



OPEN

Dynamic polyphenolic profiling of soybean seeds and leaves during developmental stages

Kihun Ha^{1,5}✉, Vimalraj Mani^{1,5}✉, Soyoung Park^{1,5}, Sukhee Kim¹, Sichul Lee¹,
Jae Kwang Kim², Jong Sung Lee², Jong-Yeol Lee³, Soo-Kwon Park⁴, Hye Rang Park⁴,
Jin A. Kim³ & Kong-Sik Shin¹

Soybean (*Glycine max* L.) is a globally important crop recognized for its high protein content and bioactive compounds, such as flavonoids and isoflavones, which contribute to its nutritional and health benefits. However, the metabolic profiles of soybean leaves at various growth stages remain underexplored. We investigated the polyphenolic (flavonoid and isoflavone) metabolite profiles in 10 genetically diverse soybean varieties across 6 leaf developmental stages and seeds. Sixteen flavonoids and isoflavones were identified through ultra-performance liquid chromatography-mass spectrometry. Data analysis, including principal component analysis, revealed metabolic differences among seed profiles. Partial least squares discriminant analysis (PLS-DA) explored metabolite changes across stages, while orthogonal PLS-DA separated vegetative and reproductive stages. Variable importance in projection scores highlighted naringenin, genistin, and kaempferol as key discriminative metabolites. Heatmap analysis showed flavonoids, particularly kaempferol glycosides, were abundant in the leaves, whereas isoflavones such as daidzin and genistin dominated seeds. Cultivars Hoseo, Sohwang, and Kwangan had the highest kaempferol glycoside levels. Correlating these metabolites with six key polyphenolic biosynthesis genes, chalcone synthase, chalcone isomerase, flavonol synthase, 2-hydroxyisoflavanone dehydratase, isoflavone 7-o-methyl transferase, and isoflavone synthase, revealed variety-specific regulation. This study provides key insights into developmental metabolite dynamics in soybeans and supports strategies to enhance soybean nutritional quality.

Keywords Soybean, Flavonoids, Isoflavones, UPLC-MS/MS

Soybean (*Glycine max* L.) is a globally important crop, valued for its high protein content and important role in both human nutrition and animal feed^{1,2}. In addition to its protein richness, soybeans are abundant in bioactive compounds like flavonoids and isoflavones, which are known for their health benefits, such as reducing the risk of cardiovascular diseases, osteoporosis, and certain cancers^{3–5}. As a versatile crop, soybeans are becoming increasingly important in agriculture, industry, and food production, with their global demand surging continuously⁶. With the projected global population increase, food production must significantly rise to meet future needs⁷. One of the most effective ways to address this challenge is through the development of soybean cultivars rich in valuable metabolites, such as flavonoids and isoflavones⁸.

Metabolite profiling is essential for uncovering the biochemical composition, metabolic pathways, and physiological states of soybeans and for providing key insights into their complex metabolic networks during growth and environmental responses⁹. This comprehensive metabolomic approach has enabled the identification and characterization of a wide array of primary and secondary metabolites, including amino acids, carbohydrates, lipids, and various phenolic compounds like flavonoids and isoflavones. Recent advances in analytical techniques, such as gas chromatography and high-performance liquid chromatography coupled with mass spectrometry (MS), have facilitated more detailed metabolomic studies¹⁰. While research has primarily focused

¹Plant Biomaterials and Biotechnology Division, Department of Agricultural Biology, National Institute of Agricultural Sciences, Rural Development Administration, Jeonju, Republic of Korea. ²Division of Life Sciences, College of Life Sciences and Bioengineering, Incheon National University, Incheon, Republic of Korea. ³Digital Breeding Convergence Division, Department of Agricultural Biotechnology, Rural Development Administration, National Institute of Agricultural Sciences, Jeonju, Republic of Korea. ⁴Department of Upland Crop Sciences, Upland Crop Breeding Division, National Institute of Crop and Food Sciences, Rural Development Administration, Miryang, Republic of Korea. ⁵Kihun Ha, Vimalraj Mani and Soyoung Park These authors have made equal contributions to this work. ✉email: khun329@korea.kr; vimalraj08@gmail.com

on soybean seeds due to their nutritional and economic importance¹¹, soybean leaves also contain valuable metabolites, particularly flavonoids like kaempferol glycosides^{12,13}. Flavonoids are a diverse group of secondary plant metabolites that play crucial roles in plant physiology and human health¹⁴. In plants, they are involved in pigmentation, UV protection, nitrogen fixation, and pathogen defense^{15,16}. In soybeans, isoflavones, a subgroup of flavonoids, are notable for their estrogen-like effects, which can influence hormone-related health conditions in humans¹⁷. Studies have demonstrated that soybean leaf extracts can benefit conditions such as diabetes and obesity, and flavonoid metabolites such as genistein, daidzein, and glycitein have been identified^{18,19}. Despite their significance, the metabolic profiles of soybean leaves at various growth stages remain underexplored²⁰. Understanding these changes is essential for enhancing plant biology knowledge and optimizing agricultural practices, which could reduce reliance on synthetic inputs and promote sustainable farming methods^{21,22}.

Recent research has employed principal component analysis (PCA) to monitor metabolite changes across different growth stages, identifying key flavonoids like kaempferol glycosides and genistein as important markers¹⁵. Additionally, 83 flavonoid derivatives have been identified in 4-week-old soybean leaves²³. While Park et al. investigated metabolite changes across five soybean cultivars and identified both primary and secondary metabolites²⁰, our study expands on this research by conducting a comprehensive analysis of flavonoids and isoflavones from ten soybean cultivars across six leaf developmental stages to determine optimal leaf harvest times, as well as in their seeds. The vegetative (V) stages studied included nitrogen fixation initiation (V2) by rhizobia and the development of four trifoliate leaves (V4). The reproductive (R) stages encompassed flower formation (R1) through seed maturity (R7). Specifically, R1 marks the beginning of flowering, R3 signifies the onset of pod development, R5 indicates seed development, and R7 represents seed maturity²⁴.

Thus, this research provides a comprehensive understanding of metabolite dynamics throughout the soybean life cycle, particularly regarding flavonoid and isoflavone composition in leaves and seeds. These insights have significant implications for optimizing soybean cultivation, enhancing its nutritional value, and advancing agricultural biotechnology. The findings lay a foundation for future crop improvement and the development of functional foods.

Results and discussion

Phenotypic characteristics of mature seeds and developmental stages of leaves

This study utilized 10 soybean landrace varieties Meju (MJ), Bancheongdu (BCD), KLS88035 (KLS), SLSB397-1 (SL), and Kangwon Sujib2-33(KW), and improved varieties Hoseo (HS), Suwon98 (SW), Sohwang (SH), Bosug (BS), and Kwangan (KA) at various leaf and seed developmental stages, as shown in Fig. 1. Regarding the phenotypic seed color in the landrace lines, MJ seeds were yellow, KLS were green, and BCD, SL, and KW exhibited a brown seed color. In the improved lines, HS, SW, SH, BS, and KA, the seeds were yellow. As shown in Fig. 1b, the vegetative (V2 and V4) and reproductive (R1, R3, R5, and R7) stages can be visually distinguished in the soybean plants.

Metabolites of soybean seeds and leaves

In this study, a total of 16 metabolites, including 11 flavonoids and 5 isoflavones, were identified in 10 soybean seed samples and leaves at different developmental stages through ultra-performance liquid chromatography-mass spectrometry (UPLC-MS)/MS. The identified flavonoids included apigenin, kaempferol, luteolin, naringenin, quercetin, rutin, and vitexin and four kaempferol glycosides (KFG1, KFG2, KFG3, and KFG4). However, the identified isoflavones were daidzein, daidzin, genistein, and genistin. Peak identification was supported by the Target Lynx XS software and an in-house mass spectra library. Detailed information on compound names, retention times, wavelengths, and precursor/product ions is provided in Supplementary Table S1. Compounds with available authentic standards were quantified by absolute quantification using calibration curves (Supplementary Table S2). Supplementary Figure S1 shows chromatograms of both samples and standard mixtures, validating the accuracy and reproducibility of the quantification. For the kaempferol group, quantification was carried out using UV detection at 350 nm, with peak areas normalized to an internal standard to ensure accuracy. All values are expressed as mean \pm SD from three biological replicates. A comprehensive dataset of all identified compounds across 10 germplasm lines, including six developmental stages of leaf samples and seed samples, is provided in Supplementary Table S3.

Comparative analysis of metabolite profiles in seeds of ten soybean cultivar varieties

Of the 16 identified metabolites, seven were detected in the seeds—daidzein, daidzin, genistein, genistin, naringenin, kaempferol, and quercetin. Notably, genistin and daidzin levels were significantly higher in seeds than in leaves (Supplementary Table S3). This finding aligns with those of previous studies that have highlighted the distinct metabolite profiles of soybean seeds and leaves^{20,23}. PCA was performed to evaluate the metabolite profiles of the seeds (Fig. 2A, 2B). The score plot revealed considerable variability in the metabolite profiles across the 10 soybean cultivars, as indicated by the spread of samples along the first two principal components. The first principal component (PC1) accounted for 65% of the variance and the second (PC2) explained 22.6%, collectively capturing 87.6% of the total variance. This high level of explained variance is consistent with other recent metabolomic analyses of soybean seeds¹¹. Distinct clustering patterns were observed, with SL, KW, BCD, KA, MJ, and SW exhibiting tight clusters, indicating homogeneity in their metabolite profiles. In contrast, KLS, HS, and BS displayed relatively greater dispersion, suggesting higher variability in their metabolite profiles. These varietal differences in the metabolite profiles corroborate the findings of Park et al.²⁰, who reported significant metabolite variations among soybean cultivars. The observed differences in seed metabolite profiles underscore the importance of cultivar selection in breeding programs aimed at enhancing specific metabolite content, as highlighted by recent studies^{4,5}. Furthermore, the distinct metabolite profiles of seeds compared with those of

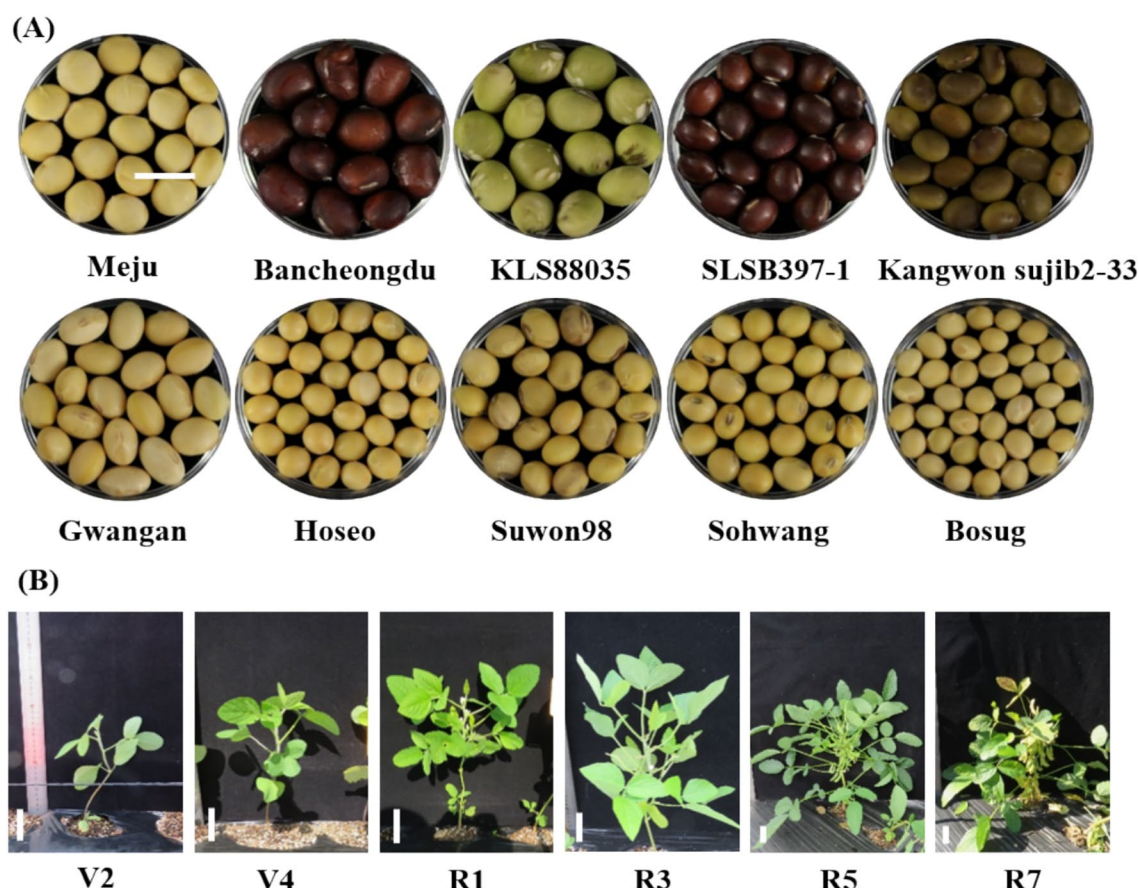


Fig. 1. Phenotypic characteristics of seeds of 10 soybean core collection cultivars (A) and leaf developmental stages (V2, V4, R1, R3, R5, and R7) of variety Kwangan (B). (A. bar = 1 cm; B. bar = 5 cm).

leaves emphasize the tissue-specific nature of metabolite accumulation in soybeans, a phenomenon that has also been noted in recent comprehensive metabolomic analyses³.

Variation in metabolite profiles across growth stages

PCA of metabolite profiles during the vegetative growth stages did not reveal any clustering based on cultivar, suggesting minimal varietal differences during this phase. However, distinct variations between the V2 and V4 stages were evident, indicating unique metabolomic shifts as the plants progressed through the vegetative phase (Supplementary Fig. 2). These results align with those of Park et al.²⁰, who also observed substantial fluctuations in metabolite composition within soybean leaves as they progressed through various developmental stages. Notably, varieties such as KA, SH, and HS consistently displayed high concentrations of kaempferol glycosides, suggesting the presence of variety-specific metabolic traits (Fig. 3). In the reproductive stages, partial least squares discriminant analysis (PLS-DA) revealed clear distinctions in metabolite profiles across different developmental stages of soybean plants, with the first two principal components explaining 40.8% of the total variation (Supplementary Fig. 2). This method successfully categorized the reproductive stages into three distinct clusters, consistent with the stage-specific metabolite changes observed by Song et al.¹⁵.

Further PLS-DA analysis across the entire growth period showed that the V4 stage was markedly distinct from other stages, indicating a unique metabolite profile (Fig. 2C, 2D). This is in line with the findings of Lee et al.²³, who reported significant flavonoid variations in young soybean leaves. Key metabolites contributing to the differentiation of growth stages included naringenin, genistin, and daidzin, with rutin and quercetin also playing crucial roles (Fig. 2D).

Notably, our findings on isoflavone content differed from those reported by Song et al.¹⁵. While they observed a rise in daidzein and genistein synthesis during the R5–R6 stages, our study found that these isoflavones were the most abundant during the vegetative stages and decreased by the R5 stage. This discrepancy highlights the complex nature of metabolite dynamics in soybeans, which can be influenced by factors such as cultivar type and environmental conditions. The Variable importance in projection (VIP) scores further underscored the significance of certain metabolites across different growth stages, with naringenin emerging as the most critical metabolite (Fig. 2G). The use of VIP scores to identify key metabolites is consistent with the methods employed by Kim et al.²⁵ in their metabolomic analysis of soybean cultivars. These findings enhance our understanding of the dynamic metabolite changes in soybean leaves during growth stages, complementing previous research on cultivar-specific metabolite variations and the relationship between growth stages and flavonoid composition.

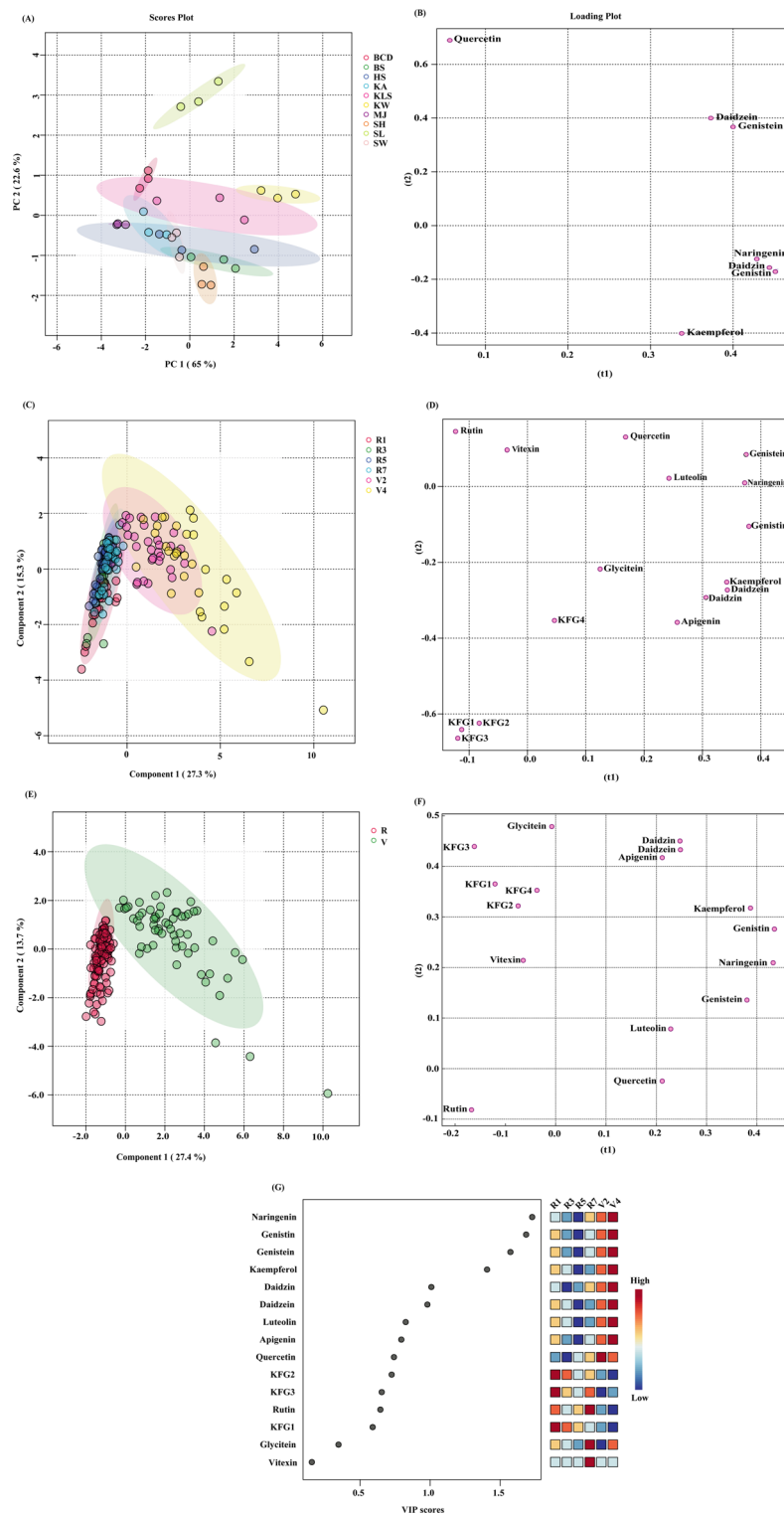


Fig. 2. Principal component analysis (PCA) score plot (A) and loading plot (B) of soybean seed in 10 cultivars. Partial least squares discriminant analysis score plot (C) and loading plot (D) of soybean leaf according to six growth stages, orthogonal partial least squares discriminant analysis score plot (E) and loading plot (F) of soybean leaf between vegetative stages and reproductive stages and variable importance in projection (VIP) scores associated with metabolites (G).

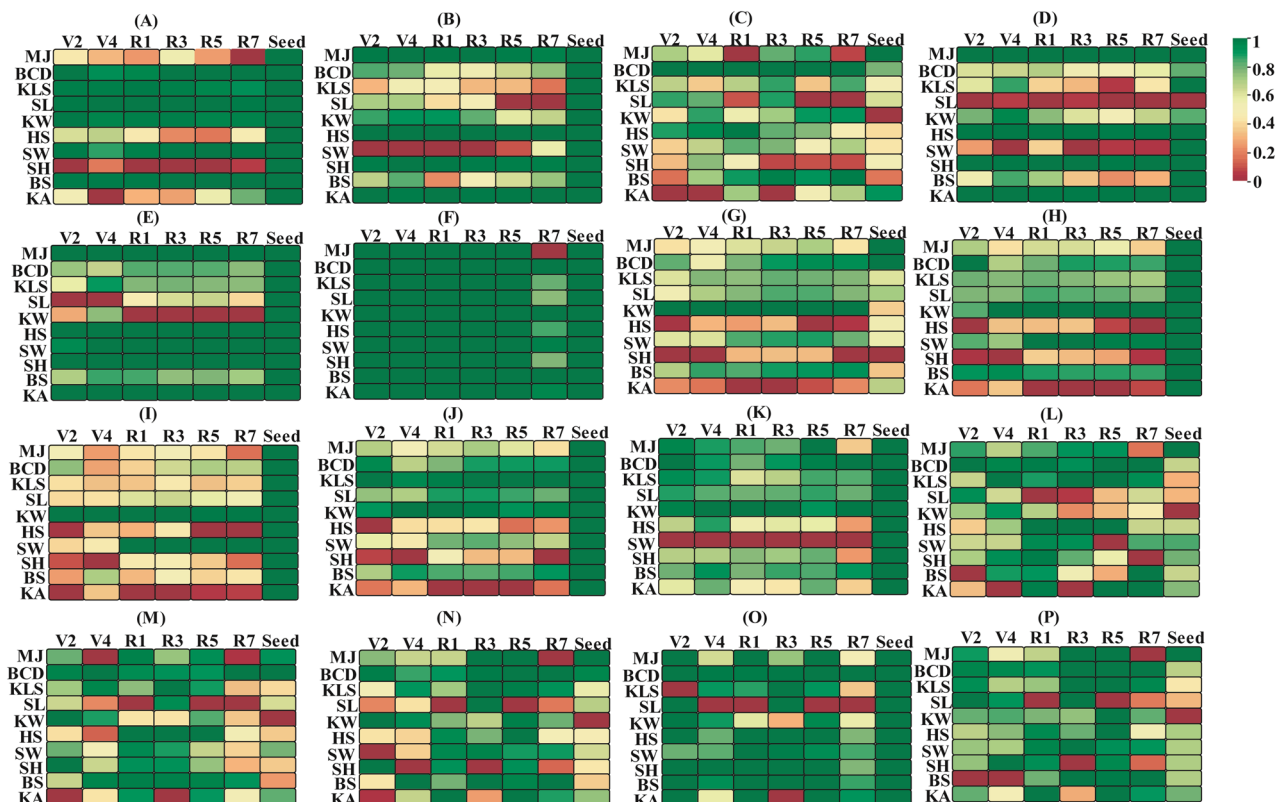


Fig. 3. Heat maps comparing the levels of flavonoids (A–K) and isoflavones (L–P) in vegetative and reproductive stages of 10 soybean cultivars. Peak areas were normalized to construct a comparative heat map using minimum and maximum algorithms. The deeper the red color, the higher the peak area of the metabolites; the deeper the green color, the lower the peak area of the metabolites.

OPLS-DA analysis of metabolite profile differences between vegetative and reproductive growth stages

Orthogonal partial least squares discriminant analysis (OPLS-DA) was employed to distinguish the metabolite profiles between the V and R growth stages of the soybeans (Fig. 2E, 2F). The predictive component accounted for 21.8% of the variation, while the orthogonal component (to) accounted for 17.3%, collectively representing 39.1% of the total variance. This analysis revealed a clear separation between the V and R stages, which is consistent with previous studies showing distinct metabolic changes during soybean development²⁰. The V stage was characterized by higher concentrations of flavonoids such as daidzin, genistin, apigenin, kaempferol, genistein, naringenin, luteolin, and quercetin. These findings align with those of Lee et al.²³, who identified 83 flavonoid derivatives in young soybean leaves. In contrast, rutin was more prominent during the R stage, particularly during the reproductive phase, supporting the observations of Song et al.¹⁵ of stage-specific flavonoid accumulation. These findings extend our understanding of the metabolite dynamics in soybean leaves across growth stages, complementing previous research on cultivar-specific metabolite variations²⁰ and the relationship between seed coat color and flavonoid composition²⁶.

Heat map analysis

Our heat map analysis (Fig. 3) based on the peak areas of metabolites revealed distinct patterns of flavonoid accumulation across both seed and leaf developmental stages in soybean. Notably, four KFGs were detected alongside apigenin, glycitein, luteolin, and vitexin in the leaf stages but were absent in the seeds. This pattern is consistent with the findings of Park et al.²⁰, who also reported the stage-specific accumulation of flavonoids in soybean leaves. In contrast, isoflavones like daidzein, daidzin, genistein, and genistin were more abundant in seeds during the V stages but showed reduced levels during the R stages. This trend aligns with the observations of Kim et al.²⁵ regarding isoflavone variation across soybean developmental stages. Kaempferol and quercetin exhibited greater abundance in the V stages compared to their levels in the R stages, while rutin concentrations were substantially higher during the R stages. This differential accumulation pattern supports the findings of Chen et al.²⁷, who reported stage-specific shifts in metabolite changes in soybean, particularly noting shifts in flavonoid profiles between the vegetative and reproductive phases. These results emphasize the dynamic nature of secondary metabolite synthesis in soybeans, as highlighted by recent metabolomic studies.

Variations in metabolite content across soybean varieties: a focus on Kwangan soybeans

Our metabolomic analysis of 10 soybean cultivars revealed distinct patterns of flavonoid content across different growth stages. Kwangan soybeans, a common variety in Korea, consistently showed higher levels of kaempferol throughout all growth stages than the other varieties studied. This result aligns with that of previous reports highlighting the variability in flavonoid content among soybean cultivars²³. Kwangan is widely cultivated in Korea and recognized for its commercial importance and versatility. It is particularly valued for its adaptability to Korean growing conditions and its high yield potential, making it one of the most commonly produced soybean cultivars in the country. Rich in protein and oil, it contributes to various food products, including traditional Korean soy-based foods such as tofu, soy milk, and fermented soybean products.

Notably, for certain metabolites, particularly luteolin, quercetin, and rutin, the other cultivars exhibited higher concentrations than KA. Varieties such as BCD, KLS, SL, KW, SW, and BS contained significant amounts of luteolin and quercetin, whereas MJ, HS, and SH varieties, similar to KA, had low concentrations (Fig. 4). This variation in flavonoid profiles among cultivars has been reported previously²⁰. Among the varieties with high levels of these metabolites, only SW and BS had yellow seed coats, whereas the others exhibited brown and green seed coats. This observation supports the findings of Kim et al.²⁵, suggesting that seed coat color may influence flavonoid composition.

Rutin was identified as the most prominent metabolite during the R stage, with significantly higher concentrations than during the V stage. This stage-specific accumulation of flavonoids has been reported previously¹⁵. The cultivars BCD, KLS, SL, KW, and BS contained substantial amounts of rutin, with KW exhibiting the highest concentration. Our findings support and extend previous research indicating that soybeans with non-yellow seed coats, including black, green, brown, and white, tend to have higher luteolin, quercetin, and rutin content than yellow-seeded varieties²⁶. Notably, despite its yellow seed coat, BS contained high levels of these compounds, demonstrating that this correlation is not absolute.

Despite including both improved and landrace lines in our study, we did not observe significant differences in metabolite levels across growth stages between these types. Additionally, when examining the metabolite content differences among the five soybean clades identified by Kim et al.²⁵, no notable variations were detected. These results suggest that factors beyond genetic lineage or improvement status may play a crucial role in determining the flavonoid profiles in soybeans.

Kwangan's significance extends beyond its agronomic traits. It has been utilized as a model variety in genetic modification studies aimed at enhancing nutritional characteristics, such as the development of β -carotene-enhanced soybeans²⁸. Its strong agronomic performance under local conditions has made it a popular choice among farmers and an important subject in breeding programs focused on improving disease resistance and other desirable traits. Additionally, KA prominence in research is underscored by its frequent use in studies investigating metabolite profiles and genetic improvements²⁹. Overall, KA plays a crucial role in both the agricultural landscape and food industry in Korea, and our study provides new insights into its flavonoid profile compared to other varieties.

Gene expression analysis of key polyphenolic biosynthetic genes

In addition to metabolomic profiling, we performed gene expression analysis of six key polyphenolic biosynthetic genes—*chalcone synthase (CHS1)*, *chalcone isomerase (CHI)*, *flavonol synthase (FNS)*, *2-hydroxyisoflavanone dehydratase (HID)*, *isoflavone O-methyltransferase (ISMAT)*, and *isoflavone synthase (IFS1)*—which are central components of the flavonoid and isoflavonoid biosynthetic pathways in *G. max* (Fig. 5). These genes were analyzed through qRT-PCR across different developmental stages and in seeds. *CHS* and *CHI* are responsible for the initiation of flavonoid skeleton formation, *FNS* branches the intermediates toward flavone derivatives, *HID* and *IFS1* drive isoflavone backbone formation, and *ISMAT* mediates late-stage methylation of isoflavones. These genes represent critical regulatory nodes that govern flux through the flavonoid/isoflavonoid metabolic network, and their expression patterns likely contribute to the observed variation in metabolite accumulation.

Integrating metabolomic data with gene expression data from qPCR provides a comprehensive understanding of the biological processes in soybeans, revealing the connections between metabolic pathways and gene regulation²⁷. Studies have shown correlations between metabolite levels and the expression of key genes in specific pathways such as flavonoid biosynthesis and isoflavone production³⁰. These correlations can be developmental stage-specific³¹ or vary in response to environmental factors³². Comparative analyses of different soybean cultivars have revealed how metabolite–gene expression correlations may differ, providing insights into breeding programs³³. Recent research has demonstrated the correlations between primary metabolite content and seed protein and oil traits in near-isogenic soybean lines, with 28 annotated metabolites showing significant correlations with seed traits²⁶. Additionally, integrated transcriptomic and metabolomic analyses have been used to investigate the regulatory networks controlling soybean seed metabolism³⁴ and to classify soybeans from different habitats³⁵.

We examined the *CHS* gene, a type III polyketide synthase that catalyzes the first step in flavonoid biosynthesis. In *G. max*, 14 unique *CHS* genes (*GmCHS1*–*GmCHS14*) have been identified³⁶. From these genes, we selected *CHS1* for correlation with our metabolite data, and observed its expression at the R7 leaf stage. Isoflavone-related genes, including *IFS*, *HID*, and *ISMAT*, were highly expressed in the seeds and R7 and V4 leaves, which corresponded with high levels of daidzin, genistin, and glycitin in our metabolite data. Additionally, *FNS*, which is involved in flavonol biosynthesis, showed high expression at the R1, V4, and R3 stages, particularly in the leaves, which aligned with the high flavonoid levels observed in our metabolomic data (Fig. 5). Finally, our gene expression results were compared to previously published transcriptomic data³⁷, which are consistent with our findings, further validating the expression patterns observed in our study.

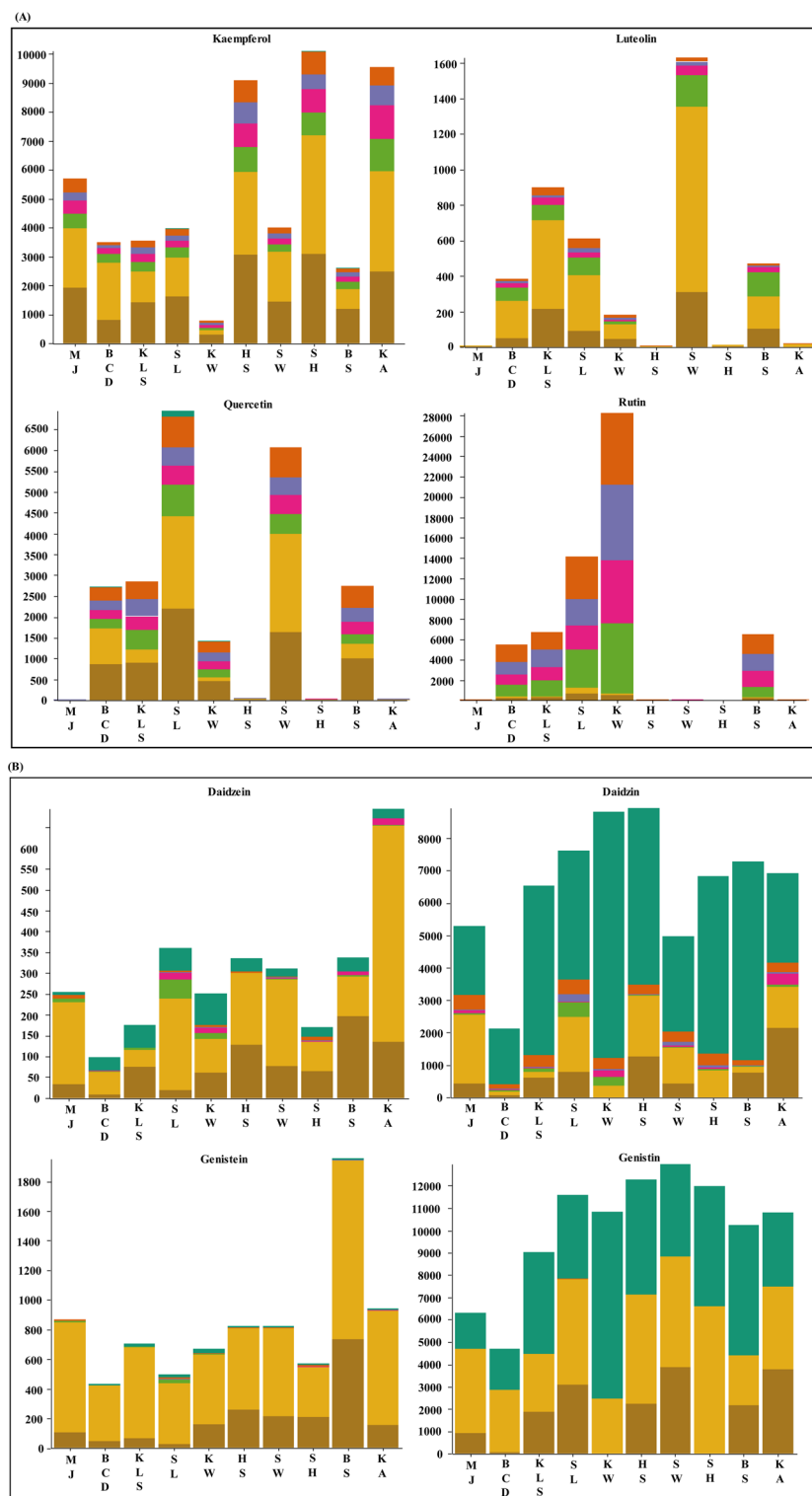


Fig. 4. Quantitative comparison of flavonoids (A, kaempferol, luteolin, quercetin, and rutin) and isoflavones (B, daidzein, daidzin, genistein, and genistin) across soybean cultivars and developmental stages. Flavonoids were predominantly detected in leaves, with minimal or no presence in seeds. Isoflavone glycosides (daidzin and genistin) showed peak concentrations in seeds, while aglycones (daidzein and genistein) were highest at the V4 stage. Data represent mean concentrations ($\mu\text{g/g}$ dry weight).

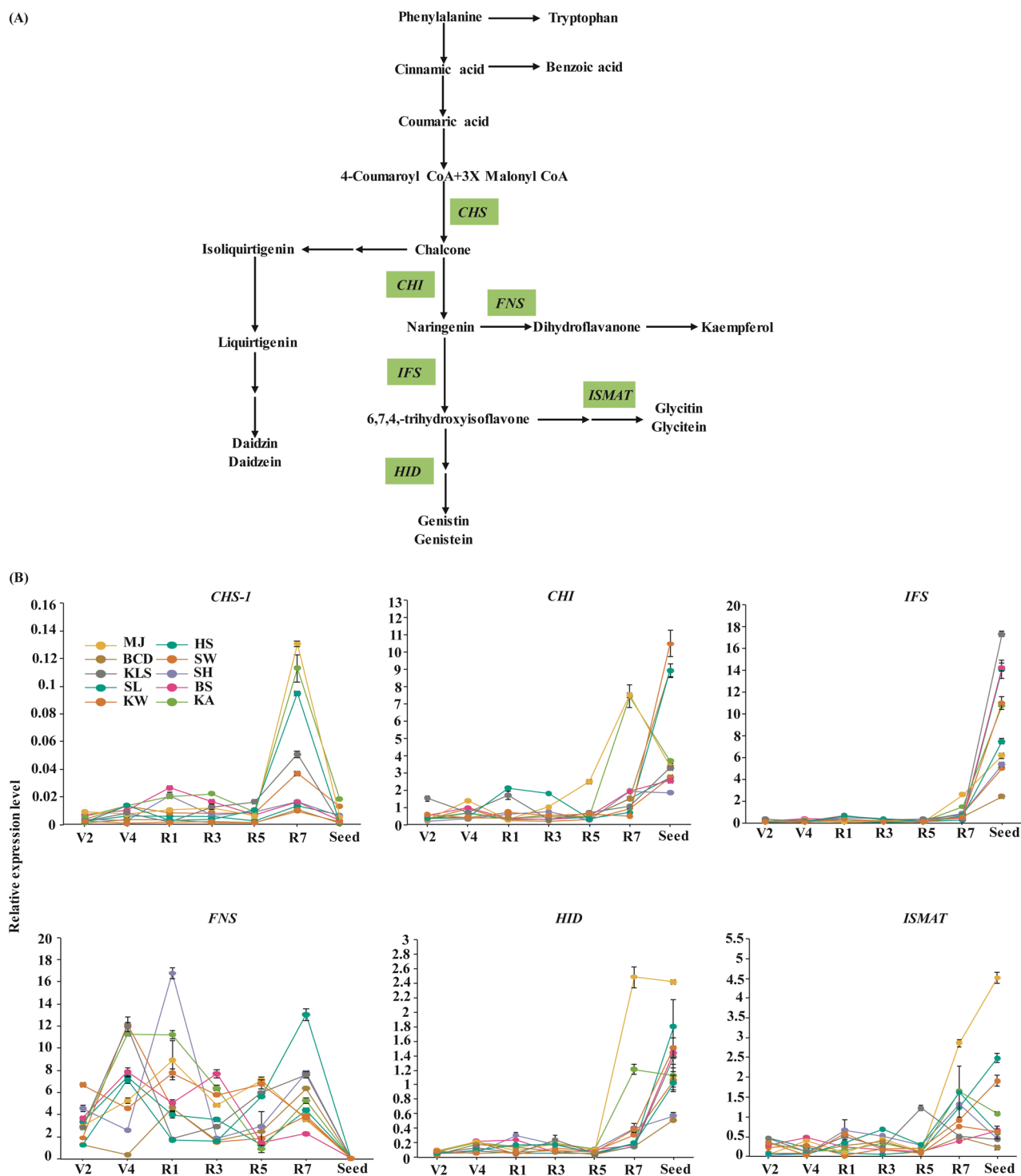


Fig. 5. Simplified polyphenolic metabolic pathway (A) and gene expression levels of biosynthesis gene for flavonoids and isoflavones (B). *CHS*, chalcone synthase; *CHI*, chalcone isomerase; *FNS*, flavonol synthase; *IFS*, isoflavone synthase; *ISMAT*, isoflavone 7-*o*-methyl transferase; *HID*, 2-hydroxyisoflavonone dehydratase. Three biological replicates were performed, and error bars represent standard deviation (SD).

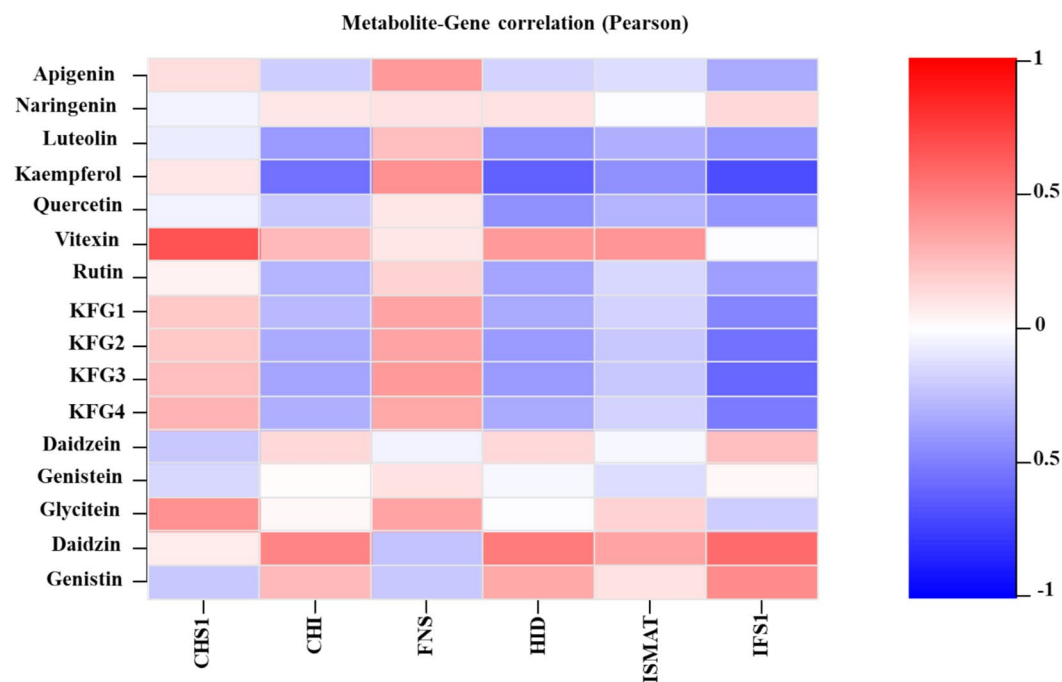
Correlation analysis integrating metabolite and gene expression data

To elucidate the regulatory relationships between metabolite accumulation and gene expression, we performed Pearson correlation analysis between 16 flavonoid/isoflavonoid metabolites and six key polyphenolic biosynthetic genes (*CHS1*, *CHI*, *FNS*, *HID*, *ISMAT*, and *IFS1*). The resulting correlation matrix revealed distinct positive and negative association patterns across developmental stages and varieties. Several metabolites, such as kaempferol-

group compounds and daidzein derivatives, showed strong positive correlations with *FNS* and *IFS1* expression, while others, including early flavanones, displayed weak or inverse correlations.

We visualized these associations as a heatmap (Fig. 6A), highlighting the correlation coefficients (r) and their statistical significance following Benjamini–Hochberg false discovery rate (FDR) correction. To further examine the most robust associations, we filtered significant metabolite–gene pairs (FDR < 0.05) and constructed

(A)



(B)

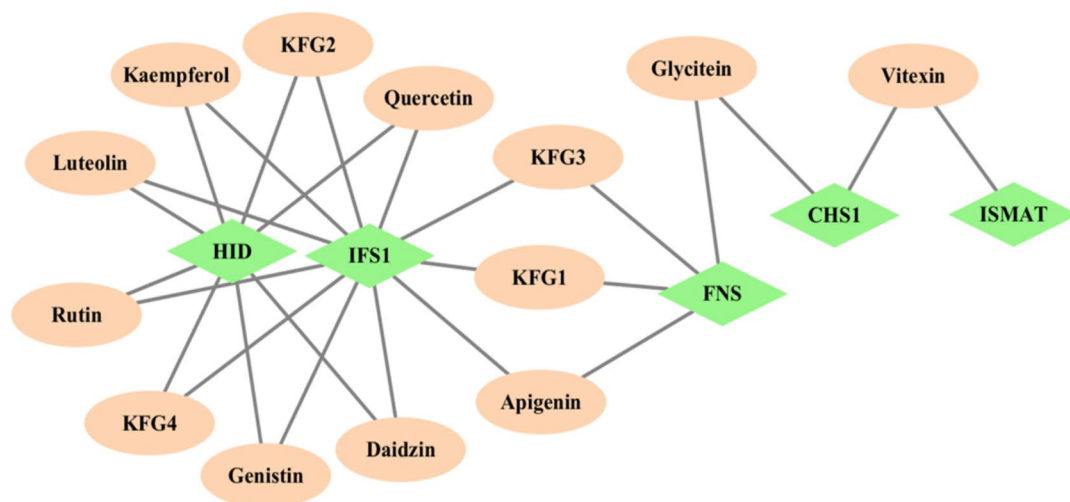


Fig. 6. Correlation between flavonoid/isoflavonoid metabolites and key polyphenolic biosynthetic genes. **(A)** Heatmap showing Pearson correlation coefficients (r) between 16 identified metabolites and six biosynthetic genes (*CHS1*, *CHI*, *FNS*, *HID*, *ISMAT*, and *IFS1*). Colors indicate correlation strength (blue = negative, red = positive). Statistical significance was assessed using two-sided tests and corrected using the Benjamini–Hochberg false discovery rate (FDR). **(B)** Network diagram showing significant metabolite–gene correlations (FDR < 0.05). Nodes denote metabolites (circles) and genes (diamonds); node size reflects degree, and node color indicates community grouping. Edge color represents correlation sign (red = positive, blue = negative) and edge width corresponds to $|r|$. Isolates (degree = 0)—naringenin, daidzein, and genistein—are not shown to indicate measured metabolites without significant associations at the chosen thresholds. This network highlights modules of tightly associated metabolites and genes Table 1.

Resources name	Improved or Land race lines	Subpopulation (Clade)*	Genome Coverages (%)*	Seed Color
Meju (MJ)	Landrace	IV	89.80	Yellow
Bancheongdu (BCD)	Landrace	II	91.11	Brown
KLS88035 (KLS)	Landrace	I	82.33	Green
SLSB397-1 (SL)	Landrace	III	92.86	Brown
Kangwon sujib-33 (KW)	Landrace	V	92.29	Brown
Hoseo (HS)	Improved	I	83.25	Yellow
Suwon98 (SW)	Improved	III	92.04	Yellow
Sohwang (SH)	Improved	IV	87.86	Yellow
Bosug (BS)	Improved	I	93.08	Yellow
Kwangan (KA)	Improved	I	92.54	Yellow

Table 1. List of 10 soybean cultivars selected based on the genome coverage with subpopulation clade. *Note: Clade and genome coverage are based on reference ²⁶.

a correlation network (Fig. 6B). In this network, nodes represent metabolites or genes, and edges represent significant positive (red) or negative (blue) correlations, with edge width scaled to $|r|$ and node size proportional to degree. This network clearly grouped metabolites and genes into coherent clusters, illustrating potential co-regulation modules underlying polyphenolic biosynthesis.

Conclusions

This study conducted a comprehensive metabolite profiling of 10 diverse soybean varieties across six leaf developmental stages and in seeds, with a particular focus on flavonoids and isoflavones. Using UPLC-MS/MS, we identified 16 key compounds that provided valuable insights into the dynamic metabolite landscape of soybeans. Through multivariate analyses, including PCA, PLS-DA, and OPLS-DA, we observed clear distinctions in metabolite profiles between the vegetative and reproductive stages, as well as among different soybean varieties. Notably, naringenin, genistin, and kaempferol were identified as the most significant metabolites for differentiating between varieties and developmental stages, based on VIP scores. This study establishes a basis for optimizing soybean cultivation and improving nutritional quality by revealing metabolic profiles across growth stages and varieties, providing insights for targeted breeding programs.

Materials and methods

Plant materials

From a core collection of 430 soybean cultivars³⁸, ten were selected for flavonoid and isoflavone profiling. These included both landrace varieties (Meju, Bancheongdu, KLS88035, SLSB397-1, Kangwon Sujib2-33) and improved varieties (Hoseo, Suwon98, Sohwang, Bosug, Kwangan)²⁵. The selection was based on a genomic coverage of over 80%, ensuring comprehensive genetic representation. The cultivars were provided by the National Institute of Crop Science in Wanju, Jeollabuk-do, Korea. Planting, cultivation, and sampling were conducted at the Metabolic Engineering Division, Department of Agricultural Biotechnology, National Institute of Agricultural Sciences, Rural Development Administration, 370 Nongsaeongmyeong-ro, Jeonju 54,874, Republic of Korea (35°49'51"N, 127°03'55"E). Each soybean cultivar was planted in a greenhouse on June 22, 2023. On July 3, 2023, seedlings were transferred to an isolated field. Field planting was conducted at 15-cm intervals between seedlings on six 80 × 0.8-m rows, with 20 seedlings planted for each cultivar. Leaves were collected from the most recently developed trifoliate leaves at six distinct growth stages (V2, V4, R1, R3, R5, and R7). Leaves were flash-frozen in liquid nitrogen immediately after collection to prevent metabolic changes. The frozen leaves were lyophilized, ground into a fine powder, and stored at -80 °C until further analysis. All experimental procedures involving plant material complied with RDA, Korean guidelines, and applicable legislation.

Flavonoid and isoflavone extraction

Flavonoids and isoflavones were extracted and analyzed as previously described²⁰. In summary, 25 mg of ground leaf or seed powder was weighed and combined with 200 μL of 1.2 M hydrochloric acid in 80% ethanol (v/v). Internal standards were then added: 50 μL of 2.5 ppm galangin (Sigma-Aldrich, St. Louis, MO, USA; CAS.No: 92342) for flavonoids and 0.5 ppm of 6-methoxyflavone (Sigma-Aldrich; CAS.No: 419737) for isoflavones. The mixture was vigorously stirred at 1200 rpm for 2 h at 26 °C using a Thermomixer Comfort (Eppendorf AG, Hamburg, Germany). Subsequently, the samples were centrifuged at 13,000 × g for 10 min at 25 °C in a TOMY MX-307 centrifuge. The supernatant was passed through a 0.5-μm filter (Advantec Toyo Roshi Kaisha Ltd., Tokyo, Japan) at room temperature.

UPLC-MS/MS analysis of metabolites

Flavonoid and isoflavone compounds were separated and analyzed using an ACQUITY UPLC CSH C18 column (1.7 μm, 150 mm × 2.1 mm; Waters, Milford, MA, USA). The mobile phase consisted of 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B), at a flow rate of 0.3 mL/min. Gradient elution was applied as follows: 0–20 min, 95–75% A; 20–28 min, 75–60% A; 28–30 min, 60–10% A; 30–32 min, 10% A; 32–35 min, 10–95% A; and 35–40 min, 95% A. The injection volume was 1 μL, and the column temperature

was maintained at 30 °C. The UPLC system was connected to a Xevo TQ-S micro mass spectrometer (Waters) operating in the positive ion mode, with nitrogen as the desolvation gas. The desolvation temperature was set to 500 °C at a flow rate of 1,000 L h⁻¹, with a source temperature of 120 °C. The capillary and cone voltages were 0.5 kV and 5 V, respectively. Collision energy for MS/MS was 15–30 V, and data were acquired in multiple reaction monitoring (MRM) mode. Raw MS data were processed using the TargetLynx XS applications manager.

Prior to sample analysis, each target metabolite was confirmed using authentic reference standards, and individual calibration curves were constructed across appropriate concentration ranges (0.002–10 µg mL⁻¹). Standard solutions were analyzed under the same UPLC-MS conditions as the samples, and calibration curves were generated by plotting peak area versus concentration. All calibration curves showed excellent linearity ($R^2 \geq 0.994$), with consistent slopes and intercepts, indicating high sensitivity and reproducibility. Retention times and precursor/product ion transitions were stable across injections, with retention time drift < 0.1 min. Calibration standards were injected immediately before the biological samples to verify system suitability and instrument stability. Subsequently, solvent blanks were run after the highest-concentration standards to ensure the absence of carryover.

The quantification of individual flavonoid metabolites was performed using the external calibration curves derived from authentic standards (Supplementary Table S2). For the kaempferol group, quantification was carried out by UV detection at 350 nm, and peak areas were normalized to an internal standard to correct for injection and instrumental variability. For all analytes, the ratio of the peak area of each metabolite to that of the internal standard was used for concentration estimation based on the selected ions. Representative chromatograms of the authentic standards and sample mixtures are shown in Supplementary Figure S1, confirming the selectivity and reproducibility of the quantification procedure.

RNA extraction and cDNA synthesis

Total RNA was extracted from 10 soybean cultivar leaves at six developmental stages (V2, V4, R1, R3, R5, and R7), and from their seeds, using a Spectrum Plant Total RNA Kit (Sigma-Aldrich), following the manufacturer's guidelines. The purified RNA was dissolved in DEPC-treated water, and 1 µg of RNA was utilized for cDNA synthesis. This synthesis was conducted using a Bio-Rad cDNA synthesis kit (Bio-Rad Laboratories, Hercules, CA, USA) in a thermal cycler. The procedure commenced with an initial denaturation at 72 °C for 3 min, followed by cooling to 4 °C. Subsequently, the master mix was added to a pre-cooled tube and incubated at 42 °C for 4 h. The reaction was concluded by heating the mixture to 70 °C for 10 min. The samples were then stored at -20 °C until further analysis.

Gene expression analysis via quantitative real-time PCR

Gene expression levels were assessed via quantitative real-time polymerase chain reaction (qRT-PCR). This analysis was conducted on a CFX96 Real-Time PCR detection system using the SYBR Premix (Bio-Rad Laboratories). The PCR conditions included an initial denaturation at 95 °C for 5 min, followed by 45 cycles of 15 s at 95 °C for denaturation and 30 s at 60 °C for annealing. Primer pairs were designed using the PrimerQuest Tool from Integrated DNA Technologies (Coralville, IA, USA). Gene expression levels were normalized to the internal reference gene, *Glycine max* L *Actin* (NM_001252731.3). Relative quantification was performed using the Livak and Schmittgen $2^{-\Delta\Delta CT}$ method to evaluate quantitative differences between replicates³⁹. The accession numbers and primer sequences used in this study are listed in Supplementary Table 4.

Data analysis

The data were analyzed using Metaboanalyst 6.0, accessed on September 5, 2024 <https://www.metaboanalyst.ca>⁴⁰, with PCA, PLS-DA, and OPLS-DA to identify significant variables by evaluating the VIP scores. Data normalization was performed to ensure robust and reliable results. PCA, PLS-DA, and OPLS-DA were specifically used to discriminate between different growth stages based on metabolite profiles, with VIP scores indicating the significance of each metabolite. Heatmaps and quantification of selected metabolites were created using the online tool Chiplot (<https://www.chiplot.online/>)^{41,42} to visualize the abundance of key metabolites across different growth stages, highlighting stage-specific patterns in the distribution of flavonoids and isoflavones. Pairwise Pearson correlation coefficients (r) and p-values were computed using R (version 4.5.1). The resulting edge and node tables were exported from R and visualized as a correlation network using the Cytoscape platform. The experiments included three biological and three technical replicates to ensure reliable results.

Data availability

All raw and processed data supporting the findings of this study are included in the manuscript and Supplementary Information files. Additional data are available from the corresponding author upon reasonable request.

Received: 29 July 2025; Accepted: 3 October 2025

Published online: 10 November 2025

References

- Hartman, G. L., West, E. D. & Herman, T. K. Crops that feed the world 2. Soybean—Worldwide production, use, and constraints caused by pathogens and pests. *Food Secur.* **3**, 5–17. <https://doi.org/10.1007/s12571-010-0108-x> (2011).
- Ncube, E., Mohale, K. & Nogemane, N. Metabolomics as a prospective tool for soybean (*Glycine max*) crop improvement. *Curr. Issues Mol. Biol.* **44**, 4181–4196. <https://doi.org/10.3390/cimb44090287> (2022).
- Sahin, I., Bilir, B., Ali, S., Sahin, K. & Kucuk, O. Soy isoflavones in integrative oncology: Increased efficacy and decreased toxicity of cancer therapy. *Integr. Cancer Ther.* **18**, 1534735419835310. <https://doi.org/10.1177/1534735419835310> (2019).

4. Xie, C. L., Park, K. H., Kang, S. S., Cho, K. M. & Lee, D. H. Isoflavone-enriched soybean leaves attenuate ovariectomy-induced osteoporosis in rats by anti-inflammatory activity. *J. Sci. Food Agric.* **101**, 1499–1506. <https://doi.org/10.1002/jsfa.10763> (2021).
5. Zuo, X., Zhao, R., Wu, M., Wan, Q. & Li, T. Soy consumption and the risk of type 2 diabetes and cardiovascular diseases: A systematic review and meta-analysis. *Nutrients* **15**, 1358. <https://doi.org/10.3390/nu15061358> (2023).
6. Liu, S., Zhang, M., Feng, F. & Tian, Z. Toward a “green revolution” for soybean. *Mol. Plant* **13**, 688–697. <https://doi.org/10.1016/j.molp.2020.03.002> (2020).
7. Bahar, N. H. A. et al. Meeting the food security challenge for nine billion people in 2050: What impact on forests?. *Glob. Environ. Change* **62**, 102056. <https://doi.org/10.1016/j.gloenvcha.2020.102056> (2020).
8. Feng, Z., Ding, C., Li, W., Wang, D. & Cui, D. Applications of metabolomics in the research of soybean plant under abiotic stress. *Food Chem.* **310**, 125914. <https://doi.org/10.1016/j.foodchem.2019.125914> (2020).
9. Chaudhary, J., Deshmukh, R., Mir, Z. A. & Bhat, J. A. Metabolomics: An emerging technology for soybean improvement. In *Biotechnology Products in Everyday Life* 175–186 (Springer International Publishing, Cham, 2018).
10. Mani, V., Lee, S. K., Yeo, Y. & Hahn, B. S. A metabolic perspective and opportunities in pharmacologically important safflower. *Metabolites* **10**, 253. <https://doi.org/10.3390/metabo10060253> (2020).
11. Rha, C. S. et al. Statistical discrimination using different machine learning models reveals dissimilar key compounds of soybean leaves in targeted polyphenol-metric metabolomics in terms of traits and cultivation. *Food Chem.* **404**, 134454. <https://doi.org/10.1016/j.foodchem.2022.134454> (2023).
12. Lee, J. H. et al. Characterization of isoflavones accumulation in developing leaves of soybean (*Glycine max*) cultivars. *J. Korean Soc. Appl. Biol. Chem.* **52**, 139–143. <https://doi.org/10.3839/jksabc.2009.026> (2009).
13. Bragagnolo, F. S., Funari, C. S., Ibáñez, E. & Cifuentes, A. Metabolomics as a tool to study underused soy parts: In search of bioactive compounds. *Foods* **10**, 1308. <https://doi.org/10.3390/foods10061308> (2021).
14. Liu, W. et al. The flavonoid biosynthesis network in plants. *Int. J. Mol. Sci.* **22**, 12824. <https://doi.org/10.3390/ijms222312824> (2021).
15. Song, H. H. et al. Metabolomics investigation of flavonoid synthesis in soybean leaves depending on the growth stage. *Metabolomics* **10**, 833–841. <https://doi.org/10.1007/s11306-014-0640-3> (2014).
16. Wen, W., Alseekh, S. & Fernie, A. R. Conservation and diversification of flavonoid metabolism in the plant kingdom. *Curr. Opin. Plant Biol.* **55**, 100–108. <https://doi.org/10.1016/j.pbi.2020.04.004> (2020).
17. Chen, L. R., Ko, N. Y. & Chen, K. H. Isoflavone supplements for menopausal women: A systematic review. *Nutrients* **11**, 2649. <https://doi.org/10.3390/nu11112649> (2019).
18. Li, H. et al. Anti-obesity effects of soy leaf via regulation of adipogenic transcription factors and fat oxidation in diet-induced obese mice and 3T3-L1 adipocytes. *J. Med. Food* **18**, 899–908. <https://doi.org/10.1089/jmf.2014.3388> (2015).
19. Hong, J. et al. Inhibitory characteristics of flavonoids from soybean (*Glycine max* [L.] Merr.) leaf against pancreatic lipase. *Food Biosci.* **56**, 103311. <https://doi.org/10.1016/j.fbio.2023.103311> (2023).
20. Park, Y. J. et al. Metabolite changes in soybean (*Glycine max*) leaves during the entire growth period. *ACS Omega* **8**, 41718–41727. <https://doi.org/10.1021/acsomega.3c06043> (2023).
21. Jung, J. W. et al. Metabolomic variability of different soybean genotypes: β -carotene-enhanced (*Glycine max*), wild (*Glycine soja*), and hybrid (*Glycine max* \times *Glycine soja*) soybeans. *Foods* **10**, 2421. <https://doi.org/10.3390/foods10102421> (2021).
22. Ho, H. M. et al. Difference in flavonoid and isoflavone profile between soybean and soy leaf. *Biomed. Pharmacother.* **56**, 289–295. [https://doi.org/10.1016/S0753-3322\(02\)00191-9](https://doi.org/10.1016/S0753-3322(02)00191-9) (2002).
23. Lee, S. et al. Comprehensive characterization of flavonoid derivatives in young leaves of core-collected soybean (*Glycine max* L.) cultivars based on high-resolution mass spectrometry. *Sci. Rep.* **12**, 14678. <https://doi.org/10.1038/s41598-022-18226-4> (2022).
24. Nleya, T., Sexton, P., Gustafson, K. & Miller, J. M. *Soybean growth stages 1–11* (Best Management Practices for soybean production, IGrow soybean, 2013).
25. Kim, M. S. et al. The patterns of deleterious mutations during the domestication of soybean. *Nat. Commun.* **12**, 97. <https://doi.org/10.1038/s41467-020-20337-3> (2021).
26. Wang, N. et al. Flavonoid components of different color magnolia flowers and their relationship to cultivar selections. *HortScience* **54**, 404–408. <https://doi.org/10.21273/HORTSCI13462-18> (2019).
27. Chen, Z. et al. Integrative analysis of metabolome and transcriptome reveals the improvements of seed quality in vegetable soybean (*Glycine max* (L.) Merr.). *Phytochemistry* **200**, 113216. <https://doi.org/10.1016/j.phytochem.2022.113216> (2022).
28. Kim, M. J. et al. Genetic modification of the soybean to enhance the β -carotene content through seed-specific expression. *PLoS ONE* **7**, e48287. <https://doi.org/10.1371/journal.pone.0048287> (2012).
29. Park, C. H. et al. Effects of light-emitting diodes on the accumulation of glucosinolates and phenolic compounds in sprouting canola (*Brassica napus* L.). *Foods* **8**, 76. <https://doi.org/10.3390/foods8020076> (2019).
30. Yin, Y. C. et al. Overexpressing chalcone synthase (CHS) gene enhanced flavonoids accumulation in *Glycyrrhiza uralensis* hairy roots. *Bot. Lett.* **167**, 219–231. <https://doi.org/10.1080/23818107.2019.1702896> (2019).
31. Zhang, Y. et al. Integrated lipidomic and transcriptomic analyses reveal the mechanism of lipid biosynthesis and accumulation during seed development in sesame. *Front. Plant Sci.* **14**, 1211040. <https://doi.org/10.3389/fpls.2023.1211040> (2023).
32. Das, A., Rushton, P. J. & Rohila, J. S. Metabolomic profiling of soybeans (*Glycine max* L.) reveals the importance of sugar and nitrogen metabolism under drought and heat stress. *Plants (Basel)* **6**, 21. <https://doi.org/10.3390/plants6020021> (2017).
33. Lin, H. et al. Seed metabolomic study reveals significant metabolite variations and correlations among different soybean cultivars. *J. Integr. Plant Biol.* **56**, 826–836. <https://doi.org/10.1111/jipb.12228> (2014).
34. Rao, D. E. & Chaitanya, K. V. Morphological and physiological responses of seven different soybean (*Glycine max* L. Merr.) cultivars to drought stress. *J. Crop Sci. Biotechnol.* **22**, 355–362. <https://doi.org/10.1007/s12892-019-0088-0> (2019).
35. Wang, J., Zheng, Q., Wang, C. & Zhou, A. Classification of soybeans from different habitats based on metabolomic–transcriptomic integration. *Appl. Biol. Chem.* **67**, 30. <https://doi.org/10.1186/s13765-024-00882-x> (2024).
36. Anguraj Vadivel, A. K., Krysiak, K., Tian, G. & Dhaubhadel, S. Genome-wide identification and localization of chalcone synthase family in soybean (*Glycine max* [L.] Merr.). *BMC Plant Biol.* **18**, 325. <https://doi.org/10.1186/s12870-018-1569-x> (2018).
37. Liu, Y. et al. Soyomics: A deeply integrated database on soybean multi-omics. *Mol. Plant* **16**, 794–797. <https://doi.org/10.1016/j.molp.2023.03.011> (2023).
38. Jeong, N. et al. Korean soybean core collection: Genotypic and phenotypic diversity population structure and genome-wide association study. *PLoS ONE* **14**, e0224074. <https://doi.org/10.1371/journal.pone.0224074> (2019).
39. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method. *Methods* **25**, 402–408. <https://doi.org/10.1006/meth.2001.1262> (2001).
40. Pang, Z., Chong, J., Li, S. & Xia, J. MetaboAnalystR 3.0: Toward an optimized workflow for global metabolomics. *Metabolites* **10**, 186. <https://doi.org/10.3390/metabo10050186> (2020).
41. Ji, X., Tang, J. & Zhang, J. Effects of salt stress on the morphology, growth and physiological parameters of *Juglans microcarpa* L. seedlings. *Plants (Basel)* **11**, 2381. <https://doi.org/10.3390/plants11182381> (2022).
42. Li, X. et al. Physiological, biochemical, and genomic elucidation of the *Ensifer adhaerens* M8 strain with simultaneous arsenic oxidation and chromium reduction. *J. Hazard. Mater.* **441**, 129862. <https://doi.org/10.1016/j.jhazmat.2022.129862> (2023).

Acknowledgements

The authors would like to thank Sunyim Bae and Moon Ja Choi for assistance with sample collection and arrangements, Editage (www.editage.co.kr) for English language editing.

Author contributions

Vimalraj Mani, Soyoung Park and Kihun Ha conceived and designed the conceptualization; Vimalraj Mani, Soyoung Park, Sukhee Kim and Kihun Ha collected samples and performed the experiments; Vimalraj Mani, Kihun Ha and Soyoung Park wrote the manuscript. Sichul Lee, Jae Kwang Kim, Jong Sung Lee and Kong-Sik Shin provided feedback and helped improve the final manuscript. Jong-Yeol Lee performed project administration. Soo-Kwon Park, Hye Rang Park and Jin A. Kim performed review and editing. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the Rural Development Administration, Republic of Korea (Project No. PJ01674901). This study was supported by the 2025 Postdoctoral Fellowship Program (V.M.) of the National Institute of Agricultural Sciences, RDA, Republic of Korea.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-23063-2>.

Correspondence and requests for materials should be addressed to K.H. or V.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025