comparisons (Fig. 3B). Across all four accessions, downregulated DEGs consistently outnumbered upregulated DEGs in the majority of consecutive phase transitions, with GD showing the most pronounced bias (Down/Up ratio: 1.72–2.30 across P1 $\rightarrow$ P2,

P2→P3, and P3→P4). ZJ and GZ showed more moderate but consistent Down-biased ratios (ZJ: 1.26–1.27; GZ: 1.22–1.60). In the YN P2→P3 transition, upregulated genes substantially exceeded downregulated genes (1,269 Up vs 708 Down; Down/Up = 0.56), indicating a stage-specific transcriptional activation program unique to this accession and developmental window. KEGG enrichment analysis of the P1→P2 transition showed that both shared and divergent pathway responses among accessions (Fig. 3C). Phenylpropanoid biosynthesis was the only pathway significantly enriched in all four accessions, indicating a common feature of early transcriptional reprogramming. Apart from this shared response, these enriched pathways fell into two groups. GD and ZJ were both enriched in photosynthesis-related pathways (photosynthesis and photosynthesis–antenna proteins), starch and sucrose metabolism, and glutathione metabolism. Photosynthesis was absent from GZ, while photosynthesis–antenna proteins were absent from GD and GZ. YN and GZ shared enrichment for plant–pathogen interaction, the MAPK signalling pathway, cutin–suberin–wax biosynthesis, and plant hormone signal transduction, the latter being most strongly enriched in GZ. YN also showed a notably broader enrichment profile (15 significant pathways), including flavonoid biosynthesis, stilbenoid biosynthesis, and multiple amino acid metabolic pathways, reflecting extensive transcriptional activation at this transition. In later transitions, phenylpropanoid biosynthesis and cutin–suberin–wax biosynthesis recurred across multiple transitions in GD, YN, and ZJ, while GZ and ZJ showed a marked reduction in enriched pathways at P2→P3 (GZ: 9→1; ZJ: 6→4; Supplementary Table S5-6). KEGG enrichment analysis of the six pairwise comparisons at 150 DAT identified both recurrent and comparison-specific signals (Fig. 3D). Plant–pathogen interaction was the most broadly enriched pathway, appearing in four of the six contrasts (GD vs YN, ZJ vs YN, ZJ vs GZ, and YN vs GZ). Phenylpropanoid biosynthesis and stilbenoid, diarylheptanoid and gingerol biosynthesis each appeared in three comparisons, all of which involved ZJ (GD vs ZJ, ZJ vs YN, ZJ vs GZ), pointing to ZJ-specific variation in phenolic secondary metabolism at this stage. The GD vs ZJ contrast additionally showed enrichment for phenylalanine, tyrosine and tryptophan biosynthesis and cysteine and methionine metabolism, extending the amino acid–phenylpropanoid metabolic divergence between these two accessions. MAPK signalling and plant–pathogen interaction were co-enriched in ZJ vs YN, ZJ vs GZ, further highlighting the ZJ-centred signalling divergence. GD vs GZ was enriched for fatty acid elongation and glycine–serine–threonine metabolism, and ZJ vs GZ showed enrichment for flavonoid biosynthesis, alpha-linolenic acid metabolism, and plant hormone signal transduction. GO analysis identified significant terms in three contrasts: photosynthesis and light harvesting in GD vs YN and ZJ vs YN; transcription factor activity terms in GD vs ZJ; and carbohydrate metabolism and photosynthesis-related terms in ZJ vs YN (Supplementary Table S7). Variance partitioning confirmed significant contributions of both accession identity and developmental stage to transcriptomic divergence (permutation $p < 0.001$ for both; Supplementary Figure S5 and Table S9), with the greatest transcriptomic flux occurring during the 0–90 DAT window across all accessions (Supplementary Figure S6).

(A) Principal component analysis (PCA) of the RNA-seq data across all accessions and developmental time points (The phase division information is shown in Supplementary Table S3). **(B)** Overview of differentially expressed genes (DEGs) associated with consecutive phase transitions within each accession, together with P4 vs P1 for GD and YN; P3 vs P1 for ZJ and GZ. **(C)** KEGG enrichment analysis of DEGs identified in the early-phase transition (P2 vs P1) within each accession. The x-axis indicates the

rich factor; bubble size represents the number of DEGs mapped to each pathway; bubble color indicates enrichment significance ($-\log_{10} P$). **(D)** Cross-accession KEGG enrichment analysis of DEGs at 150 DAT derived from six pairwise comparisons among GD, ZJ, YN, and GZ. Bubble size and color are defined as in **(C)**.

Cross-accession transcriptomic divergence at 150 DAT reveals distinct regulatory programs associated with geo-authenticity

The number of DEGs differed markedly across the six pairwise comparisons at 150 DAT (Fig. 4A), ranging from 1,925 (YN vs GZ) to 4,163 (ZJ vs YN). ZJ vs YN yielded the highest total DEG count, with upregulated genes exceeding downregulated genes (2,571 up vs 1,592 down), while the three comparisons involving GD ranked second to fourth (GD vs YN: 3,491; GD vs ZJ: 3,323; GD vs GZ: 1,973). YN vs GZ showed the lowest DEGs and was the only contrast in which downregulated genes outnumbered upregulated genes (1,109 down vs 816 up), indicating relatively limited transcriptional divergence between these two accessions at this stage. Hierarchical clustering partitioned the 7,657 DEGs at 150 DAT into nine expression clusters (G-C1 to G-C9; Fig. 4B). G-C5 was characterized by ZJ-specific activation and GD-specific suppression at 150 DAT, while G-C6 showed the opposite mode, with YN strongly activated and GD suppressed. Differential alternative splicing and TF-family representation among DEGs were assessed across all six pairwise comparisons among the four accessions (Fig. 4C and Supplementary Table S8). Across the six pairwise comparisons, the number of significant differential AS events ranged from 445 (ZJ vs GZ) to 1,964 (GD vs YN) ($FDR < 0.05$, $|\Delta PSI| > 0.1$). Skipped exon (SE) events were the most frequent AS type in five of the six comparisons, accounting for 47.7–59.5% of total. The ZJ vs GZ comparison was the sole exception, with no SE events were detected in this contrast. Instead, alternative 3' and 5' splice site usage (A3SS: 43.6%; A5SS: 26.1%) predominated. Among the 7,657 unique DEGs, 580 unique genes were classified as TFs, representing 51 families and 1,310 TF-DEG occurrences across the six contrasts. When each occurrence was assigned to the accession showing higher expression in the corresponding comparison, YN had the highest cumulative count (431), followed by GD (338), ZJ (296), and GZ (245). The most represented families differed among accessions: WRKY, bHLH, ERF, and NAC predominated in YN; NAC, bHLH, and MYB in GD; bHLH and MYB in ZJ; and ERF in GZ. These patterns indicate accession-dependent TF-family representation among upregulated DEGs rather than direct evidence of TF-family representation. The expression trajectories of 150 DAT DEGs were examined across the full 0–240 DAT window (Fig. 4D). Four clusters displayed constitutively accession-biased expression patterns across all seven time points. G-C1 and G-C4 both maintained YN-dominant expression throughout, with GD and ZJ consistently suppressed, respectively. G-C7 showed a stable GZ-dominant signature, with YN suppressed at all stages. G-C8 displayed GD-dominant expression throughout. Two further clusters showed a stable suppression pattern in one accession despite variable dominance among the others. G-C6 maintained GD suppression across all time points, with the dominant signal alternating between GZ (0–60 DAT) and ZJ (90–240 DAT). G-C9 showed consistently suppressed expression in YN across all time points, while ZJ, GD, and GZ showed broadly similar intermediate levels. The remaining three clusters showed more dynamic stage-associated patterns. G-C3 maintained GD-dominant

expression with ZJ most strongly suppressed at 150 DAT, but showed less consistent patterns at other stages. G-C5 transitioned from GZ dominance in early development (0–60 DAT) to ZJ dominance from 90 DAT onwards, with the ZJ peak at 150 DAT. G-C2 showed persistent YN dominance through most of the developmental window, with ZJ progressively declining from 60 DAT to a minimum at 150 DAT before partially recovering at 240 DAT. In summary, transcriptional divergence among provenances at 150 DAT involved differential gene expression, alternative splicing, and accession-dependent TF-family representation. The stage-dependent trajectories of the expression clusters, together with the accession-biased distribution of TF families among upregulated DEGs, indicate coordinated transcriptional shifts organized in both time and provenance.

(A) DEG burden across pairwise comparisons among accessions at 150 DAT. Up- and down-regulated genes are defined according to the numerator and denominator specified for each pairwise contrast. **(B)** Heatmap of transcriptomic variation across the four provenance-defined accessions at 150 DAT. **(C)** Accession-associated transcriptional differences at 150 DAT at the levels of alternative splicing and TF-family representation. Lower panel: stacked bar chart showing the number of significant differential AS events ($FDR < 0.05$, $|\Delta PSI| > 0.1$) per pairwise comparison, classified by AS type. Upper panel: grouped bar chart showing cumulative accession-upregulated TF-DEG occurrences for the ten most represented TF families across the six pairwise comparisons. **(D)** Developmental expression trajectories of 150 DAT DEGs across accessions (G-C1 to G-C9; $n = 7,657$ total).

Multi-omics integration reveals the transcriptional basis of metabolite-defined quality variation through co-expression module analysis

PCA was performed separately on the metabolomic and transcriptomic datasets (Fig. 5A; full module–trait correlation statistics at phase-point resolution in Supplementary Table S10). In the metabolomic PCA ($PC1 = 8.8\%$, $PC2 = 4.7\%$), individual samples showed considerable overlap across accessions, reflecting the high within-group variability characteristic of untargeted metabolomic profiling. Centroid trajectories nevertheless revealed that the major developmental transitions differed among accessions in both magnitude and direction. Accession-level separation was clearer in the transcriptomic PCA ($PC1 = 11.0\%$, $PC2 = 9.9\%$) than in the metabolome. Four modules showed nominal associations with the metabolite-derived quality indices at the phase level (Fig. 5B). MEcyan correlated positively with both EI and HSS. MEtan was also positively associated with EI. MELightgreen showed significant positive correlations with both HSS and EI, whereas MEgreenyellow was negatively associated with HSS. No module showed a significant association with SI. These exploratory module–trait relationships are consistent with a modular organization of transcriptional variation associated with EI and HSS. The temporal dynamics of representative modules were examined across developmental phases (Fig. 5C). GD and YN were each resolved into four phases, while ZJ and GZ were each resolved into three (see Table S3 for phase definitions). MEtan and MEcyan both showed elevated eigengene values at P1, followed by a sharp decline at P2 that was maintained through later phases; this early-phase activation was most pronounced in YN and GD. Given their positive correlations with both EI and HSS, this pattern points to the preferential

establishment of efficacy-related transcriptional programs during early stem development, at least in GD and YN. MEgreenyellow displayed the opposite trajectory, with generally low values at P1 and a later increase in YN (P4) and GZ (P2–P3), consistent with its negative association with HSS. Module–metabolite correlation analysis using six representative EI-associated compounds confirmed and extended these associations (Fig. 5D). Naringenin correlated significantly with three modules (MElightgreen, MEtan, MEcyan; $p = 0.556–0.631$), suggesting that flavonoid-related regulation is distributed across multiple co-expression programs. MEtan showed the strongest individual module–metabolite correlation (methyl 4-hydroxycinnamate: $p = 0.842$, $p < 0.001$), and MEcyan was associated with moupinamide and afzelechin. MElightgreen was correlated with dopamine ($p = +0.719$) and quercetin ($p = +0.587$), indicating a broader association with the phenolic metabolic network. By contrast, MEgreenyellow showed no significant associations with any of the six compounds, the metabolic basis of its inverse association with HSS therefore remains unresolved. Nine modules were selected for eigengene network analysis and formed a structured, interconnected network rather than operating independently (Fig. 5E). In addition to the four core modules shown in Fig. 5C, five additional modules (MEturquoise, MElightcyan, MEblue, MEpurple, MEbrown) were included based on significant pairwise eigengene correlations with at least one core module. MElightgreen, although a core module, occupied a more peripheral position in the network, connected only to MEcyan and MEturquoise, suggesting a relatively independent transcriptional program among the four trait-associated modules. MEcyan and MEtan were positively correlated, and both shared positive associations with MEturquoise and MElightcyan, defining a coherent positive co-expression cluster. MEgreenyellow, together with MEpurple and MEbrown, formed a separate positive cluster. MEblue was exclusively negatively correlated with all four of its network partners, making it a transcriptional antagonist within the network. Negative edges between MEcyan and MEgreenyellow and between MEtan and MEgreenyellow indicate opposing regulatory dynamics between the two positive clusters, consistent with their contrasting associations with EI- and HSS-related indices. Cross-accession concordance in module activity was assessed from eigengene trajectories across the seven shared DAT points (Fig. 5F). MEtan showed the most consistent temporal pattern, with positive correlations across all six accession pairs and the strongest concordance for GD–ZJ ($p = 0.893$) and GD–GZ ($p = 0.857$). By contrast, MEcyan and MElightgreen showed greater pairwise variability; for example, MElightgreen was highly concordant between GD and ZJ ($p = 0.857$) but showed little similarity between GD and GZ ($p = 0.071$). MEgreenyellow displayed the weakest overall concordance, including a slight negative correlation between GD and GZ. These results indicate that MEtan followed a relatively stable temporal activity pattern across accessions, whereas MElightgreen and MEgreenyellow were more accession dependent. Collectively, these results indicate that quality-related transcriptional programs in *D. officinale* are organized into a structured co-expression network, in which individual modules display distinct associations with EI- and HSS-related indices, characteristic developmental trajectories, and variable conservation across genetic backgrounds.

Figure 4 | Divergence among provenance-defined accessions at 150 DAT reveals distinct transcriptional programs associated with geo-authenticity.

(A) DEG burden across pairwise comparisons among accessions at 150 DAT. Up- and down-regulated genes are defined according to the numerator and denominator specified for each pairwise contrast. **(B)**

Heatmap of transcriptomic variation across the four provenance-defined accessions at 150 DAT. **(C)** Accession-associated transcriptional differences at 150 DAT at the levels of alternative splicing and TF-family representation. Lower panel: stacked bar chart showing the number of significant differential AS events ($FDR < 0.05$, $|\Delta PSI| > 0.1$) per pairwise comparison, classified by AS type. Upper panel: grouped bar chart showing cumulative accession-upregulated TF-DEG occurrences for the ten most represented TF families across the six pairwise comparisons. **(D)** Developmental expression trajectories of 150 DAT DEGs across accessions (G-C1 to G-C9; $n = 7,657$ total).

Figure 5 | Multi-omics integration reveals the transcriptional basis of metabolite-defined quality variation through co-expression module analysis

(A) Phase-level co-variation between transcriptomic and metabolomic trajectories across accessions. PCA of phase centroids was performed on transcriptomic and metabolomic datasets separately. **(B)** Spearman correlations between module eigengenes and EIz, SIz, and HSS across 84 biological samples. Asterisks indicate nominal significance ($P < 0.05$, Spearman test). **(C)** Phase trajectories of representative module eigengenes across developmental phases within each accession. Modules were selected based on their nominal associations with EI (MEtan, MElightgreen, and MEcyan) and HSS (MElightgreen, MEgreenyellow, and MEcyan). **(D)** Bubble plot of Spearman correlations between module eigengenes and six representative EI-associated compounds spanning multiple chemical classes. Bubble color indicates Spearman's ρ , and bubble size indicates $-\log_{10}(P)$. **(E)** Eigengene correlation network of selected modules. Nodes represent module eigengenes; edges represent pairwise Spearman correlations ($|\rho| \geq 0.35$, $P < 0.05$). Edge width is proportional to $|\rho|$. **(F)** Cross-accession concordance of module eigengene trajectories. Higher ρ values indicate greater similarity in the temporal activity pattern of the module between accessions.

Integration of module–trait associations reveals the transcriptional basis of quality and agronomic trait variation in *D. officinale*

Module–trait relationships were assessed across 28 accessions $\times$ DAT means and interpreted as exploratory because significance was based on nominal P values (Fig. 6A; Supplementary Table S11). MEgreen showed the strongest inverse association with stem diameter ($\rho = -0.622$, $P < 0.001$), whereas MEtan was positively associated with ZR and GA₃, MESalmon was negatively associated with IBA, IAA, and stem length, and MElightgreen was positively associated with EI and stem diameter. Additional nominal associations with stem diameter and hormone traits were distributed across several modules. Four representative relationships are shown in Fig. 6B: MEgreen–stem diameter, MEtan–ZR ($\rho = 0.541$, $P = 0.003$), MElightgreen–EI, and MESalmon–IBA. MEgreen was most strongly associated with stem diameter. MEtan correlated positively with ZR content ($\rho = 0.541$, $P = 0.003$), pointing to a connection with cytokinin-mediated developmental regulation. MElightgreen was positively associated with EI, implicating this module in efficacy-related metabolite accumulation. MESalmon correlated negatively with IBA content. To further characterize the functional identity of trait-associated modules, gene set enrichment analysis was performed on the four selected module (Fig. 6C). MEtan was enriched in phenylpropanoid biosynthesis and microtubule-based processes. These enrichments are consistent with its positive

associations with ZR and GA₃, may suggesting involvement in both cytoskeletal organization and secondary metabolic regulation. MEgreen was strongly enriched in photosynthesis-related terms and chlorophyll biosynthesis. MESalmon was enriched in aminoacyl-tRNA biosynthesis and pyruvate metabolism, implicating translational control and central carbon metabolism as functional features of this auxin-associated module. MELightgreen was enriched in glycolysis/gluconeogenesis and tyrosine metabolism, consistent with its positive association with EI. These enrichment patterns provide functional context for the module–trait associations in Fig. 6A, and prompted promoter cis-element analysis of seven candidate hub genes across the four modules (Fig. 6D). bHLH-binding sites (142 occurrences), cytokinin-responsive elements (49 occurrences), and MYB-binding motifs (40 occurrences) were the most frequent and were present in all seven gene promoters. WRKY-binding and ABA-responsive elements were detected in six and five genes, respectively; auxin- and GA-responsive motifs occurred sparsely. The co-occurrence of multiple motif classes within individual promoters suggests that these hub genes are subject to combinatorial regulation involving both hormone-responsive and TF-mediated inputs. Based on the module–trait associations, functional enrichment profiles, and promoter cis-element patterns described above, a working model of quality formation in *D. officinale* is proposed (Fig. 6E).

Four co-expression modules receive distinct upstream inputs and converge on metabolic allocation between EI- and SI-associated chemical dimensions. MEtan and MESalmon are linked to hormonal signals (ZR, GA₃, IAA, IBA), with candidate regulatory TFs including WRKYs, LBDs, MYBs, and NACs (MEtan) and AP2/ERFs, ARR, and kinases (MESalmon). MEgreen and MELightgreen are associated with developmental and metabolic signals, respectively, with TF enrichment profiles reflecting photosynthetic and glycolytic programs. Candidate hub genes within each module – Cesa8-like and CAD2-like (MEtan), MPEC-like (MEgreen), FBA-like and PDC1-like (MELightgreen), and eIF4G-like and CUL1-like (MESalmon) – represent putative functional nodes linking transcriptional output to downstream metabolism. Taken together, these results implicate transcriptional module activity in the spatiotemporal metabolic variation underlying geo-authenticity in *D. officinale*. The partially coordinated, accession-specific organisation of hormone-responsive, developmental, and metabolic regulatory programs offers a transcriptomic basis for the accession- and stage-dependent divergence of both agronomic traits and metabolite-defined quality indices – the latter captured through EI, SI, and their integrated balance as HSS – collectively reflecting the multi-dimensional nature of geo-authenticity quality as operationally defined in this study.

(A) Spearman correlations between module eigengenes and nine trait variables – including metabolite-based quality indices (EI, SI, HSS), stem morphology (stem length, stem diameter), and phytohormones (IAA, IBA, ZR, GA₃) – calculated across 28 accessions × DAT-point means. **(B)** Representative module eigengene–trait relationships scatter plots. Associations are shown for MEgreen vs. stem diameter, MEtan vs. ZR content, MELightgreen vs. EI, and MESalmon vs. IBA content (n = 28 accession × DAT-point means). Lines indicate fitted regression trends. **(C)** Functional enrichment of trait-associated modules. Significantly enriched KEGG pathways and GO Biological Process terms ($P < 0.05$, Fisher's test) are shown for MEtan, MEgreen, MESalmon, and MELightgreen. **(D)** Promoter cis-element analysis of candidate hub genes. Heatmap (left) shows counts of nine cis-regulatory motif categories in 2 kb upstream promoter regions of seven hub genes across four modules. Positional distribution of motifs is shown on the right;

triangles indicate strand orientation (▲ +, ▼ -). **(E)** Working model of transcriptional regulation underlying quality formation in *D. officinale*. Four co-expression modules (MEtan, MEgreen, MELightgreen, MEsalmon) are proposed to link hormonal and developmental signals to metabolic allocation between EI-associated (phenylpropanoid and lipid-related) and SI-associated (alkaloid and organonitrogen) chemical pools. The resulting EI–SI balance is captured by HSS and reflects accession-dependent developmental trajectories. Hub genes within each module represent candidate regulatory nodes. Dashed arrows indicate inferred rather than experimentally validated relationships.

Discussion

Geo-authenticity in medicinal plants is generally understood as origin-dependent differences in morphology, chemical composition, and therapeutic properties, reflecting the combined influence of genetic factors and long-term environmental adaptation in specific producing regions [9, 17, 18, 19]. In *D. officinale*, previous studies have shown that materials from different producing regions have differences in agronomic traits, chemical fingerprints, metabolite composition, and stress responses, yet these studies have largely been confined to single-point comparisons that capture the endpoint of quality differentiation without addressing how it is progressively established over development [20, 21]. The molecular and developmental basis by which provenance shapes medicinal quality of *D. officinale* has therefore remained poorly resolved. By comparing four provenance-defined accessions under a common-garden framework across seven developmental time points, the present study finds that geo-authenticity in *D. officinale* is not a fixed provenance label but a temporally structured quality phenotype that emerges through the coordinated differentiation of developmental, transcriptional, and metabolic programs.

The chemical basis of geo-authenticity is quantitative, not qualitative

In medicinal plants, secondary metabolites are major contributors to the pharmacological activities and quality characteristics of medicinal plants [22], their accumulation is influenced by geographical origin and developmental stage [23–27]. When quality was assessed dynamically rather than at a single point, provenance-dependent differences became more apparent. HSS trajectories diverged markedly: ZJ peaked at 90 DAT, YN maintained elevated values through 150 DAT, GD peaked late at 240 DAT, and GZ showed suppressed intermediate-stage values. Overall HSS was significantly higher in ZJ and YN than in GZ, while SI differed among accessions only at 90 and 240 DAT, consistent with alkaloid accumulation being subject to tighter developmental constraint than the broader EI pool. The class-level composition of both EI- and SI-associated metabolite pools was broadly conserved across accessions and developmental stages, with provenance-associated differences expressed primarily as shifts in the relative abundance of individual compounds within a stable chemical framework. In the EI pool, fatty acids and derivatives dominated (60–66%) across all accessions, with terpenoids enriched in ZJ, phenylpropanoids and flavonoids in YN, and lipid-related compounds in GD. The SI pool was similarly stable in class composition. This compositional invariance is reflected in the metabolomic PCA, where samples overlapped suggesting that the metabolome acts as an integrative and buffered layer, reflecting the

cumulative outcome of biosynthesis, transport, storage, and turnover ^[28]. Variance partitioning indicating that ontogenetic programming is the dominant driver of metabolomic variation across the development stage. This suggests that provenance-associated metabolic differences in *D. officinale* are not uniformly expressed throughout development but emerge preferentially at specific stages — a dynamic that would be invisible to single-point comparisons and that motivates the time-resolved sampling strategy adopted here. Despite the limited accession-level separation in metabolomic PCA, OPLS-DA of 114 representative EI compounds identified 84 with significant provenance-associated patterns, confirming that accession effects on individual metabolite abundance are pervasive but distributed across many chemical dimensions rather than concentrated along the leading axes of variance. These findings are consistent with the view that geo-authenticity reflects quantitative rather than qualitative metabolic divergence — a pattern also reported for *Atractylodes lancea*, where chemotypes are defined by differential marker accumulation rather than class-level compositional shifts ^[9] — and suggest that harvest timing should be calibrated to provenance rather than applied uniformly.

Developmental programming and provenance interact to shape the transcriptome

During long-term process selection and environmental adaptation, geographically distinct plant populations may establish differentiated regulatory networks, and their molecular basis can be revealed through transcriptome analysis ^[18, 29, 30]. Variance partitioning of the metabolome confirmed that developmental stage accounted for a greater proportion of individual metabolite variation (mean partial R^2 = 11.0%) than accession identity (5.5%), and PCA of the transcriptome resolved samples primarily along developmental rather than provenance axes. These findings are broadly consistent with transcriptomic studies of other medicinal plants in which stage-driven transcriptional transitions represent the dominant source of gene expression variation, with genotypic effects becoming most apparent at specific developmental windows ^[10,31]. In this study, the greatest transcriptomic flux occurred during the 0–90 DAT window, coinciding with the period of most rapid morphophysiological change: early stem elongation, declining ZR concentrations, and the P1→P2 pathway reprogramming that engaged phenylpropanoid, photosynthetic, and defence-related programs across all four accessions. The enrichment of phenylpropanoid biosynthesis at P1→P2 in all four accessions, and its recurrence across multiple later transitions in GD, YN, and ZJ, points to this pathway as a constitutive feature of the developmental transcriptional programme in *D. officinale* stems regardless of provenance. The accession-specific overlay on this shared programme — photosynthesis and carbon metabolism enriched in GD and ZJ but not YN and GZ, and plant hormone signal transduction most strongly enriched in GZ — suggests that provenance encodes a preferential allocation of transcriptional resources to distinct metabolic strategies during otherwise shared developmental transitions. A comparable partition between conserved pathway activation and genotype-specific regulatory emphasis has been described in ecotype comparisons of *Salvia* and *Paeonia*, where the developmental timing of secondary metabolite biosynthesis is broadly conserved but the magnitude and metabolic context of pathway activation differ across origins ^[10]. At 150 DAT, YN showed the broadest transcriptional response and the highest cumulative number of accession-

upregulated TF–DEG occurrences, with WRKY, bHLH, ERF, and NAC prominently represented; GD was enriched in NAC, bHLH, and MYB, ZJ in bHLH and MYB, and GZ in ERF. Although these counts reflect TF-family representation among pairwise DEGs rather than absolute TF abundance or activity, their concordance with DEG load, expression-cluster patterns, and alternative-splicing profiles supports distinct accession-specific transcriptional and post-transcriptional programs. Notably, YN achieves this broad secondary metabolic profile without a proportional increase in WRKY or NAC activation; these two families were most strongly upregulated in ZJ. This pattern suggests that phenylpropanoid accumulation in YN may rely more on constitutive transcriptional programs than on defence-responsive regulatory circuits. By contrast, GD's transcriptional dominance in fatty acid- and lipid-associated gene clusters, combined with lower overall TF-family representation, points to a more metabolically focused regulatory strategy centred on lipid metabolism rather than broad secondary metabolic remodelling.

GZ showed the least transcriptional divergence from other accessions, in line with its undifferentiated metabolic fingerprint in OPLS-DA. Beyond differential gene expression, the ZJ vs GZ contrast was marked by a qualitatively distinct AS profile, dominated by alternative splice site usage rather than the exon-skipping events prevalent in all other pairwise comparisons. This shift in AS type distribution suggests that transcriptional divergence between ZJ and GZ may extend beyond mRNA abundance to the post-transcriptional processing of specific regulatory genes, though the functional relevance of this pattern remains unclear. The accession-specific patterns in DEG magnitude, cluster expression, TF-family representation, and AS profiles point to a coordinated reconfiguration of distinct regulatory programs in each accession at 150 DAT. This fits with the idea that defence signalling and phenylpropanoid metabolism are major channels through which adaptive secondary metabolic diversity is distributed among plant genotypes [3, 18].

Co-expression modules are associated with hormonal, developmental, and metabolic variation

WGCNA identified 20 non-grey modules in the *D. officinale* transcriptome, four of which showed nominal associations with quality indices or agronomic traits. MEtan was associated with ZR and GA₃ and enriched for phenylpropanoid biosynthesis (KEGG, $p < 0.001$). Its hub genes, Cesa8-like (LOC110110394, kME = + 0.539) and CAD2-like (LOC110116647, kME = + 0.516), connect cell wall formation to monolignol biosynthesis. CesaA proteins are core components of the cellulose synthase complex, while CAD enzymes catalyze the reduction of cinnamaldehydes to monolignols – the penultimate step of lignin biosynthesis. In *Arabidopsis* stems, CAD-C and CAD-D largely determine lignin composition [33]. That these two genes are co-regulated within a single ZR/GA₃-associated module raises the possibility that cytokinin and gibberellin signals coordinate cell wall reinforcement and phenylpropanoid flux as part of a shared developmental programme rather than through independent regulatory inputs. The observed strong eigengene–metabolite correlations between MEtan and phenylpropanoid compounds – methyl 4-hydroxycinnamate, moupinamide, and naringenin— provide metabolic support for this interpretation and are consistent with comparative transcriptome studies of *Salvia miltiorrhiza* varieties showing that

cytokinin and gibberellin-related signalling converges on phenylpropanoid pathway gene activation during secondary metabolite accumulation ^[31].

MEgreen was negatively correlated with stem diameter and strongly enriched for photosynthesis-related pathways (KEGG, $p < 10^{-44}$); its hub gene MPEC-like (LOC110113521, kME = + 0.279) is annotated to chlorophyll biosynthesis and porphyrin metabolism. The inverse association between MEgreen activity and stem diameter may reflect differences in carbon allocation between photosynthetic and structural programs. Such competition between photosynthetic and growth-associated programs has been documented in woody species, where phenylpropanoid biosynthesis and phytohormone signalling jointly regulate the transition between cell division and secondary wall deposition ^[34]. YN maintained the largest stem diameter throughout development while simultaneously showing the broadest phenylpropanoid enrichment, that suggests its developmental programme does not trade off structural growth against secondary metabolite accumulation to the same extent as the other accessions. While GZ showed more coordinated trajectories between MEgreen activity and stem diameter, consistent with a tighter coupling between photosynthetic resource allocation and radial growth in this accession. MELightgreen, associated with EI ($p = +0.441$, $p = 0.019$) and enriched for glycolysis/gluconeogenesis, carries hub genes FBA-like and PDC1-like that encode central carbon metabolic enzymes. Loss of plastidial FBA function in *Arabidopsis* significantly reduced shikimate-derived aromatic amino acid pools including phenylalanine and tyrosine ^[35], suggesting that FBA-like activity gates carbon entry into both phenylpropanoid and alkaloid biosynthetic routes. The positive correlation between MELightgreen and EI suggests that primary carbon partitioning through glycolysis may contribute to the regulation of efficacy-associated secondary metabolite accumulation, potentially by influencing the availability of shikimate pathway precursors for phenylpropanoid and alkaloid biosynthesis ^[36].

MEsalmon was negatively associated with both IAA and IBA and enriched for aminoacyl-tRNA biosynthesis and pyruvate metabolism. Its hub genes eIF4G-like and CUL1-like implicate translational control and SCF-mediated proteolysis in auxin signal transduction: the CUL1-containing SCF^{TIR1} complex ubiquitinates AUX/IAA repressors for 26S proteasome-dependent degradation, thereby permitting ARF-mediated auxin-responsive transcription ^[36, 37, 38]. The co-regulation of CUL1-like with a translation initiation factor within MEsalmon aligns with evidence that auxin signalling intersects with translational control through the TOR pathway to influence the production of developmental regulators ^[39]. This annotation-based interpretation raises the possibility

consistent with a carbon allocation trade-off The strong eigengene–metabolite correlations that MEsalmon may coordinate both protein degradation and translational regulation of auxin-responsive gene expression during stem development. That MEsalmon is suppressed during the auxin-dominated early growth phase – IAA and IBA rose sharply across all accessions by 30 DAT – and may be reinstated as auxin concentrations plateau in later development suggests it represents a component of the regulatory switch governing the developmental transition from rapid elongation to secondary metabolite accumulation in *D. officinale*.

The exploratory module–trait correlations described above identified candidate relationships between individual modules and phenotypic or metabolic variables. The WGCNA analysis in Fig. 5 addresses a complementary set of questions: how these modules relate to one another, how their trajectories vary across accessions, and what such network-level patterns can tell us about the transcriptional basis of geo-authenticity. The eigengene network resolved a positive co-expression cluster formed by MEtan and MEcyan, both peaking early in development (Supplementary Figure S9). These two modules were opposed by MEgreenyellow, which rose in later phases and was negatively correlated with both modules. This opposite pattern, in which a quality-promoting transcriptional cluster is counterbalanced by a module with opposite temporal dynamics, suggests that EI accumulation requires not only activation of biosynthetic pathways but also concurrent suppression of competing programs — a regulatory logic increasingly recognised in plant secondary metabolism [3, 18]. Cross-accession conservation analysis reinforces this view. MEtan was the most conserved module across accession pairs, while MEgreenyellow was the least conserved. This pattern suggests that provenance-specific quality differences reside less in the core biosynthetic modules than in the accession-specific tuning of suppressive regulatory programs. In this respect, geo-authenticity in *D. officinale* can be viewed as a shared developmental scaffold upon which accession-specific variation in antagonistic regulatory balance is superimposed.

The HSS framework and its implications for quality assessment

EI and SI were integrated into a Harvest Suitability Score (HSS) that tracks the balance between efficacy- and safety-associated metabolite pools at each time point. Across the full developmental window, HSS was significantly higher in ZJ and YN than in GZ, yet these differences did not reach significance at any single time point, highlighting the limits of single-point comparisons when assessing quality differences among accessions. The HSS trajectories diverged here is a direct expression of geo-authenticity at the metabolite-quality level. GD peaked late, ZJ showed a clear mid-developmental maximum, and YN maintained elevated values through 150 DAT before declining. Provenance comparisons in medicinal plant research are most commonly conducted at a single developmental stage, typically the conventional harvest point. The present data suggest that this practice may miss accession-level quality differences that are real and reproducible at the population level. Developmentally resolved, multi-point assessment is therefore not simply a refinement of existing approaches but a necessary condition for characterising provenance-dependent quality variation in species where metabolite accumulation is temporally regulated.

Together, our findings support a systems-level model in which geo-authenticity in *D. officinale* emerges from the interaction between developmental timing, transcriptional regulatory architecture, and metabolite allocation. Under common cultivation conditions, the four provenance-defined accessions retained distinct trajectories in morphology, anatomy, phytohormones, transcriptomic programs, and metabolite profiles. This indicates that provenance-associated quality differentiation is rooted in dynamic developmental regulation rather than static geographic origin alone. The partially coordinated, accession-specific organization of hormone-responsive, developmental, and metabolic programs provides

a transcriptomic basis for the accession- and stage-dependent divergence of agronomic traits and quality indices – defined here through EI, SI, and their integrated balance as HSS – capturing the multidimensional nature of geo-authenticity quality as operationally defined in this study. This study was conducted on tissue-culture-derived material under controlled cultivation conditions, which, while enabling direct developmental comparison among accessions, may not fully recapitulate the secondary metabolite profiles or transcriptional programs expressed in field-grown plants from traditional geo-authenticity producing regions. The regulatory relationships identified here remain correlative and have not yet been functionally validated.

Conclusion

This study provides a multi-dimensional characterisation of developmental and provenance-associated variation in *D. officinale* stems, integrating untargeted metabolomics and whole-transcriptome sequencing across four accessions and seven time points under controlled cultivation conditions. These results indicate that geo-authenticity in this species is not a static attribute of geographical origin. Instead, it is a quality phenotype that takes shape over developmental time, arising from shared developmental programs and accession-specific regulatory divergence. At the metabolic level, the EI and SI pools retained similar class compositions across accessions and stages, with provenance-associated differences appearing largely as quantitative variation in individual metabolites rather than as shifts in compound categories. At the transcriptional level, developmental stage dominated gene-expression variation, yet accession-specific programs emerged at key developmental transitions—most prominently at 150 DAT, where each accession exhibited a distinct combination of DEG burden, pathway enrichment, TF-family representation among upregulated DEGs, and alternative-splicing profiles. Co-expression network analysis prioritized four modules showing nominal associations with quality indices, agronomic traits, or phytohormones. MEtan was associated with ZR- and GA₃-mediated hormonal signals to phenylpropanoid and cell wall biosynthesis. MEgreen showed an inverse association with stem diameter and was enriched for photosynthesis-related functions. MELightgreen was associated with primary carbon partitioning and EI accumulation. MESalmon showed inverse associations with auxin-responsive proteolysis and contained candidate genes involved in proteolysis and translational regulation. The network-level features of the WGCNA further illuminate the transcriptional basis of geo-authenticity in metabolism. The opposing dynamics between MEtan/MEcyan and MEgreenyellow suggest that EI accumulation depends on concurrent suppression of competing transcriptional programmes, and the higher cross-accession conservation of MEtan relative to MEgreenyellow implies that provenance-specific quality variations are encoded in the tuning of suppressive regulatory balance rather than in core biosynthetic modules. Overall, these results indicated that geo-authenticity in *D. officinale*, when examined under common-garden conditions, is a developmentally regulated phenotype. A broadly shared transcriptional scaffold is modulated by accession-specific regulatory programs, generating distinct metabolic outcomes at specific developmental stages. This work provides a foundation for evidence-based optimisation of harvest timing and cultivation strategies in *D. officinale*, and offers an approach for the molecular characterisation of geo-authenticity in other medicinal plant species.

Materials and methods

Plant materials, experimental design, and sample collection

Four geographically distinct accessions of *Dendrobium officinale* originating from Guangdong, Zhejiang, Guizhou, and Yunnan were used in this study and are hereafter referred to as the provenance-defined accessions GD, ZJ, GZ, and YN. The materials were obtained as tissue-cultured plantlets and all plant materials were formally identified as *D. officinale* by Prof. De-Po Yang. Voucher specimens representing the four accessions were deposited in the collection of the Guangzhou Caizhilin Intangible Cultural Heritage Museum, Guangzhou, China, under voucher numbers CZLFYWHGDO001-4. All materials used in this study were artificially propagated and obtained from cultivated sources; no plants were collected from the wild or from protected areas. Four accessions were supplied by Guangzhou Caizhilin Pharmaceutical Co., Ltd. (Guangzhou, China), with documented cultivation origins in Shaoguan, Guangdong; Wenzhou, Zhejiang; Tongren, Guizhou; and Wenshan, Yunnan, respectively. All plants were cultivated in a greenhouse at Guangzhou Caizhilin Pharmaceutical Co., Ltd., Guangzhou, Guangdong, China, under natural light, with temperature maintained at $25 \pm 3^\circ\text{C}$ and relative humidity at $60 \pm 5\%$. To minimize interference of pre-cultivation conditions, rooted tissue-cultured seedlings were pre-cultured on half-strength Murashige and Skoog (1/2 MS) medium for 60 days before transplanting into plastic pots (8 cm in diameter $\times$ 15 cm in height) filled with pine bark (5–12 mm particle size). Sampling was performed at 0, 30, 60, 90, 150, 210, and 240 days after transplanting (DAT). At each sampling point, eight plants from each accession were randomly selected for analysis, and three biological replicates per accession were collected for metabolomic and transcriptomic analyses, yielding a total of 84 omics samples (Supplementary Table S1). After harvest, sample surfaces were rinsed with running water to remove dust, and residual moisture was blotted off with paper towels. Stem tissue from the middle portion of each sample, excluding the apical and basal regions, was excised, immediately frozen in liquid nitrogen, and stored at -80°C until further use. Developmental phases were defined independently for each accession from variance-stabilized transcriptomic profiles. In the WGCNA analysis, GD and YN were classified into four phases, whereas ZJ and GZ were classified into three phases. Detailed phase groupings are provided in Table S3.

Measurement of morphological traits, stem anatomy, and scanning electron microscopy (SEM)

Morphological traits were documented using an i850 scanner (Microtek, China). Stem length and stem diameter were measured with a vernier caliper, and sample weight was recorded at the same time. At least eight biological replicates were analyzed for each accession at each sampling time point. For anatomical and micromorphological observations, stem segments approximately 5 mm in length were sampled for transverse sectioning and scanning electron microscopy (SEM). Histological samples were prepared by standard paraffin embedding, sectioned with a rotary microtome, and the resulting stem cross-sections were observed and photographed under a light microscope (CX43, Olympus, Japan). For SEM, samples were fixed in glutaraldehyde at 4°C for 4 h and then prepared according to published

procedures^[40]. Observations were carried out using a Sigma 300 scanning electron microscope (Carl Zeiss, Germany).

Determination of endogenous phytohormones

At each sampling time point, at least three plants were sampled from each provenance-defined accession. Stem tissues from the middle portion of the plants were collected, immediately frozen in liquid nitrogen, and stored at -80°C . The contents of indole-3-acetic acid (IAA), indole-3-butyric acid (IBA), zeatin riboside (ZR), and gibberellic acid (GA_3) were quantified using ELISA kits (Shanghai MLBIO, China; ml095279, ml077218, ml077225, and ml077232, respectively). Values represent relative trends across accessions and stages.

Untargeted metabolomic analysis and calculation of EI, SI, and HSS

Metabolites were extracted from 20 mg of sample, followed by homogenization, sonicated at room temperature for 30 min, and then incubated in an ice bath for 30 min. After centrifugation at 12,000 rpm for 10 min at 4°C , the supernatant was collected, dried, and reconstituted in 200 μL of 50% acetonitrile containing 2-chloro-L-phenylalanine (4 ppm). The solution was filtered through a 0.22 μm membrane and subjected to LC–MS analysis. Chromatographic separation was performed on a Vanquish UHPLC system (Thermo Fisher Scientific, USA) using an ACQUITY UPLC HSS T3 column (2.1×100 mm, 1.8 μm ; Waters, Milford, MA, USA)^[41]. The raw metabolomics data have been deposited in the China National Center for Bioinformation (CNCB) under BioProject accession PRJCA067216. Sample-level extremes in PCA score space were screened using squared Mahalanobis distances calculated from PC1 and PC2 after $\log_2(x + 1)$ transformation and feature-wise autoscaling. Five samples exceeded the 97.5% χ^2 threshold for two dimensions ($D^2 > 7.378$). Four of these samples had total ion counts within the biological-sample mean ± 2 SD range, and QC replicates showed a coefficient of variation of 2.0%, confirming instrument stability throughout the analytical run. These samples were therefore classified as biological outliers reflecting genuine metabolic variation and were retained in all downstream analyses. The efficacy index (EI), safety index (SI), and harvest suitability score (HSS) were calculated for each sample based on metabolite peak areas, with positive- and negative-ion mode data combined prior to analysis. Raw metabolite abundances were log-transformed as $\log_2(x + 1)$ to reduce heteroscedasticity, and each metabolite was subsequently standardized across biological samples within the same batch using feature-wise z-score normalization. Metabolites were classified into efficacy-associated (EI) and safety-associated (SI) groups based on a combination of chemical taxonomy, reported bioactivity, and structural characteristics. Superclass annotation was initially obtained from standard metabolite databases and manually curated. Given the chemical diversity within certain superclasses, particularly organoheterocyclic compounds, assignment to EI or SI was performed at the subclass or compound level rather than relying solely on superclass annotation. Specifically, metabolites were categorized into EI if they belonged to classes commonly associated with bioactive secondary metabolism in *Dendrobium* and related medicinal plants, including phenylpropanoids and polyketides, benzenoids, lignans and neolignans, lipids and lipid-like molecules, organic oxygen compounds, and selected organoheterocyclic subclasses with documented

phytochemical activity (e.g., indoles, benzopyrans, quinolines, and lactones). Metabolites were assigned to SI if they were enriched in nitrogen-containing or structurally related classes, including alkaloids and derivatives, organic nitrogen compounds, and selected organoheterocyclic subclasses with potential safety relevance (e.g., tropane alkaloids, pyrrolizidines, benzimidazoles, and triazines), as well as organohalogen, hydrocarbon, and organometallic compounds. EI was defined as the sum of normalized abundances of efficacy-related metabolite classes, including lipids and lipid-like molecules (n = 268), organic oxygen compounds (n = 109), phenylpropanoids and polyketides (n = 108), benzenoids (n = 91), organoheterocyclic compounds with phytochemical activity (n = 90), and lignans, neolignans and related compounds (n = 17), together with a small number of annotated metabolites from additional superclasses reassigned to EI based on documented bioactivity in *D. officinale* (n = 6; total EI n = 690). SI was initially annotated as comprising 46 metabolite features across four chemical superclasses. Following conservative filtering to exclude compounds whose presence in tissue-culture-derived material could not be attributed to endogenous biosynthesis – including agrochemical residues, pharmaceutical metabolites, synthetic heterocycles, organohalogen compounds, and industrial amines (n = 23 excluded). The final SI pool comprised 23 metabolites: alkaloids and derivatives (n = 22) and organic nitrogen compounds (n = 1, Linoleoyl ethanolamide). Assignment of metabolites to chemical classes was based on the metabolite annotation table (Table S4). To obtain a harvest-oriented composite metric, EI and SI were further standardized across samples within each batch by z-transformation, and HSS was calculated as:

$$\sqrt{\text{varvec}H\text{varvec}S\text{varvec}S} = \sqrt{\text{varvec}z(\text{varvec}E\text{varvec}I)} \otimes \sqrt{\text{varvec}z(\text{varvec}S\text{varvec}I)}$$

HSS was developed as a relative composite index to describe the balance between EI- and SI-associated chemical dimensions during the sampled developmental period. In the present study, HSS is used for longitudinal comparison of metabolite-defined quality states and should not be interpreted as evidence of an optimal harvest time. The EI and SI are operational indices constructed for comparative analysis within the present dataset and should not be interpreted as direct measurements of pharmacological efficacy or toxicological safety.

RNA extraction, sequencing, and transcriptomic analysis

Total RNA was extracted from *D. officinale* samples using the RNeasy Pure Plant Kit (Qiagen, DP441) according to the manufacturer's instructions. RNA concentration and purity were determined using a NanoDrop spectrophotometer (Thermo Scientific, USA), and RNA integrity was assessed with an Agilent 2100 Bioanalyzer (Agilent Technologies, USA). Three biological replicates were included for each treatment. Sequencing libraries were constructed and purified using the AMPure XP system, and library quality was evaluated using the Agilent High Sensitivity DNA Kit on the Agilent 2100 Bioanalyzer. The libraries were then sequenced on the Illumina NovaSeq 6000 platform by Shanghai Personal Biotechnology Co., Ltd.. The raw sequencing data have been deposited in the China National Center for Bioinformation (CNCB) under BioProject accession PRJCA067216. Read count data were subjected to variance-stabilising transformation (VST) using DESeq2 (v1.38.0). Genes passing a minimum expression threshold following VST (15,858 genes) were retained for principal component analysis (PCA), which was performed using the plotPCA function in DESeq2 and visualised with ggplot2 (v3.5.0). Differential

expression analysis was performed between consecutive developmental phases within each accession and between accession pairs at 150 DAT, using DESeq2 (v1.38.0) with default parameters. Genes with $|\log_2 \text{ fold change}| > 1$ and $\text{FDR} < 0.05$ (Benjamini–Hochberg correction) were defined as differentially expressed. The full set of 22,566 genes was used as the background for all comparisons. For visualization, clustering, and WGCNA input, expression values were additionally normalized to $\log_2(\text{FPKM} + 1)$. For functional annotation, DEGs were mapped to Gene Ontology (GO) terms. GO enrichment analysis was performed using the topGO package (v2.40.0), and significance was assessed using the hypergeometric test, with $P < 0.05$ considered statistically significant. KEGG pathway enrichment analysis was conducted using the clusterProfiler package (v3.16.1), and pathways with $P < 0.05$ were considered significantly enriched. TF genes within the DEG set were identified by matching gene annotations against the PlantTFDB (v5.0) classification. Candidate transcription factors were identified from the 150 DAT DEG set by cross-referencing gene annotations. TF genes within the 150-DAT DEG set were identified by matching unique Gene IDs to the *D. catenatum* TF reference list obtained from PlantTFDB TExt. Of the 7,657 unique DEGs, 580 were classified as TF genes (7.58%), representing 51 TF families. Unique TF genes and cumulative TF-DEG occurrences across the six pairwise comparisons were summarized separately. For accession-level summaries, each TF-DEG occurrence was assigned to the accession showing higher expression in the corresponding pairwise contrast. Alternative splicing (AS) was analyzed using rMATS-turbo (version [version]) with coordinate-sorted BAM files and the corresponding GTF annotation as input; biological replicates within each group were analyzed jointly. Five AS types were evaluated: skipped exon (SE), alternative 5' splice site (A5SS), alternative 3' splice site (A3SS), mutually exclusive exons (MXE), and retained intron (RI), and events were quantified using junction-spanning reads. The inclusion level of each event was expressed as percent spliced in (PSI), calculated as $(I/L_I)/(I/L_I + S/L_S)$, where I and S denote reads supporting the inclusion and skipping isoforms, respectively, and L_I and L_S are their effective lengths. Differential AS events were identified using the rMATS likelihood-ratio test, with Benjamini–Hochberg-adjusted $\text{FDR} < 0.05$ and $|\Delta\text{PSI}| \geq 0.10$ as the significance criteria.

Multi-omics integration analysis of transcriptomic programs with metabolite-derived quality indices on phase-level

PCA was performed separately on $\log_2(\text{FPKM} + 1)$ -transformed transcriptomic data and on $\log_2(\text{peak area} + 1)$ -transformed metabolomic data, both standardized by z-scoring prior to PCA. For each accession $\times$ stage combination, stage centroids were calculated as the mean of biological replicate PC scores. Centroid trajectories were plotted in PC1–PC2 space for each accession across the seven developmental time points. Metabolomic indices (EI, SI, HSS) were accordingly re-aggregated at the transcriptomically defined phase. Co-expression network construction was performed using the WGCNA R package^[43] (v1.71). Of 22,566 expressed genes in the normalized count matrix, the 5,000 genes with the highest variance were retained for network construction, following established precedent for large multi-condition transcriptomic datasets^[43, 44]. A signed weighted co-expression network was constructed using a soft-thresholding power of $\beta = 9$, selected as the lowest value at which the scale-free topology fit index R^2 exceeded 0.85, as determined by the pickSoftThreshold function (Supplementary Figure S7). Modules

were identified using the `blockwiseModules` function, yielding 20 modules (Supplementary Figure S8); genes not meeting the co-expression coherence threshold were assigned to the grey module and excluded from downstream analyses. The 5,000 input genes comprising 3,450 assigned genes, with the remaining 1,550 genes (31.0%) designated as the grey module – a proportion consistent with values reported for multi-accession, multi-timepoint transcriptomic datasets of comparable complexity [44, 45]. Module eigengenes (MEs) were defined as the first principal component of the expression matrix for each module. Module eigengenes were correlated with EIz, SIz, and HSS using Spearman's rank correlation across 14 accessions × developmental-phase observations, comprising four phases each for GD and YN and three phases each for ZJ and GZ. Nominal *P* values were used for exploratory module screening, and Benjamini–Hochberg-adjusted values are reported in Supplementary Table S10; associations with nominal *P* < 0.05 were considered hypothesis-generating. ME values were averaged within each accession × developmental-phase group. Associations between MEs and the abundances of selected EI- or SI-associated metabolites were likewise assessed across the 84 biological samples using Spearman's rank correlation. An eigengene correlation network was constructed from pairwise Spearman correlations among MEs, with edges retained at $|p| \geq 0.35$ and nominal *P* < 0.05. Cross-accession concordance in module activity was assessed by calculating pairwise Spearman correlations between seven-point ME trajectories averaged across biological replicates for each accession × DAT combination.

Multi-omics integration and module–trait association analysis

Spearman's rank correlations were calculated between module eigengenes and nine traits using accession × DAT means (*N* = 28; four accessions × seven DAT points), with biological replicates averaged within each group. Nominal *P* values were used for exploratory screening, while Benjamini–Hochberg-adjusted values are provided in Supplementary Table S11; associations with nominal *P* < 0.05 were considered hypothesis-generating rather than confirmatory. Gene set enrichment analysis of four trait-associated modules (MEtan, MEgreen, MEsalmon, MELightgreen) was performed using Fisher's exact test against a background of all 22,566 expressed genes, with KEGG pathway and GO Biological Process annotations derived from the genome annotation file. A minimum overlap of two genes per term was required. Multiple testing correction was applied within each module using the Benjamini–Hochberg procedure. Terms with adjusted *p* < 0.05 were retained for visualisation. Intramodular hub genes were identified by ranking all genes within each trait-associated module by absolute kME (module membership), defined as the Pearson correlation between each gene's expression profile and the module eigengene across all 84 samples. Candidate hub genes were initially selected from the top 15 genes by absolute kME within each module and further prioritised based on functional annotation relevance to the associated phenotypic traits, as assessed by KEGG pathway assignment, GO Biological Process annotation, and BLAST-derived gene descriptions. Promoter sequences (2 kb upstream of the transcription start site, TSS) of seven candidate hub genes were extracted from the *D. officinale* genome assembly (ASM160598v2) using TSS coordinates from the GFF3 annotation file. For genes on the minus strand, the reverse complement of the extracted sequence was used. Eleven cis-regulatory motifs spanning hormone-responsive (GA-responsive, auxin-responsive, cytokinin-responsive, ABA-responsive),

TF-binding (MYB-binding, WRKY-binding, bHLH-binding, NAC-binding), and stress/development-associated categories were identified by scanning both strands using regular expression pattern matching.

Statistical analysis

Morphological traits were analysed by one-way ANOVA with Tukey's post-hoc test within each accession; phytohormone concentrations by two-way ANOVA (accession $\times$ DAT) with Tukey's comparisons at each DAT ($\alpha = 0.05$). Data are presented as mean $\pm$ SE unless otherwise stated. Per-metabolite variance partitioning used accession identity and developmental stage as predictors, with partial R^2 significance assessed by permutation (999 iterations). Sample quality was evaluated by Mahalanobis distance in PC1–PC2 space ($\chi^2_2 = 5.024$ threshold); samples exceeding this threshold but with total ion counts within the population mean ± 2 SD and QC CV $\leq 2.0\%$ were retained as biological outliers rather than analytical failures. EI, SI, and HSS were analyzed separately by two-way analysis of variance, with accession (GD, ZJ, YN, and GZ), days after transplanting (DAT; 0, 30, 60, 90, 150, 210, and 240), and their interaction included as fixed effects. Each accession $\times$ DAT combination comprised three biological replicates, giving 84 observations for each index. Model residuals were assessed for normality using the Shapiro–Wilk test, and homogeneity of variance among the 28 accessions $\times$ DAT groups was evaluated using the median-centered Brown–Forsythe test. No substantial departures from these assumptions were detected for any of the three indices. Where the accession $\times$ DAT interaction was significant, simple effects were examined using the residual variance from the full factorial model, followed by Tukey-adjusted pairwise comparisons. Partial η^2 was calculated as a measure of effect size. Statistical significance was set at $P < 0.05$. As a robustness check, the analyses were repeated using a Freedman–Lane permutation procedure with 9,999 permutations, which produced the same conclusions regarding statistical significance. OPLS-DA was performed for each accession against the remaining three using 114 EI-classified metabolites; VIP > 1 in at least one comparison defined significance. Differential gene expression was defined as $|\log_2$ fold change| > 1 and FDR < 0.05 (Benjamini–Hochberg correction; DESeq2 v1.38.0); differential alternative splicing events additionally required $|\Delta\text{PSI}| > 0.1$. KEGG pathway and GO Biological Process enrichment were assessed using clusterProfiler (v3.16.1) and topGO (v2.40.0), respectively, at $P < 0.05$. WGCNA module enrichment used Fisher's exact test against a background of 22,566 expressed genes, with Benjamini–Hochberg correction applied within each module (adjusted $P < 0.05$). Module eigengene associations with quality indices, individual metabolites, morphological traits, and phytohormones were assessed using Spearman's rank correlation. Nominal P values were used to identify exploratory associations. These correlation analyses were considered hypothesis-generating. Pearson correlation was used only for the calculation of kME between individual genes and their corresponding module eigengenes.

Declarations

Ethics approval and consent to participate

Not applicable. This study did not involve human participants, human data, or experimental animals.

Consent for publication

Not applicable.

Availability of data and materials

The raw sequencing and metabolomics data generated in this study have been deposited in the China National Center for Bioinformation (CNCB) under BioProject accession PRJCA067216, and are currently under curation. The data will be made publicly available upon publication. All other data supporting the findings are available within the article and its supplementary materials, or from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This work was supported by National Key Research and Development Program of China (2022YFD1600300)

Authors' contributions

CF: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **XC:** Methodology, Formal analysis, Writing – review & editing. **JL:** Investigation, Validation, Data curation. **YC:** Resources, Investigation. **EM:** Resources, Investigation. **DY:** Methodology. **ZZ:** Conceptualization, Supervision, Project administration, Funding acquisition, Writing – review & editing. All authors have read and approved the final manuscript.

References

1. Li PY, Li L, Wang YZ. Traditional uses, chemical compositions and pharmacological activities of *Dendrobium*: a review. *J Ethnopharmacol.* 2023;310:116382.
2. Chinese Pharmacopoeia Commission. Pharmacopoeia of the People's Republic of China. 2025 ed. Beijing: China Medical Science Press; 2025. p. 92, 282.
3. Yang L, Yang Y, Huang L, Cui X, Liu Y. From single- to multi-omics: future research trends in medicinal plants. *Brief Bioinform.* 2023;24(1):bbac485.
4. Hao DD, Huang LQ, Yuan Y, Jin Y. Application of numerical taxonomy in study of Dao-di herbs and their producing areas. *China J Chin Mater Med.* 2019;44(17):3633–6.
5. Xue ZZ. Study on Dao-di property of *Magnolia officinalis* based on the analysis of secondary metabolites [dissertation]. Beijing: China Academy of Chinese Medical Sciences; 2019.
6. Jangpangi D, Patni B, Chandola V, Chandra S. Medicinal plants in a changing climate: understanding the links between environmental stress and secondary metabolite synthesis. *Front Plant Sci.* 2025;16:1587337.

7. Venkatasai NN, Shetty DN, Vinay CM, Sekar M, Muthusamy A, Rai PS. A comprehensive review of factors affecting growth and secondary metabolites in hydroponically grown medicinal plants. *Planta*. 2025;261(3):48.
8. He S, Li M, Chen Y, Liu Q. Environmental and biological factors influence flavonoid and saponin contents in *Ziziphi Spinosae Semen* in the Taihang Mountains. *Ind Crops Prod*. 2025;231:121216.
9. Zhang C, Wang H, Lyu C, Wang Y, Sun J, Zhang Y, et al. Authenticating the geographic origins of *Atractylodes lancea* rhizome chemotypes in China through metabolite marker identification. *Front Plant Sci*. 2023;14:1237800.
10. Yu H, Liao J, Jiang Y, Zhong M, Tao S, Chai S, et al. Ecotype-specific phenolic acid accumulation and root softness in *Salvia miltiorrhiza* are driven by environmental and genetic factors. *Plant Biotechnol J*. 2025;23(6):2224–41.
11. Huang LQ, Guo LP, Hu J, Shao AJ. Molecular mechanism and genetic basis of geoherbalsim. *China J Chin Mater Med*. 2008;33(20):2303–8.
12. Jiang W, Wu ZG, Tao ZM. Herbalogical study on *Dendrobium officinale*. *J Chin Med Mater*. 2014;37(4):697–9.
13. Liu CB, Wen WE, Li ZG, Luo TT, Xu DL, Qian G, et al. Correlation analysis between endophytic fungal community structure and agronomic trait differences in *Dendrobium officinale* stems from different producing areas. *Chin Tradit Herb Drugs*. 2023;54(4):1236–42.
14. Zheng P, Wen M, Wang X, Liu QY, He X, Zuo YJ. Establishment of fingerprint and chemical pattern recognition of fresh *Dendrobium officinale* from different producing areas. *J Chin Med Mater*. 2024;47(12):3071–5.
15. Ding YL, Lin LY, Chen DQ, Xu H, Wang ZT. Content determination of six flavonoids in *Dendrobium officinale* stems from different producing areas, cultivation and processing methods by QAMS combined with dual-wavelength method. *China J Chin Mater Med*. 2021;46(14):3605–13.
16. Yuan LW. Tissue culture and cold resistance physiological research of three provenances of *Dendrobium officinale* [master's thesis]. Xinyang: Xinyang Normal University; 2012.
17. Liu C, Cao F, Diao Y, He Y, Cai S. Geographical origin identification of *Dendrobium officinale* using variational inference-enhanced deep learning. *Foods*. 2025;14(19):3361.
18. Kroymann J. Natural diversity and adaptation in plant secondary metabolism. *Curr Opin Plant Biol*. 2011;14(3):246–51.
19. Brinckmann JA. Geographical indications for medicinal plants: globalization, climate change, quality and market implications for geo-authentic botanicals. *World J Tradit Chin Med*. 2015;1(1):16–23.
20. Lan Q, Liu C, Wu Z, Ni C, Li J, Huang C, et al. Does the metabolome of wild-like *Dendrobium officinale* of different origins have regional differences? *Molecules*. 2022;27(20):7024.
21. Su JY, Yang SB, Wang YZ. Geographical origin identification of *Dendrobium officinale* based on FT-NIR and ATR-FTIR spectroscopy. *J Instrum Anal*. 2025;44(6):1005–15.
22. Li Y, Kong D, Fu Y, Sussman MR, Wu H. The effect of developmental and environmental factors on secondary metabolites in medicinal plants. *Plant Physiol Biochem*. 2020;148:80–9.

23. Managa LR, du Toit ES, Prinsloo G. Variations in the leaf metabolite profile between hydroponic and field-grown *Moringa oleifera* Lam. genotypes. *Biochem Syst Ecol.* 2021;97:104302.
24. Obul Z, Qiu YJ, Zhao YQ, Zhu J, Wang GP, Song WD, et al. Quality evaluation of *Arnebia euchroma* with different growth years based on metabolomics and antioxidant activity. *Front Plant Sci.* 2026;17:1735489.
25. Delporte C, Noret N, Vanhaverbeke C, Hardy OJ, Martin JF, Tremblay-Franco M, et al. Does the phytochemical diversity of wild plants like the *Erythrophleum* genus correlate with geographical origin? *Molecules.* 2021;26(6):1668.
26. Si C, Zeng D, Yu Z, Teixeira da Silva JA, Duan J, He C, et al. Transcriptomic and metabolomic analyses reveal the main metabolites in *Dendrobium officinale* leaves during the harvesting period. *Plant Physiol Biochem.* 2022;190:24–34.
27. Zhang SB, Zhou KJ, Zhang Z, Lu RR, Li X, Li XH. Suitable harvesting period of cultivated *Dendrobium officinale* in Yunnan. *China J Chin Mater Med.* 2015;40(18):3549–52.
28. Fiehn O. Metabolomics-the link between genotypes and phenotypes. *Plant Mol Biol.* 2002;48:155–71.
29. Xu Q, Zhu C, Fan Y, Song Z, Xing S, Liu W, et al. Population transcriptomics uncovers the regulation of gene expression variation in adaptation to changing environment. *Sci Rep.* 2016;6:25536.
30. Liu W, Chen S, Yuan X, Zhang Z, Duan K, Li L. Metabolome and transcriptome analyses provide insights into the podophyllotoxin content difference in different *Sinopodophyllum hexandrum* provenances. *Front Plant Sci.* 2026;16:1722118.
31. Zhou J, Liu R, Shuai M, Yan ZY, Chen X. Comparative transcriptome analyses of different *Salvia miltiorrhiza* varieties during the accumulation of tanshinones. *PeerJ.* 2021;9:e12300.
32. Liu S, Xing M, Yin X. MAPK regulates secondary metabolism and abiotic stress in horticultural and medicinal plants. *Hortic Res.* 2026;13:uhaf350.
33. Sibout R, Eudes A, Mouille G, Pollet B, Lapierre C, Jouanin L, et al. CINNAMYL ALCOHOL DEHYDROGENASE-C and -D are the primary genes involved in lignin biosynthesis in the floral stem of *Arabidopsis*. *Plant Cell.* 2005;17(7):2059–76.
34. Wu Y, Yan Q, Zou J, Chen H, Zhang Y, Xia H. Integrated transcriptomic, untargeted and targeted metabolomic analyses reveal seasonal regulatory mechanisms of vascular cambium activity in woody plants: insights from *Schima superba*. *Front Plant Sci.* 2026;16:1727826.
35. Carrera DA, George GM, Fischer-Stettler M, Galbier F, Eicke S, Truernit E, et al. Distinct plastid fructose biphosphate aldolases function in photosynthetic and non-photosynthetic metabolism in *Arabidopsis*. *J Exp Bot.* 2021;72(10):3739–55.
36. Gray WM, Kepinski S, Rouse D, Leyser O, Estelle M. Auxin regulates SCF^{TIR1}-dependent degradation of AUX/IAA proteins. *Nature.* 2001;414(6861):271–6.
37. Dreher K, Callis J. Ubiquitin, hormones and biotic stress in plants. *Ann Bot.* 2007;99(5):787–822.
38. de Roij M, Borst JW, Weijers D. Protein degradation in auxin response. *Plant Cell.* 2024;36(9):3025–35.

39. Schepetilnikov M, Ryabova LA. Auxin signaling in regulation of plant translation reinitiation. *Front Plant Sci.* 2017;8:1014.
40. Zhao D, Shi W, Xia X, Tang Y, Tao J. Microstructural and lignin characteristics in herbaceous peony cultivars with different stem strengths. *Sci Hortic.* 2019;253:1–8.
41. Want EJ, Masson P, Michopoulos F, Wilson ID, Theodoridis G, Plumb RS, et al. Global metabolic profiling of animal and human tissues via UPLC-MS. *Nat Protoc.* 2013;8(1):17–32.
42. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559.
43. Yang J, Ren Y, Zhang D, Chen X, Huang J, Xu Y, et al. Transcriptome-based WGCNA analysis reveals regulated metabolite fluxes between floral color and scent in *Narcissus tazetta* flower. *Int J Mol Sci.* 2021;22(15):8249.
44. Wang R, Shu P, Zhang C, et al. Integrative analyses of metabolome and genome-wide transcriptome reveal the regulatory network governing flavor formation in kiwifruit (*Actinidia chinensis*). *New Phytol.* 2022;233(1):373–89.
45. Umer MJ, Bin Safdar L, Gebremeskel H, et al. Identification of key gene networks controlling organic acid and sugar metabolism during watermelon fruit development by integrating metabolic phenotypes and gene expression profiles. *Hortic Res.* 2020;7:193.

Figures

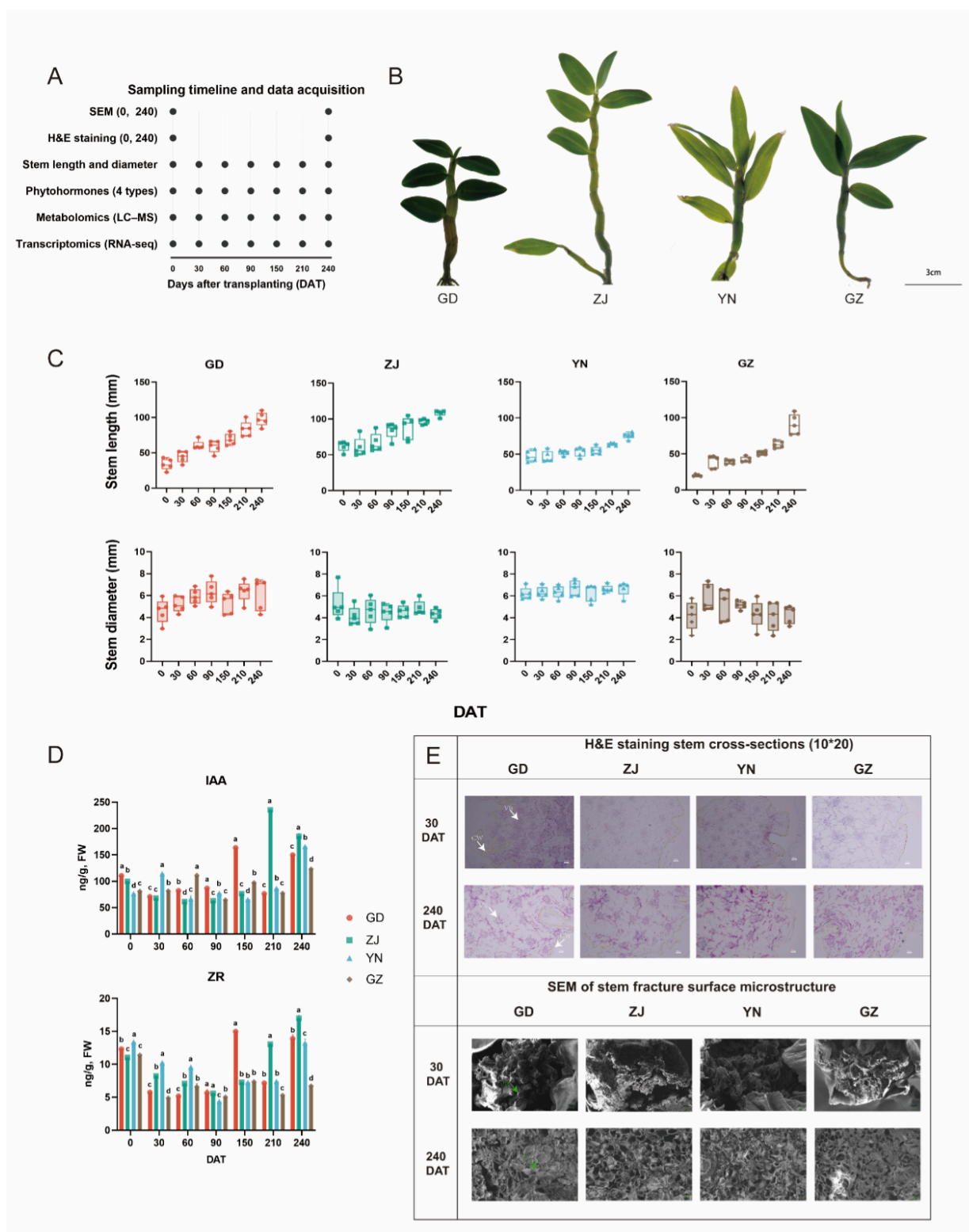


Figure 1

Sampling design, growth dynamics, endogenous hormone profiles, and stem anatomical characterization across four provenance-defined accessions.

(A) Sampling timeline and data acquisition across days after transplanting (DAT). (B) Representative morphological profiles of the four accessions at 150 DAT. (C) Stem length and stem diameter across sampling time points in each accession ($n \geq 6$ plants per accession per time point). Time-point effects were

evaluated within each accession by one-way ANOVA followed by Tukey's multiple-comparisons test ($\alpha = 0.05$). **(D)** Endogenous concentrations of indole-3-acetic acid (IAA) and zeatin riboside (ZR) in stems of four accessions ($n=3$). Data were analyzed by two-way ANOVA with accession and DAT as factors; simple main effects of accession were evaluated at each DAT by Tukey's multiple-comparisons test (adjusted $P < 0.05$). **(E)** Representative haematoxylin and eosin (H&E)-stained stem cross-sections and scanning electron microscopy (SEM) images of stem fracture surfaces at 30 and 240 DAT for each accession; scale bars as indicated. VB= vascular bundles; CW=cell wall.

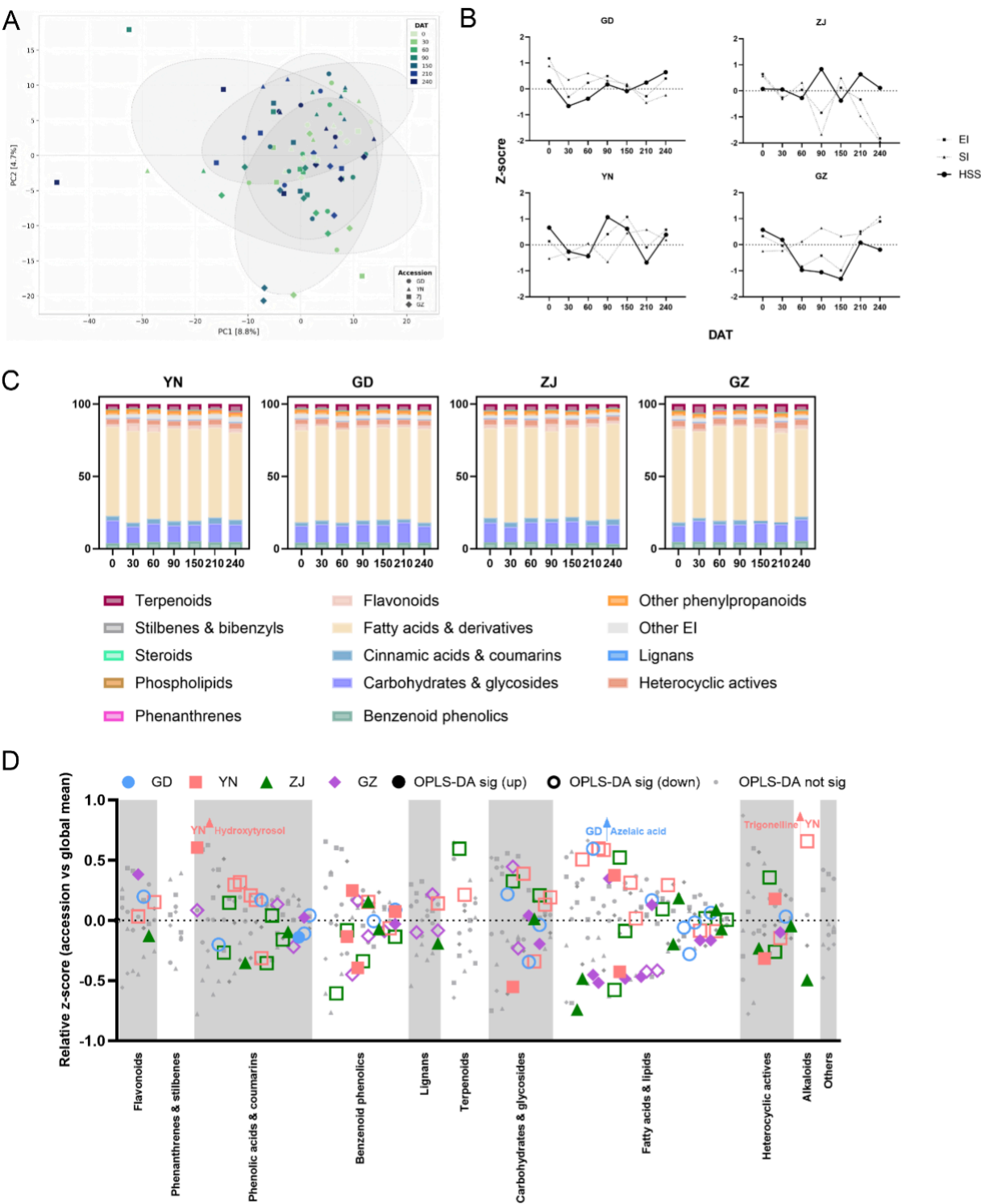


Figure 2

Metabolomic dynamics reveal accession-dependent quality formation across four provenance-defined accessions.

(A) PCA score plot of metabolomic profiles across all 84 biological samples **(B)** Temporal trajectories of EI_z, SI_z, and HSS within each accession across stages. Data are presented as mean (n = 3). A significant accession × DAT interaction was detected only for SI. Tukey-adjusted simple-effects comparisons showed that SI was lower in ZJ than in GD (P = 0.034) and GZ (P = 0.011) at 90 DAT, and lower in ZJ than in YN (P = 0.022) and GZ (P < 0.001) at 240 DAT. No significant accession, DAT, or interaction effects were detected for EI or HSS. **(C)** Superclass-level compositional breakdown of EI-contributing metabolites across DAT for each accession (stacked bars, y-axis). **(D)** Accession-specific EI metabolite enrichment profiles across chemical categories. Relative z-scores are shown for 114 EI-classified metabolites grouped by chemical category.

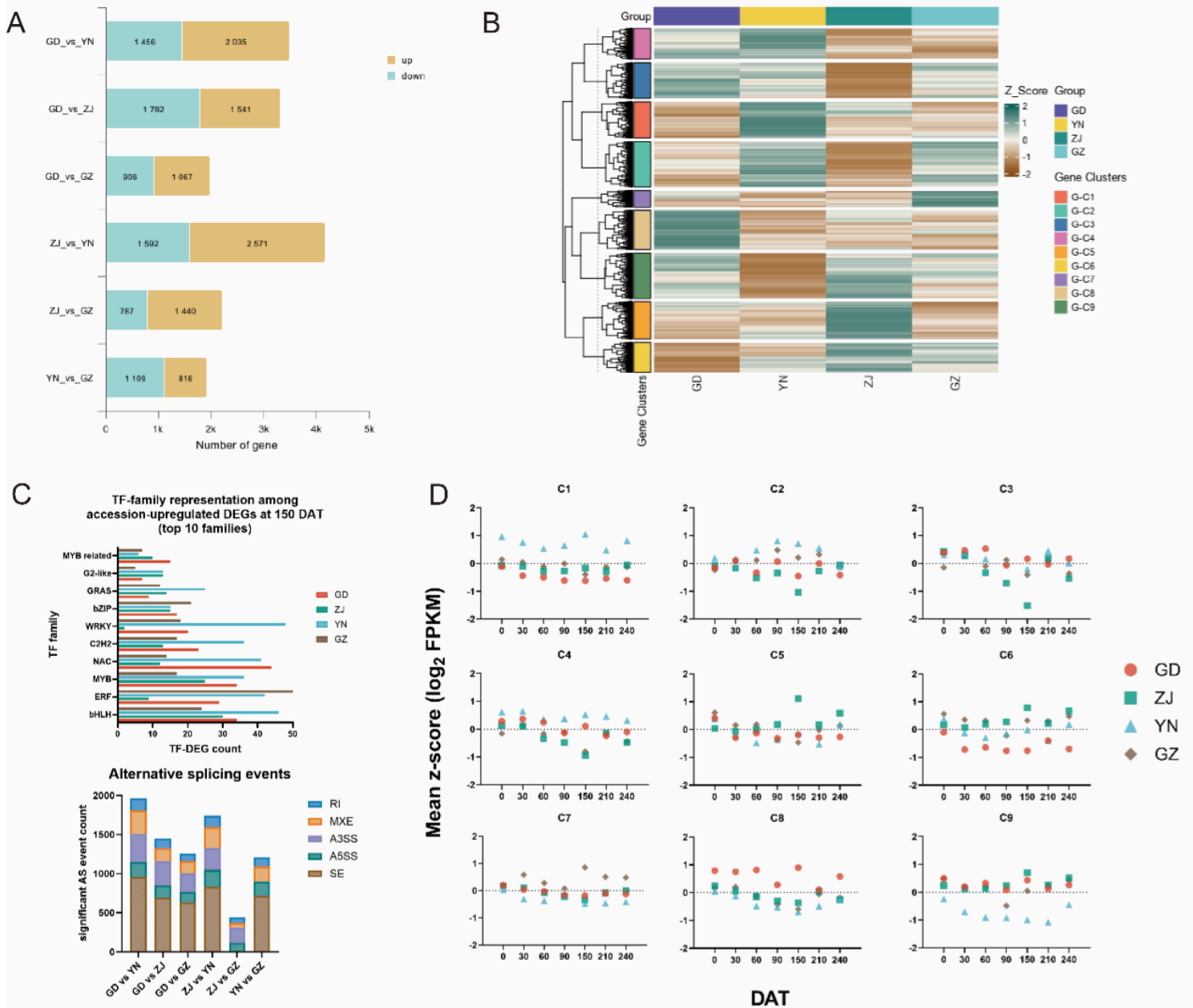


Figure 3

Divergence among provenance-defined accessions at 150 DAT reveals distinct transcriptional programs associated with geo-authenticity.

(A) DEG burden across pairwise comparisons among accessions at 150 DAT. Up- and down-regulated genes are defined according to the numerator and denominator specified for each pairwise contrast. **(B)** Heatmap of transcriptomic variation across the four provenance-defined accessions at 150 DAT. **(C)** Accession-associated transcriptional differences at 150 DAT at the levels of alternative splicing and TF-family representation. Lower panel: stacked bar chart showing the number of significant differential AS events (FDR < 0.05, $|\Delta\text{PSI}| > 0.1$) per pairwise comparison, classified by AS type. Upper panel: grouped bar chart showing cumulative accession-upregulated TF-DEG occurrences for the ten most represented TF

families across the six pairwise comparisons. **(D)**Developmental expression trajectories of 150 DAT DEGs across accessions (G-C1 to G-C9; n = 7,657 total).

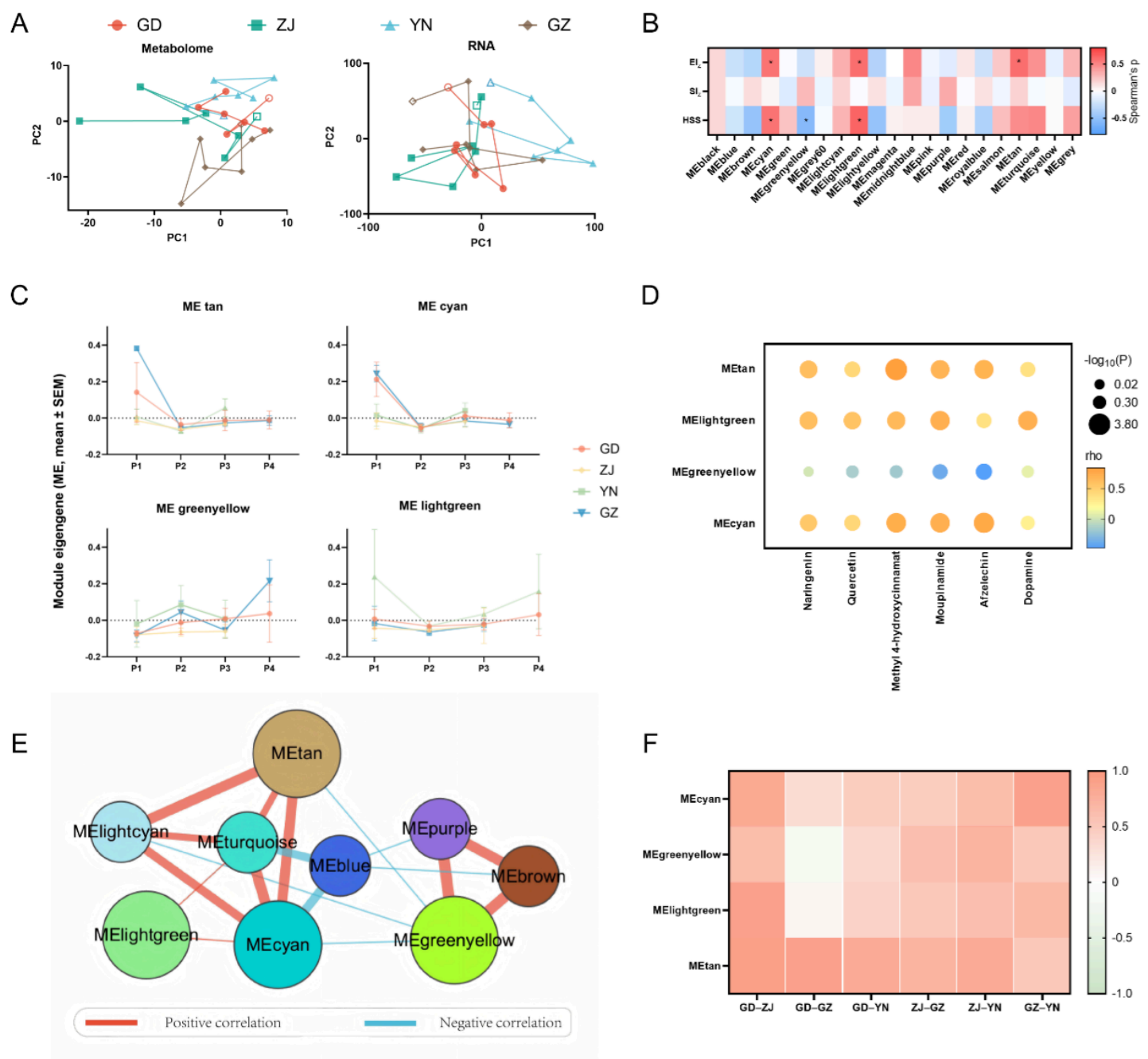
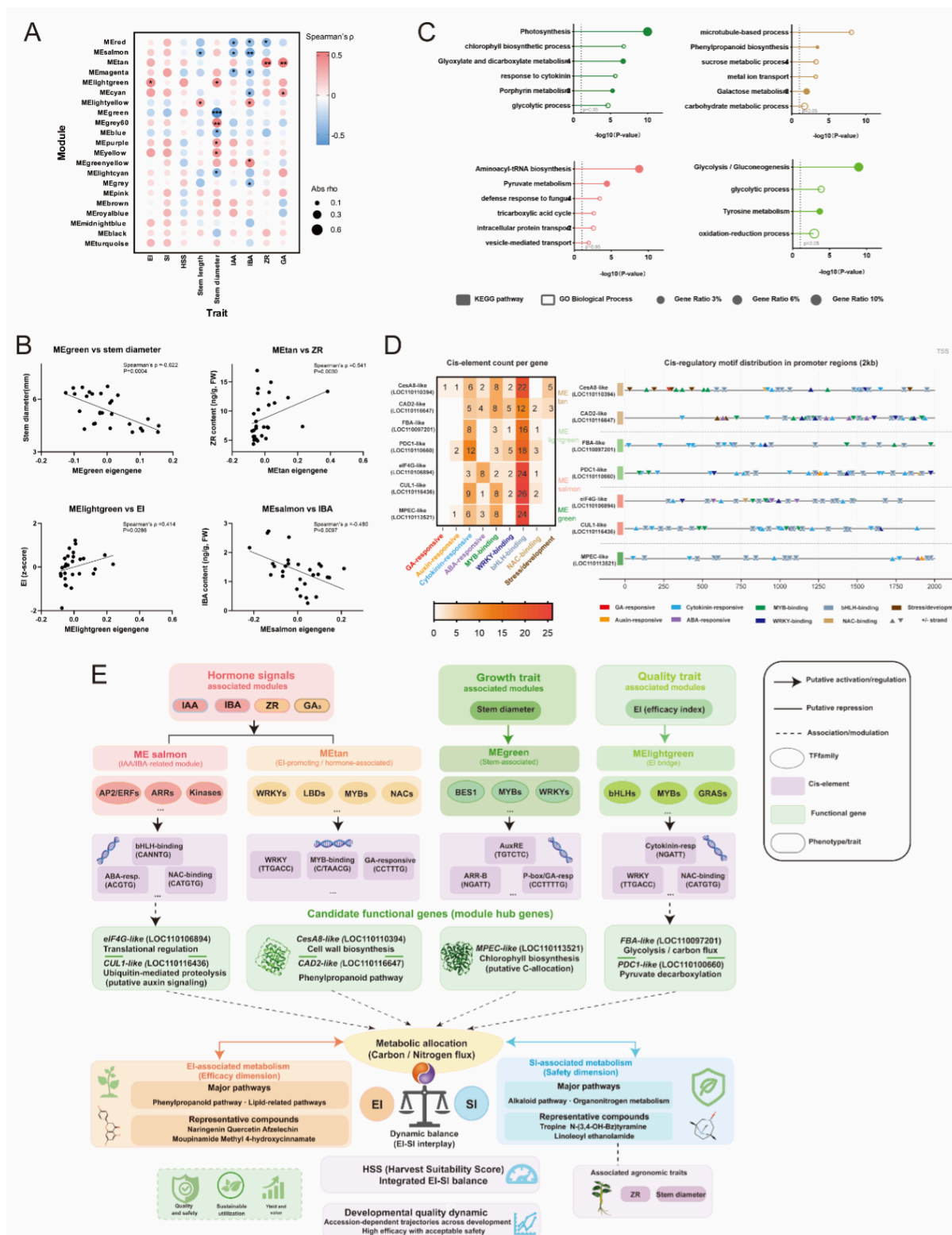


Figure 4

Multi-omics integration reveals the transcriptional basis of metabolite-defined quality variation through co-expression module analysis

(A) Phase-level co-variation between transcriptomic and metabolomic trajectories across accessions. PCA of phase centroids was performed on transcriptomic and metabolomic datasets separately. **(B)** Spearman correlations between module eigengenes and EIz, SIz, and HSS across 84 biological samples.

Asterisks indicate nominal significance ($P < 0.05$, Spearman test). **(C)** Phase trajectories of representative module eigengenes across developmental phases within each accession. Modules were selected based on their nominal associations with EI (MEtan, MElightgreen, and MEcyan) and HSS (MElightgreen, MEgreenyellow, and MEcyan). **(D)** Bubble plot of Spearman correlations between module eigengenes and six representative EI-associated compounds spanning multiple chemical classes. Bubble color indicates Spearman's ρ , and bubble size indicates $-\log_{10}(P)$. **(E)** Eigengene correlation network of selected modules. Nodes represent module eigengenes; edges represent pairwise Spearman correlations ($|\rho| \geq 0.35$, $P < 0.05$). Edge width is proportional to $|\rho|$. **(F)** Cross-accession concordance of module eigengene trajectories. Higher ρ values indicate greater similarity in the temporal activity pattern of the module between accessions.



eigengene–trait relationships scatter plots. Associations are shown for MEgreen vs. stem diameter, MEtan vs. ZR content, MELightgreen vs. EI, and MESalmon vs. IBA content ($n = 28$ accession $\times$ DAT-point means). Lines indicate fitted regression trends. **(C)** Functional enrichment of trait-associated modules. Significantly enriched KEGG pathways and GO Biological Process terms ($P < 0.05$, Fisher's test) are shown for MEtan, MEgreen, MESalmon, and MELightgreen. **(D)** Promoter cis-element analysis of candidate hub genes. Heatmap (left) shows counts of nine cis-regulatory motif categories in 2 kb upstream promoter regions of seven hub genes across four modules. Positional distribution of motifs is shown on the right; triangles indicate strand orientation ($\blacktriangle +$, $\blacktriangledown -$). **(E)** Working model of transcriptional regulation underlying quality formation in *D. officinale*. Four co-expression modules (MEtan, MEgreen, MELightgreen, MESalmon) are proposed to link hormonal and developmental signals to metabolic allocation between EI-associated (phenylpropanoid and lipid-related) and SI-associated (alkaloid and organonitrogen) chemical pools. The resulting EI–SI balance is captured by HSS and reflects accession-dependent developmental trajectories. Hub genes within each module represent candidate regulatory nodes. Dashed arrows indicate inferred rather than experimentally validated relationships.

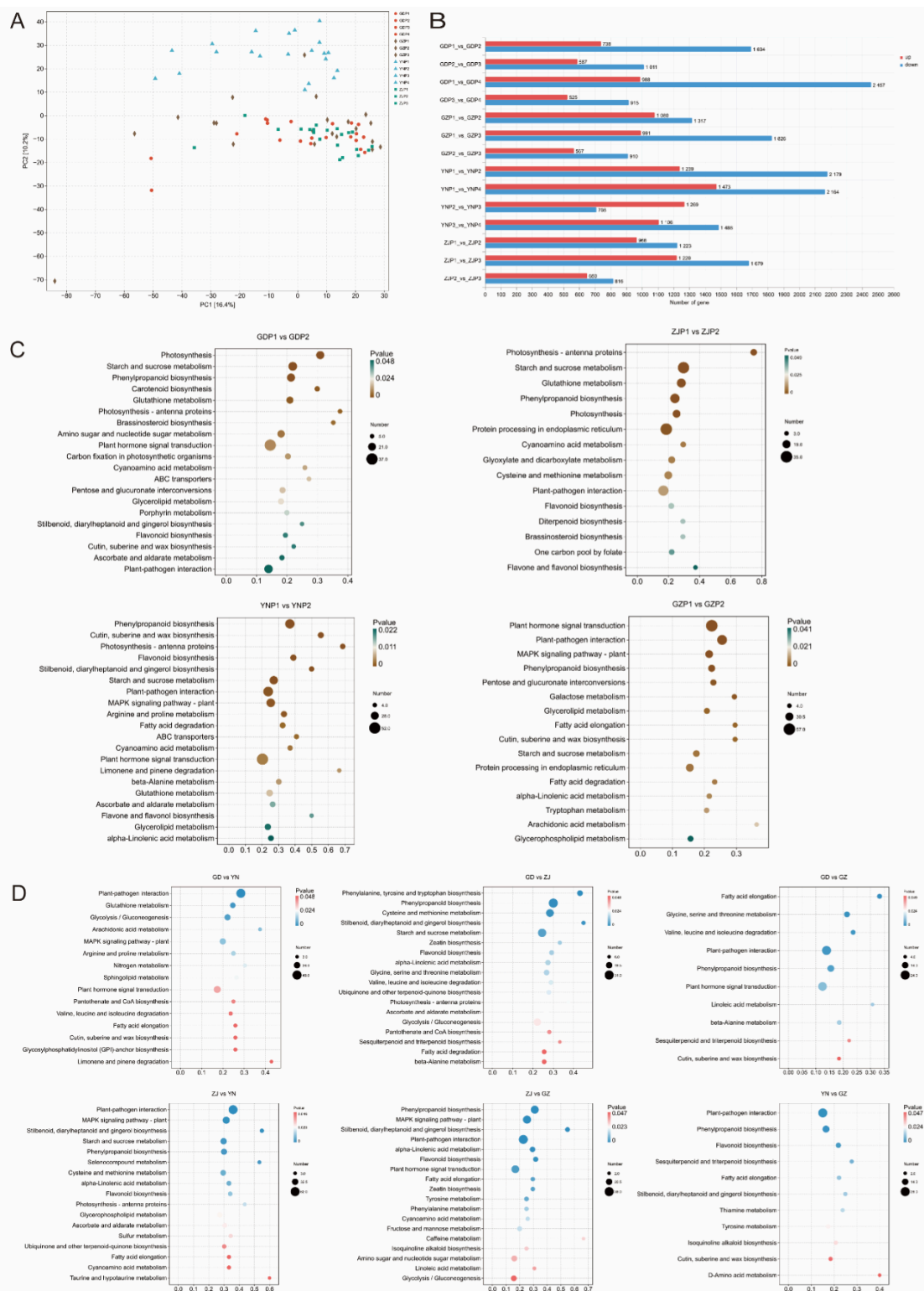


Figure 6

Time- and accession-dependent transcriptomic patterns reveal an early developmental switch and transcriptomic divergence point at 150 DAT.

(A) Principal component analysis (PCA) of the RNA-seq data across all accessions and developmental time points (The phase division information is shown in Supplementary Table S3). **(B)** Overview of differentially expressed genes (DEGs) associated with consecutive phase transitions within each

accession, together with P4 vs P1 for GD and YN; P3 vs P1 for ZJ and GZ. **(C)** KEGG enrichment analysis of DEGs identified in the early-phase transition (P2 vs P1) within each accession. The x-axis indicates the rich factor; bubble size represents the number of DEGs mapped to each pathway; bubble color indicates enrichment significance ($-\log_{10} P$). **(D)** Cross-accession KEGG enrichment analysis of DEGs at 150 DAT derived from six pairwise comparisons among GD, ZJ, YN, and GZ. Bubble size and color are defined as in **(C)**.

Supplementary Files

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

- [SupplementTable.xlsx](#)
- [supplementfigure.pdf](#)