Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Differences in Pigment Composition Between Differently Seeded Cultivated and Wild Soybeans

Qi Yu
Nanjing Institute of Environmental Sciences, Ministry of Ecology and Environment of China
Xin Yin
Nanjing Institute of Environmental Sciences, Ministry of Ecology and Environment of China
Li Zhang
Nanjing Institute of Environmental Sciences, Ministry of Ecology and Environment of China
Zhentao Ren
Institute of Food and Nutrition Development, Ministry of Agriculture and Rural Affairs
Laipan Liu
Nanjing Institute of Environmental Sciences, Ministry of Ecology and Environment of China
Wenjing Shen
Nanjing Institute of Environmental Sciences, Ministry of Ecology and Environment of China
Zhixiang Fang
Nanjing Institute of Environmental Sciences, Ministry of Ecology and Environment of China
Lanzhi Han
Nanjing Institute of Environmental Sciences, Ministry of Ecology and Environment of China
Xiaojing Wang
Hinggan League Institute of Agricultural and Husbandry Sciences
Biao Liu
liubiao@nies.org

Nanjing Institute of Environmental Sciences, Ministry of Ecology and Environment of China

Research Article

Keywords: soybean, pigments, LC-MS/MS, health-promoting functions, MALDI-MSI

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11059442/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

Natural pigments hold significant value in the food and health industries, yet their composition and spatial distribution in seed tissues remain largely unknown. Using liquid chromatography-tandem mass spectrometry, 131 pigments were identified across three cultivated soybean (*Glycine max*) varieties (yellow-, green-, and black-seeded) and wild soybeans (*G. soja* black-seeded). Most metabolites belonged to the flavonoid superfamily, including anthocyanidins, anthocyanins, and procyanidins. Matrix-assisted laser desorption/ionization mass spectrometry imaging combined with multivariate statistics and targeted/semiquantitative assays, revealed that black-seeded soybeans shared a pigment profile more similar to *G. soja* than to light-seeded counterparts. Dark-seeded soybeans exhibited markedly greater pigment diversity and higher overall abundance. These results identify dark-seeded soybeans as valuable sources of bioactive compounds and provide the first comprehensive comparison of pigment composition and spatial distribution between *G. max* and *G. soja*, offering insights for breeding pigment-enriched cultivars.

1. Introduction

The diversity of natural pigments, which are primarily derived from the secondary metabolites of plants, animals, and microorganisms, shapes the vibrant landscapes of nature and plays essential roles in biological survival and human health [1, 2]. Although synthetic pigments dominate multiple industries due to their stable supply and low cost, concerns regarding their safety and the environmental burden associated with their production have prompted renewed focus on natural pigments. Indeed, with advancements in separation and identification technologies and process engineering, natural pigments are increasingly being applied in the fields of food, cosmetics, and medicine [3, 4].

Cultivated soybean (*Glycine max* L. Merr.) is a globally important crop, valued for its high-quality protein, balanced nutritional profile, abundant oil content, and extensive industrial applications. It was domesticated from its wild progenitor, *G. soja* Sieb. et Zucc., which remains an important reservoir of genetic diversity [5, 6]. The seed coats of *G. soja* are predominantly brown/black in color, although domestication and selection have produced a diverse palette of seed coat colors in *G. max*, including yellow, green, brown, black, and mottled types, many of which represent potential sources of natural pigments [7]. Previous research has consistently reported that black soybeans accumulate higher amounts of antioxidant metabolites, particularly anthocyanins and procyanidins, than yellow soybeans [8], with the majority of anthocyanins localized in the palisade layer of the seed coat [9]. Anthocyanins exhibit diverse biological functions, including antioxidant, anti-inflammatory, and metabolism-modulating effects [10–12]. The principal anthocyanins in soybean are glycosides of delphinidin, cyanidin, and petunidin, which are accompanied by minor pigments such as malvidin and peonidin [13, 14]. However, many commercial processing streams remove the seed coat, leaving these bioactive pigments underutilized [15, 16].

Most previous studies have focused on biosynthetic pathways within a single seed coat color or a limited set of cultivated accessions, often without direct comparison to *G. soja*. Moreover, comprehensive mapping of pigment composition and spatial distribution across seed tissues remains limited. In this study, pigment-related metabolite profiles of the seed coats and embryos of three *G. max* varieties (yellow-, green-, and black-seeded) and *G. soja* are analyzed, including quantification of key compounds and visualization of their tissue distributions using matrix-assisted laser desorption/ionization mass spectrometry imaging (MALDI-MSI). Specifically, this study aims to (i) compare pigment composition and content between seed coats and embryos across different color classes of *G. max* and *G. soja*; (ii) identify and quantify the differential pigments that cause these variations between *G. soja* and *G. max* with different seed coat colors; and (iii) reveal spatial localization patterns of representative pigments that underpin the observed color differences and potential functional attributes. Moreover, this work aims to analyze the pigment differences between *G. max* and *G. soja*, clarify the specific relationship between pigment composition and seed coat color, and highlight the roles of pigments in plant resistance and human health functions.

2. Materials and Methods

2.1 Plant Materials

Four soybean seed samples with representative colors were collected from the Huang-Huai-Hai ecological region of China. These included three *G. max* varieties-yellow seeded (YS), green seeded (GS), and black seeded (BS), as well as a brown-seeded *G. soja* accession (WS). All plant material collection was performed with permission from the competent authority, and field sampling complied with relevant national regulations of China. Seed coats and embryos were separated as detailed in Figure S1 and Table S1. All seeds were cultivated under uniform conditions in Nanjing, Jiangsu Province, China. For each variety, three individual plants were selected, and progeny seeds were sampled. All seeds were stored at 4°C until further use. For each variety, three biological replicates (each replicate consisting of seeds from three individual plants) were employed to ensure statistical reliability.

2.2 Chemicals and Reagents

All chemicals and reagents employed in the current study were of chromatographic purity or analytical grade. Acetone, aluminum chloride, diethyl ether, ethanol, ethyl acetate, formic acid, hydrochloric acid, methanol, sodium carbonate, sodium hydroxide, sodium nitrite, and water

were obtained from Thermo Fisher Scientific (Waltham, MA, USA). The MALDI-MS matrix, 2-mercaptobenzothiazole, was purchased from Sigma Aldrich (St. Louis, Missouri, USA).

2.2 Liquid chromatography-tandem mass spectrometry

For each biological replicate of the different soybean varieties, the seed coat or embryo was sampled (200 mg) and ground into a homogeneous fine powder by ball milling prior to transfer to 2 mL centrifuge tubes. Subsequently, the sample was suspended in methanol/water (80:20, v/v; 500 μL) and homogenized for 30 s using a Precellys tissue homogenizer (Bertin Technologies, France). The resulting mixture was subjected to centrifugation, and the supernatant was collected and concentrated (Savant SPD111 SpeedVac concentrator, Thermo Scientific, Waltham, MA, USA). After subsequent freeze-drying, the residual metabolite particles were suspended in 50% acetonitrile (500 μL). Prior to analysis by liquid chromatography–tandem mass spectrometry (LC-MS/MS), the mixture was filtered through a membrane to ensure the absence of particulate matter.

LC-MS/MS was performed using a Waters ACQUITY ultra-performance liquid chromatography (UPLC) system (Thermo Fisher Scientific, Bremen, Germany) coupled with an Orbitrap Fusion Lumos Tribrid mass spectrometer and an electrospray ionization (ESI) source. A binary gradient was employed, with 0.01% formic acid in water as solvent A and acetonitrile/isopropanol (1:1) as solvent B. The gradient of solvent B was ramped from 5% to 45% within 5 min, further increased to 100% within the next 20 min, and then retained constant for 10 min. Metabolite separation was achieved using a Waters ACQUITY UPLC BEH C18 column (50 mm × 2.1 mm i.d., 1.7 μm; Beverly, MA, USA). The flow rate and column temperature were 0.35 mL min^{-1} and 45°C, respectively. Mass spectra were acquired over an *m/z* range of 30-1500 for broad spectrum metabolite coverage. The MS/MS experiments were performed using collision-induced dissociation (CID) with a 35% standardized collision energy to generate informative fragment patterns.

Metabolite identification was achieved by cross-referencing the acquired MS/MS spectra against established repositories, including HMDB (https://hmdb.ca/), METLIN (https://metlin.scripps.edu/), and MassBank (https://massbank.eu/MassBank/). Multivariate statistical analysis including principal component analysis (PCA), was conducted using the MetaboAnalyst 4.0 platform (https://www.metaboanalyst.ca/) to visualize the LC-MS/MS data [17].

2.3 MALDI-MSI

Evenly sized soybean seeds were carefully selected and tissue samples were sectioned using a Leica CM1950 cryostat (Leica Microsystems Inc., Wetzlar, Germany) at -20°C, with the sectioning direction parallel to the soybean cotyledons. The section thickness was set to 20 μm. After sectioning, slices were transferred onto specialized conductive slides for MALDI-2 (Bruker Daltonics, Billerica, MA, USA) using a pre-cooled brush. The slices were gently warmed with finger heat to facilitate adhesion to the slides and then placed in a vacuum drying chamber for 30 min. A matrix solution of acetonitrile/water (70:30, v/v) containing 0.1% trifluoroacetic acid was sprayed uniformly onto the tissue slices using a TM-Sprayer M3 matrix sprayer (HTX Technologies, Orange, NC, USA). The spraying flow rate and nitrogen gas pressure were set at 120 μL min^{-1} and 10 psi to ensure even and thorough coating. The matrix-coated conductive slides were placed onto the target plate of a timsTOF flex MALDI 2 instrument (Bruker Daltonics) and an imaging resolution of 50 μm was employed. The metabolite mass range was set to 50-1500 Da. Tissue slices were scanned using a laser energy of 62% and a laser frequency of 1000 Hz, collecting 100 scans per pixel to generate the raw data files.

Bruker FlexAnalysis 3.4 software was employed for MS peak visualization, peak matching, batch internal calibration, and applying a signal-to-noise (S/N) ratio of 3 for single isotope peaks. Endogenous flavonoid ion distributions were mapped using Bruker FlexImaging 4.1 with a strict mass filter width of 10 ppm.

2.4 Chemical Assays

Seed coats and embryos of the different color types were manually peeled, freeze-dried, and ground into a uniform powder. Total anthocyanins in the tissue samples were quantified using the pH differential method, while the total flavonoid content (TFC) was determined using the $Al(NO_3)_3$ complexation spectrophotometric assay. Procyanidins were measured using the vanillin-HCl method. Carotenoids were extracted from tissues using an ethanol/acetone solution, while the chlorophyll components were extracted using 95% ethanol. These compounds were quantified separately using spectrophotometry.

3. Results and Discussion

3.1 Comparative LC-MS/MS Analysis of Pigment Composition in Four Soybean Types

LC-MS/MS was employed to analyze the metabolomic profiles of the seed coats and embryos from four soybean varieties, namely GS, YS, BS, and WS. Analysis was conducted in both positive and negative ion modes, identifying 3,514 and 3,576 metabolites from the seed coats and embryos of the four varieties, respectively. Among these, 131 metabolites exhibiting significant pigment-related characteristics were identified,

with 112 located in the seed coats and 105 in the embryos. The major pigment classes detected were flavonoids, quinones, and terpenes, although some indole derivatives and products of amino acid metabolism were also identified. Flavonoids were the predominant pigment group, exceeding the combined total of all other metabolite categories. Representative compounds included cyanidin, delphinidin, pelargonidin, and proanthocyanidins (Fig. 1A). Additionally, the seed coats of the four varieties contained 88, 87, 111, and 111 pigment-related metabolites, respectively, while the embryos contained 87, 81, 104, and 105, respectively (Fig. 1B). Of the 112 pigment-related metabolites identified in the seed coats, 81 were common across all four varieties (Fig. 1C). In the embryos, 77 pigment-related metabolites were common across the four varieties (Fig. 1D).

The number of pigment-related metabolites identified in the seed coats was greater than in the embryos, and dark-colored seed varieties exhibited greater pigment-related metabolite diversity in both tissues compared with light-colored varieties (Fig. 1B). Differences in pigment composition between the seed coats and the embryos were more pronounced in YS and GS, with 13 differential metabolites detected between the two tissues for both accessions, while BS and WS exhibited only one difference (Figs. 1C and 1D). In terms of pigment-related metabolite numbers, YS and GS exhibited greater similarity, whereas BS more closely resembled WS and differed markedly from the light-colored varieties, and WS contains all detected pigments.

(A) Quantities of different categories of pigment-related metabolites. (B) Histogram showing the pigment-related metabolite counts in seed coats and embryos. (C, D) Venn diagrams illustrating pigment-related metabolite overlap in seed coats (C) and embryos (D).

3.2 PLS-DA Analysis and Differential Pigment-related Metabolite MALDI-MSI

Partial least squares discriminant analysis (PLS-DA) of the seed coat components revealed significant differences between dark-colored and light-colored soybeans, with smaller differences observed within each color group. Components 1 and 2 explained 92.7% and 2.9% of the variance, respectively, accounting for a total of 95.6% (Fig. 2A). The top 10 most variable metabolites are shown in Fig. 2B. Among these, metabolites with a variable importance in projection (VIP) score > 1 included leucopelargonidin, procyanidin B3, procyanidin C1, procyanidin B2, and quercetin 3-galactoside. Among these 10 metabolites, only the yellow pigments glucoobtusifolin and chrysophanein were expressed at higher levels in YS, whereas the remaining eight (all flavonoids) were higher in dark-colored soybeans. Of the eight metabolites, six were particularly enriched in WS compared with BS, including three anthocyanin precursors-procyanidin C1, B2, and B3-that can be enzymatically converted into anthocyanins [18, 19]. This suggests that WS possesses the highest anthocyanin content, consistent with previous findings [20].

Figure 2C outlines the PLS-DA results for the pigment-related metabolites in the embryo samples. The score plot of components 1 and 2 clearly separates WS and BS from YS and GS, with some overlap between the latter two, indicating smaller differences. Components 1 and 2 explained 87.4% and 9.1% of the variance, respectively, accounting for a total of 96.5%. The six most significant metabolites, based on a VIP score > 1.0, were procyanidin, procyanidin B3, procyanidin B2, quercetin 3-galactoside, quercetin 3-O-rhamnoside, and taxifolin-3-Glucoside (Fig. 2D). Among the 10 most variable metabolites, all were more abundant in dark-colored soybeans, with the highest levels detected in the BS embryos.

To visualize the spatial distributions of the differential metabolites in the four soybean varieties, MALDI-MSI was used for *in situ* analysis, and ion images were generated for six representative metabolites (Fig. 3). The MALDI-MSI results showed that these six differential metabolites were evenly distributed throughout the embryos of dark-colored soybeans, with higher expression levels than in light-colored soybeans, consistent with the metabolomic findings (Fig. 2D). Among these metabolites, procyanidin B3, quercetin 3-galactoside, quercetin 3-O-rhamnoside, and taxifolin-3-glucoside displayed sporadic distributions in the embryos of light-colored soybeans. Azaleatin 3-glucoside was primarily localized at the outermost layer of the YS embryo, whereas in GS it was sparsely distributed across the entire embryo. In contrast, luteolin 7-malonylglucoside was present at very low levels in the embryos of light-colored soybeans. Overall, the distribution patterns of these six metabolites revealed clear similarities between YS and GS and between BS and WS, consistent with the PLS-DA results from the metabolomic data (Figs. 2A and 2C).

The above results reveal that the levels of flavonoids, their derivatives, and anthocyanins are higher in BS and WS than in YS and GS. These pigment-related metabolites not only account for the visible color differences among soybean varieties but also contribute significant biological activities, including antioxidant, anti-inflammatory, and antibacterial effects, as well as enhanced health benefits [21, 22]. Anthocyanins play a crucial role in absorbing light energy, regulating photosynthesis, and protecting plants from oxidative stress, thereby supporting overall plant health [23]. Consequently, elevated levels of flavonoids and anthocyanins contribute to enhanced environmental adaptability in BS and WS [24–27]. Compared with WS, BS embryos contain higher flavonoid levels and larger seeds, exhibiting superior edibility and agronomic traits, rendering them particularly suitable for applications in food, health products, and pharmaceuticals.

3.3 Differential Analysis of Pigments in Soybeans of Different Colors

Pairwise comparisons of the seed coats and embryos across the four soybean varieties were conducted using thresholds of *t*-test $p \leq 0.01$ and fold change ≥ 10 or ≤ 0.1. A total of 77 differential pigment-related metabolites were identified in the seed coats and 81 in the embryos. After merging the datasets, 101 unique differential metabolites were subjected to Pearson correlation-based hierarchical clustering (Fig. 4). It was revealed that BS and WS contained more enriched metabolites in both tissues compared with YS and GS, consistent with previous studies [28].

Additionally, BS exhibited the highest number of enriched metabolites. Although all four soybean embryos appeared yellow to the naked eye, with less visible variation than in the seed coats, the embryos contained a greater number of differential metabolites (81) than the seed coats (77). Moreover, the metabolites enriched in the embryos differed from those in the seed coats, highlighting the potential role of embryos in pigment biosynthesis and their contribution to nutritional and health-related traits.

Heatmap analysis (Fig. 4) revealed that BS and WS shared several enriched metabolites, particularly proanthocyanidins, which represent a class of polyphenols with strong antioxidant capacity. Proanthocyanidins are commonly found in fruits, bark, and seeds, contributing to plant defense against microbial pathogens, pests, and herbivores [29]. They also exhibit immunomodulatory, anticancer, antioxidant, free radical-scavenging, anti-inflammatory, and cardioprotective properties, providing significant health benefits [30–34]. Procyanidin C1 and B5 were enriched in BS and WS seed coats, while procyanidin B1/2/3 and C2 were enriched in both the seed coats and the embryos. Similarly, procyanidin and procyanidin B7 were enriched in BS and WS embryos (Fig. 4). Across both tissues, dark-colored soybeans consistently displayed higher proanthocyanidin levels than light-colored varieties, underscoring their superior antioxidant potential, greater resilience to biotic stress, and enrichment in health-promoting secondary metabolites [35, 36].

Quercetin, another flavonoid, was enriched in the seed coats of dark-colored soybeans. As a yellow plant pigment with a bitter taste and strong antioxidant properties, quercetin enhances tolerance to biotic and abiotic stresses [37]. Furthermore, exogenous application of quercetin promotes lignification and cell wall thickening in parenchymal and cortical layers [38], potentially explaining the thicker seed coats observed in BS and WS. Luteolin, another flavone metabolite exhibiting potent antioxidant activity [39] and free-radical scavenging properties [40], was also highly expressed in the seed coats and embryos of BS and WS.

Despite similarities in pigment composition between BS and WS, distinct metabolite patterns were observed. Specifically, hydroxylated luteolin derivatives (e.g., 8-hydroxyluteolin) and the flavonoid baicalin were more abundant in WS seed coats, whereas xanthopurpurin and (+)-taxifolin were enriched in BS seed coats. WS embryos contained high levels of demethoxycurcumin and procyanidin B6, while BS seed coats and embryos accumulated proanthocyanidin, isorhamnetin, isotheaflavin, and fisetin. In addition, purpurin, pterin, prodelphinidin a1, and rutin were enriched in BS embryos. Most of these compounds belong to the flavonoid family, contributing not only to fruit and flower pigmentation, but also playing critical roles in plant growth and development, protecting against ultraviolet radiation and microbial infection [41].

Although the content of pigment-related metabolites in light-colored soybeans was markedly lower than in dark-colored ones, certain metabolites were still enriched, particularly in YS. In the YS seed coat, chrysophanein, glycitin, and genistein were abundant, while narirutin and apigeninidin were enriched in YS embryos. Among the four varieties, GS exhibited the highest content of luteolin 4'-sulfate in the seed coat and narirutin in the embryo. Chrysophanein, a yellow pigment, has demonstrated leishmanicidal and trypanocidal activities [42], which may contribute to the enhanced resistance of *G. max* against protozoan parasites. Glycitin and genistein-widely distributed natural isoflavone glycosides in legumes-are typical phytoestrogens exhibiting antioxidant, anti-inflammatory, antimicrobial, and antiviral properties [43]. Narirutin, a flavonoid, also possesses antimicrobial and antioxidant activity [44]. Additionally, apigeninidin is highly stable, exhibiting a long half-life, low reaction rate constant, and low Q10 value, contributing to the storage stability of sorghum [45]. Its high abundance in soybean embryos may similarly enhance longevity. In light-colored soybeans, the enriched pigment-related metabolites are predominantly pale yellow, reflecting seed color. However, the relatively limited number of these enriched metabolites suggests reduced environmental resilience, indicating that these varieties are better suited to low-stress, artificial cultivation environments.

3.4 Quantitative Determination of Soybean Pigments and Antioxidant Capacity

As an important and relatively safe class of food additives, natural edible pigments have been used in the food industry for centuries. Among them, anthocyanins, procyanidins, flavonoids, carotenoids, and chlorophylls are widely applied owing to their nutritional and health-promoting properties [46–48]. To validate the results of metabolomic identification and complement the lipid-soluble pigments not fully captured by metabolomics, LC-MS was employed to detect and quantify flavonoids, anthocyanidins, procyanidins, anthocyanins, chlorophylls, and carotenoids, as detailed in Table 1. Among the four soybean seed coats, the greatest variation was observed for the TFC, with corresponding values of 2.72, 3.31, 308.14, and 186.55 mg g^{-1} DW for the YS, GS, BS, and WS seed coats, respectively. In contrast, their embryos exhibited significantly smaller differences (4.29, 5.17, 5.80, and 4.38 mg g^{-1} DW, respectively), suggesting that flavonoid differences between dark- and light-coated soybeans are primarily attributed to the seed coat. Anthocyanidin and procyanidin levels were significantly higher only in the seed coats of dark-coated soybeans, but lower in their cotyledons and in light-coated varieties, consistent with previous reports [9]. Anthocyanin concentrations in the seed coats and embryos of dark-coated soybeans were markedly higher than those in light-coated varieties, whereas chlorophyll and carotenoid contents were relatively lower. These findings indicate that dark-coated soybeans possess higher pigment levels, predominantly in their seed coats, consistent with their deeper pigmentation [49].

Flavonoids are essential secondary metabolites synthesized in nearly all plant organs in response to diverse environmental interactions. They contribute to tissue coloration, attract pollinators and seed dispersers such as insects and birds, and protect plants against pests and herbivores [50]. The high flavonoid content in black and wild soybeans suggests greater stress resistance, enabling these varieties to more

effectively withstand adverse environmental conditions. Such compositional differences may have been reduced during domestication, as seed coat color shifted from dark to light [51]. Furthermore, since natural soybean habitats are primarily riparian and wetland ecosystems, the diversity of anthocyanins and procyanidins in their seed coats may confer adaptive advantages, particularly in protecting seeds against fungal pathogens [52, 53].

To investigate the differences in antioxidant capacity among the four soybean varieties, pigments were extracted from their seed coats and cotyledons, and their free radical scavenging abilities were evaluated using the 2,2-diphenyl-1-picrylhydrazyl (DPPH) assay (Table 2). Although BS contains the highest pigment content, WS demonstrated the strongest antioxidant activity in both its seed coats and cotyledons. This suggests that WS may contain certain substances with particularly strong antioxidant capacities, such as specific anthocyanin monomers [54], or that the proportional composition and synergistic effects among its various antioxidant components are more favorable, resulting in overall enhanced free radical scavenging capacity and antioxidant potential. Furthermore, stress resistance inherent to wild species may drive the production of efficient bioactive compounds, allowing them to adapt more effectively to harsh natural environments [55]. The antioxidant capacity of pigment extractions from the seed coats of WS and BS was higher than that of their cotyledons, whereas in YS and GS, cotyledon extracts exhibited greater antioxidant activity than their seed coats, reflecting the respective pigment contents (Table 1). Overall, dark-seeded soybeans demonstrated substantially higher antioxidant capacity than light-seeded varieties, ultimately highlighting their potential nutritional and functional value.

Table 1
The pigment content in the seed coats and embryos of four types of soybeans

Soybean seeds	Flavonoids ($mg \cdot g^{-1}$ DW)		Anthocyanidins ($mg \cdot g^{-1}$ DW)		Procyanidins ($mg \cdot g^{-1}$ DW)		Anthocyanin ($mg \cdot g^{-1}$DW)		Chlorophyl ($mg \cdot 100g^{-1}$ DW)		Carotenoids ($mg \cdot 100g^{-1}$ DW)	
	Coats	Embryos	Coats	Embryos	Coats	Embryos	Coats	Embryos	Coats	Embryos	Coats	Embryos
YS	2.72 ± 0.22^{b}	4.29 ± 0.23^{b}	0.02 ± 0.00^{c}	0.01 ± 0.00^{b}	0.65 ± 0.04^{c}	0.89 ± 0.01^{bc}	2.72 ± 0.21^{c}	1.76 ± 0.11^{c}	0.90 ± 0.04^{ab}	0.80 ± 0.02^{a}	0.36 ± 0.03^{b}	0.23 ± 0.02^{b}
GS	3.31 ± 0.25^{b}	5.17 ± 0.39^{a}	0.02 ± 0.00^{c}	0.11 ± 0.01^{a}	0.66 ± 0.05^{c}	1.27 ± 0.08^{a}	3.30 ± 0.28^{c}	1.93 ± 0.11^{b}	0.13 ± 0.02^{c}	0.05 ± 0.01^{b}	0.47 ± 0.02^{a}	0.26 ± 0.01^{b}
BS	308.14 ± 23.14^{a}	5.80 ± 0.54^{a}	4.28 ± 0.37^{b}	0.03 ± 0.00^{b}	72.44 ± 7.63^{a}	0.74 ± 0.00^{c}	6.73 ± 0.51^{a}	3.85 ± 0.18^{a}	0.55 ± 0.06^{bc}	0.04 ± 0.01^{c}	0.31 ± 0.01^{bc}	0.50 ± 0.03^{ab}
WS	186.55 ± 11.63^{a}	4.38 ± 0.35^{b}	6.01 ± 0.48^{a}	0.03 ± 0.00^{b}	36.37 ± 2.89^{b}	1.03 ± 0.05^{ab}	4.45 ± 0.40^{b}	2.83 ± 0.16^{b}	0.99 ± 0.06^{a}	0.14 ± 0.04^{b}	0.30 ± 0.01^{c}	0.59 ± 0.02^{a}
Note: Values with different letters in the same column indicate significant differences at $P < 0.05$.												

Table 2
Enzyme activity determination in the seed coats and embryos of four types of soybeans

YS	Coats	Embryos
	16.61%^{c}	**34.47%^{b}**
GS	17.20%^{c}	31.20%^{c}
BS	52.35%^{b}	32.74%^{b}
WS	74.01%^{a}	53.58%^{a}
Note: Values with different letters in the same column indicate significant differences at $P < 0.05$.		

4. Conclusion

Through the analysis of pigment-related metabolites in the seed coats and embryos of yellow-, green-, and black-seeded *G. max* together with *G. soja*, this study clarified the intrinsic relationship between seed coat color and pigment composition. LC-MS/MS identified a total of 131 pigments. Dark-colored soybeans (black-seeded *G. max* and *G. soja*) contained a greater number of pigment types in both the seed coats and embryos. MALDI-MSI revealed that pigments were widely and evenly distributed in the embryos of dark-colored soybeans, while their distribution was limited in light-colored soybeans. Quantitative pigment measurements across the four soybean types aligned with metabolomic results.

Dark-coated soybeans exhibited significantly higher pigment diversity and abundance than yellow- and green-seeded soybeans. Furthermore, the elevated procyanidin and flavonoid contents in both the seed coats and the embryos of dark-coated soybeans may confer enhanced environmental adaptability and stronger bioactivities, such as antioxidant and anti-inflammatory effects. Particularly, the enrichment of anthocyanins in dark-coated soybeans suggests a critical role in protecting seeds from diseases and environmental stress. Overall, this study provides a new perspective for analyzing the metabolic mechanism responsible for soybean seed coat color and offers practical guidance for breeding soybean varieties with abundant pigment profiles and high nutritional value. Future work should focus on combining genomics, metabolomics, and functional gene analysis to further reveal the genetic regulatory network of pigment synthesis and promote breeding progress to enhance stress resistance and nutritional quality in soybeans.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the 2024 Science and Technology Support Project of the Inner Mongolia Innovation Center of Biological Breeding Technology (2024NSZC07), the Fundamental Research Funds for the Central Public-interest Scientific Institution (GYZX20260207), the National Natural Science Foundation of China (32171656), the Biological Breeding-National Science and Technology Major Project (2022ZD04021).

Author Contribution

Q. Y., conducted the main pigment composition analysis and wrote the main manuscript. X. Y., L. Z., and Z. R.,participated in some experimental design and data statistics. L. L., W. S., Z. F., and L. H., participated in the soybean breeding process and the traits' investigation. X. W., and B. L., provided the experimental methods and edited the manuscript. All authors reviewed the manuscript.

Availability of data and materials

The raw metabolomics datasets generated during this study are available from the corresponding author upon reasonable request.

References

1. Handayani I, tri septiana A, Sustriawan B. Natural pigments and antioxidants properties of annatto extract at various pH of distilled water solvent and extraction times. Food Res. 2024;8(2):489–94. https://doi.org/10.26656/fr.2017.8(2).394.
2. Nonglait DL, Gokhale JS. Review insights on the demand for natural pigments and their recovery by emerging microwave-assisted extraction (MAE). Food Bioprocess Technol. 2024;17(7):1681–705. https://doi.org/10.1007/s11947-023-03192-0.
3. Echegaray N, Guzel N, Kumar M, Guzel M, Hassoun A, Lorenzo JM. Recent advancements in natural colorants and their application as coloring in food and in intelligent food packaging. Food Chem. 2023;404:134453. https://doi.org/10.1016/j.foodchem.2022.134453.
4. Prajapati RA, Jadeja GC. Natural food colorants: Extraction and stability study. Materials Today: Proceedings, 2022. 57: pp. 2381–2395. https://doi.org/10.1016/j.matpr.2021.12.151
5. Van K, Kim MY, Shin JH, Kim KD, Lee Y-H, Lee S-H. Molecular evidence for soybean domestication, in Genomics of Plant Genetic Resources: Volume 1. Managing, sequencing and mining genetic resources, 2013, Springer Netherlands: Dordrecht. pp. 465–81. https://doi.org/10.1007/978-94-007-7572-5_19
6. Hymowitz T. On the domestication of the soybean. Econ Bot. 1970;24(4):408–21. https://doi.org/10.1007/BF02860745.
7. Song J, Liu Z, Hong H, Ma Y, Tian L, Li X, Li Y-H, Guan R, Guo Y, Qiu L-J. Identification and validation of loci governing seed coat color by combining association mapping and bulk segregation analysis in soybean. PLoS ONE. 2016;11(7):e0159064.

/https://doi.org/10.1371/journal.pone.0159064.

8. Li W-J, Odongo K, Hironao K-y, Yamashita Y, Ashida H. Polyphenol content and antioxidant activity of two black soybean varieties over five cultivation years. Food Sci Technol Res. 2024;30(3):367–76. https://doi.org/10.3136/fstr.FSTR-D-23-00221.
9. Zhang RF, Zhang FX, Zhang MW, Wei ZC, Yang CY, Zhang Y, Tang XJ, Deng YY, Chi JW. Phenolic composition and antioxidant activity in seed coats of 60 chinese black soybean (*Glycine max* L. Merr.) Varieties. J Agric Food Chem. 2011;59(11):5935–44. https://doi.org/10.1021/jf201593n.
10. Kim J-N, Han SN. Kim Anti-inflammatory and anti-diabetic effect of black soybean anthocyanins. Molecules. 2021;26. https://doi.org/10.3390/molecules26113363.
11. Lee H-J, Cho H-E, Park H-J. Germinated black soybean fermented with *Lactobacillus pentosus* SC65 alleviates DNFB-induced delayed-type hypersensitivity in C57BL/6N mice. J Ethnopharmacol. 2021;265:113236. https://doi.org/10.1016/j.jep.2020.113236.
12. Paik S-S, Jeong E, Jung SW, Ha TJ, Kang S, Sim S, Jeon JH, Chun M-H, Kim. Anthocyanins from the seed coat of black soybean reduce retinal degeneration induced by N-methyl-N-nitrosourea. Exp Eye Res. 2012;97(1):55–62. https://doi.org/10.1016/j.exer.2012.02.010.
13. Lee JH, Kang NS, Shin S-O, Shin S-H, Lim S-G, Suh D-Y, Baek I-Y, Park K-Y, Ha TJ. Characterisation of anthocyanins in the black soybean (*Glycine max* L.) by HPLC-DAD-ESI/MS analysis. Food Chem. 2009;112(1):226–31. https://doi.org/10.1016/j.foodchem.2008.05.056.
14. Xu B, Chang SKC. Antioxidant capacity of seed coat, dehulled bean, and whole black soybeans in relation to their distributions of total phenolics, phenolic acids, anthocyanins, and isoflavones. J Agric Food Chem. 2008;56(18):8365–73. https://doi.org/10.1021/jf801196d.
15. Chen Y, Shan S, Cao D, Tang D. Steam flash explosion pretreatment enhances soybean seed coat phenolic profiles and antioxidant activity. Food Chem. 2020;319:126552. https://doi.org/10.1016/j.foodchem.2020.126552.
16. Loman AA, Ju L-K. Soybean carbohydrate as fermentation feedstock for production of biofuels and value-added chemicals. Process Biochem. 2016;51(8):1046–57. https://doi.org/10.1016/j.procbio.2016.04.011.
17. Xia J, Sinelnikov I, Han B, Wishart D. MetaboAnalyst 3.0-Making metabolomics more meaningful. Nucleic Acids Res. 2015;43(W1). https://doi.org/10.1093/nar/gkv380. p. W251-257.
18. Liu W, Feng Y, Yu S, Fan Z, Li X, Li J, Yin H. The flavonoid biosynthesis network in plants. Int J Mol Sci. 2021;22. https://doi.org/10.3390/ijms222312824.
19. Yu D, Huang T, Tian B, Zhan J. Advances in biosynthesis and biological functions of proanthocyanidins in horticultural plants. Foods. 2020;9. https://doi.org/10.3390/foods9121774.
20. Zhao D, Tao J. Recent advances on the development and regulation of flower color in ornamental plants. Front Plant Sci. 2015;6. https://doi.org/10.3389/fpls.2015.00261.
21. Kan L, Nie S, Hu J, Liu Z, Xie M. Antioxidant activities and anthocyanins composition of seed coats from twenty-six kidney bean cultivars. J Funct Foods. 2016;26:622–31. https://doi.org/10.1016/j.jff.2016.08.030.
22. Sandoval-Ramírez BA, Catalán Ú, Llauradó E, Valls RM, Salamanca P, Rubió L, Yuste S, Solà R. The health benefits of anthocyanins: an umbrella review of systematic reviews and meta-analyses of observational studies and controlled clinical trials. Nutr Rev. 2022;80(6):1515–30. https://doi.org/10.1093/nutrit/nuab086.
23. Sun T, Rao S, Zhou X, Li L. Plant carotenoids: recent advances and future perspectives. Mol Hortic. 2022;2(1):3. https://doi.org/10.1186/s43897-022-00023-2.
24. Pourcel L, Routaboul JM, Cheynier V, Lepiniec L, Debeaujon I. Flavonoid oxidation in plants: from biochemical properties to physiological functions. Trends Plant Sci. 2007;12(1):29–36. https://doi.org/10.1016/j.tplants.2006.11.006.
25. Ren Z, Yin X, Liu L, Zhang L, Shen W, Fang Z, Yu Q, Qin L, Chen L, Jia R, Wang X, Liu B. Flavonoid localization in soybean seeds: Comparative analysis of wild (*Glycine soja*) and cultivated (*Glycine max*) varieties. Food Chem. 2024;456:139883. https://doi.org/10.1016/j.foodchem.2024.139883.
26. Wang X, Li Y, Wang X, Li X, Dong S. Physiology and metabonomics reveal differences in drought resistance among soybean varieties. Bot Stud. 2022;63(1):8. https://doi.org/https://doi.org/10.1186/s40529-022-00339-8.
27. Yin X, Ren Z, Jia R, Wang X, Yu Q, Zhang L, Liu L, Shen W, Fang Z, Liang J, Liu B. Metabolic profiling and spatial metabolite distribution in wild soybean (*G. soja*) and cultivated soybean (*G. max*) seeds. Food Chemistry: X. 2024;101717. https://doi.org/10.1016/j.fochx.2024.101717.
28. Koh K, Youn JE, Kim H-S. Identification of anthocyanins in black soybean (*Glycine max* (L.) Merr.) varieties. J Food Sci Technol. 2014;51(2):377–81. https://doi.org/10.1007/s13197-011-0493-y.
29. Dixon RA, Xie DY, Sharma SB. Proanthocyanidins-a final frontier in flavonoid research? New phytologist, 2005. 165(1): pp. 9–28. https://doi.org/10.1111/j.1469-8137.2004.01217.x
30. Rao LJM, Yada H, Ono H, Ohnishi-Kameyama M, Yoshida M. Occurrence of antioxidant and radical scavenging proanthocyanidins from the Indian minor spice nagkesar (*Mammea longifolia* planch and triana syn). Bioorg Med Chem. 2004;12(1):31–6. https://doi.org/10.1016/j.bmc.2003.10.052.

31. Sano T, Oda E, Yamashita T, Naemura A, Ijiri Y, Yamakoshi J, Yamamoto J. Anti-thrombotic effect of proanthocyanidin, a purified ingredient of grape seed. Thromb Res. 2005;115(1–2):115–21. https://doi.org/10.1016/j.thromres.2004.07.015.
32. Sharma SD, Meeran SM, Katiyar SK. Dietary grape seed proanthocyanidins inhibit UVB-induced oxidative stress and activation of mitogen-activated protein kinases and nuclear factor-κB signaling in in vivo SKH-1 hairless mice. Molecular cancer therapeutics, 2007. 6(3): pp. 995–1005.
33. Subarnas A, Wagner H, Phytomedicine. 2000. 7(5): pp. 401–5. https://doi.org/10.1016/s0944-7113(00)80061-6
34. Zhao M, Yang B, Wang J, Liu Y, Yu L, Jiang Y. Immunomodulatory and anticancer activities of flavonoids extracted from litchi (*Litchi chinensis* Sonn.) pericarp. Int Immunopharmacol. 2007;7(2):162–6. https://doi.org/10.1016/j.intimp.2006.09.003.
35. Cho KM, Ha TJ, Lee YB, Seo WD, Kim JY, Ryu HW, Jeong SH, Kang YM, Lee JH. Soluble phenolics and antioxidant properties of soybean (*Glycine max* L.) cultivars with varying seed coat colours. J Funct Foods. 2013;5(3):1065–76. https://doi.org/10.1016/j.jff.2013.03.002.
36. Nabavi SF, Braidy N, Gortzi O, Sobarzo-Sanchez E, Daglia M, Skalicka-Woźniak K, Nabavi SM. Luteolin as an anti-inflammatory and neuroprotective agent: A brief review. Brain Res Bull. 2015;119:1–11. https://doi.org/10.1016/j.brainresbull.2015.09.002.
37. Singh P, Arif Y, Bajguz A, Hayat S. The role of quercetin in plants. Plant Physiol Biochem. 2021;166:10–9. https://doi.org/10.1016/j.plaphy.2021.05.023.
38. Franco DM, Silva EM, Saldanha LL, Adachi SA, Schley TR, Rodrigues TM, Dokkedal AL, Nogueira FTS, de Almeida LFR. Flavonoids modify root growth and modulate expression of *SHORT-ROOT* and *HD-ZIP III*. J Plant Physiol. 2015;188:89–95. https://doi.org/10.1016/j.jplph.2015.09.009.
39. Gendrisch F, Esser PR, Schempp CM, Wölfle U. Luteolin as a modulator of skin aging and inflammation. BioFactors. 2021;47(2):170–80. https://doi.org/10.1002/biof.1699.
40. Mahmoud AM, Bautista H, Sandhu RJ, M. A., and, Hussein OE. Beneficial effects of citrus flavonoids on cardiovascular and metabolic health. Oxidative Med Cell Longev. 2019;2019(1):5484138. https://doi.org/10.1155/2019/5484138.
41. Panche AN, Diwan AD, Chandra SR. Flavonoids: an overview. J nutritional Sci. 2016;5:e47. https://doi.org/10.1017/jns.2016.41.
42. Domínguez-Carmona DB, Escalante-Erosa F, García-Sosa K, Ruiz-Pinell G, Gutierrez-Yapu D, Chan-Bacab MJ, Moo-Puc RE, Veitch NC, Giménez-Turba A, Peña-Rodríguez LM. Metabolites from roots of *Colubrina greggii* var. *yucatanensis* and evaluation of their antiprotozoan, cytotoxic and antiproliferative activities. J Braz Chem Soc. 2011;22:1279–85. https://doi.org/10.1590/S0103-50532011000700013.
43. Sharifi-Rad J, Quispe C, Imran M, Rauf A, Nadeem M, Gondal TA, Ahmad B, Atif M, Mubarak MS, Sytar O. Genistein: an integrative overview of its mode of action, pharmacological properties, and health benefits. Oxidative Med Cell Longev. 2021;2021(1):3268136. https://doi.org/10.1155/2021/3268136.
44. Jeong H, Nam S-H, Jo J-A, Cho S, Yang K-Y, Aung T, Mirzapour-Kouhdasht A, Park S-W. Extraction and purification of narirutin and hesperidin from green yuzu (*Citrus junos*) and evaluating their biological activities. Process Biochem. 2023;133:132–41. https://doi.org/10.1016/j.procbio.2023.08.015.
45. Bradwell J, Hurd M, Pangloli P, McClure A, Dia VP. Storage stability of sorghum phenolic extracts' flavones luteolin and apigenin. Lwt. 2018;97:787–93. https://doi.org/10.1016/j.lwt.2018.08.006.
46. Gogna S, Kaur J, Sharma K, Prasad R, Singh J, Bhadariya V, Kumar P, Jarial S. Spirulina-an edible cyanobacterium with potential therapeutic health benefits and toxicological consequences. J Am Nutr Association. 2023;42(6):559–72. https://doi.org/10.1080/27697061.2022.2103852.
47. Hadipour E, Taleghani A, Tayarani-Najaran N, Tayarani-Najaran Z. Biological effects of red beetroot and betalains: A review. Phytother Res. 2020;34(8):1847–67. https://doi.org/10.1002/ptr.6653.
48. Jiang T, Ghosh R, Charcosset C. Extraction, purification and applications of curcumin from plant materials-A comprehensive review. Trends Food Sci Technol. 2021;112:419–30.
49. Malenčić D, Cvejić J, Miladinović J. Polyphenol content and antioxidant properties of colored soybean seeds from central Europe. J Med Food. 2012;15(1):89–95. https://doi.org/10.1089/jmf.2010.0329.
50. Alseekh S, Ofner I, Liu Z, Osorio S, Vallarino J, Last RL, Zamir D, Tohge T, Fernie AR. Quantitative trait loci analysis of seed-specialized metabolites reveals seed-specific flavonols and differential regulation of glycoalkaloid content in tomato. Plant J. 2020;103(6):2007–24. https://doi.org/10.1111/tpj.14879.
51. Song J, Xu R, Guo Q, Wu C, Li Y, Wang X, Wang J, Qiu L-J. An omics strategy increasingly improves the discovery of genetic loci and genes for seed-coat color formation in soybean. Mol Breeding. 2023;43(9):71. https://doi.org/10.1007/s11032-023-01414-z.
52. Santinon C, Borges A, Simões M, Gonçalves ASC, Beppu MM, Vieira MGA. Visible-light photoactivated proanthocyanidin and kappa-carrageenan coating with anti-adhesive properties against clinically relevant bacteria. Int J Biol Macromol. 2024;263:130611. https://doi.org/10.1016/j.ijbiomac.2024.130611.
53. Yu X, Jing Y, Jiang Y. Preparation of proanthocyanidin-chitosan complex and its antioxidant and antibacterial properties. Iran Polym J. 2018;27(9):653–61. https://doi.org/10.1007/s13726-018-0642-5.

54. Stintzing FC, Stintzing AS, Carle R, Frei B, Wrolstad RE. Color and antioxidant properties of cyanidin-based anthocyanin pigments. J Agric Food Chem. 2002;50(21):6172–81. https://doi.org/10.1021/jf0204811.
55. Yu Q, Ren Z, Liu L, Zhang L, Yin X, Shen W, Fang Z, Jia R, Liu B. LC/MS and MALDI-MSI reveal the different distribution of terpenoids in *Glycine max* and *Glycine soja* seeds. LWT. 2025;117781. https://doi.org/10.1016/j.lwt.2025.117781.

Figures

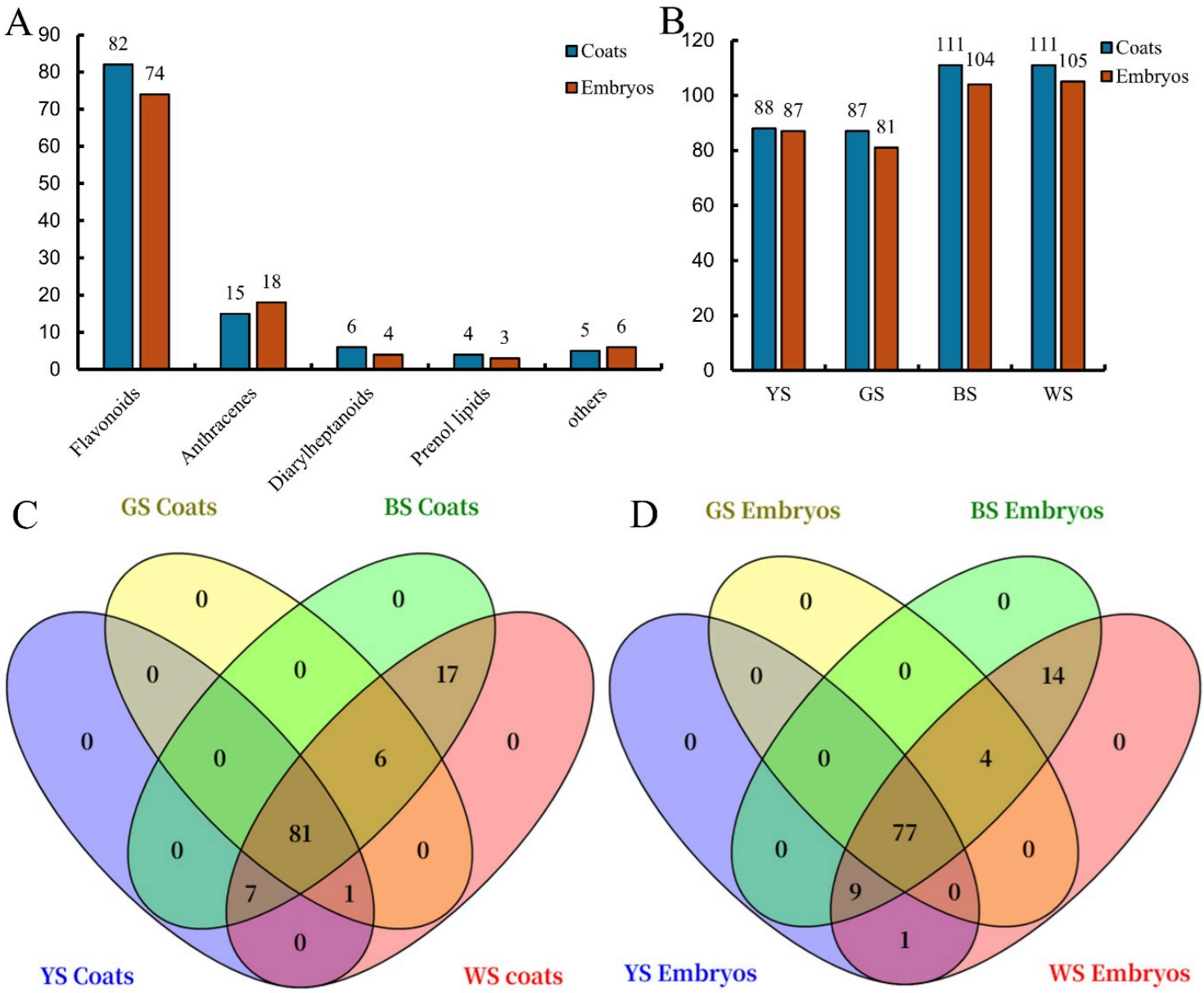


Figure 1

Detection of pigments in the YS, GS, BS, and WS seed coats and embryos.

(A) Quantities of different categories of pigment-related metabolites. (B) Histogram showing the pigment-related metabolite counts in seed coats and embryos. (C, D) Venn diagrams illustrating pigment-related metabolite overlap in seed coats (C) and embryos (D).

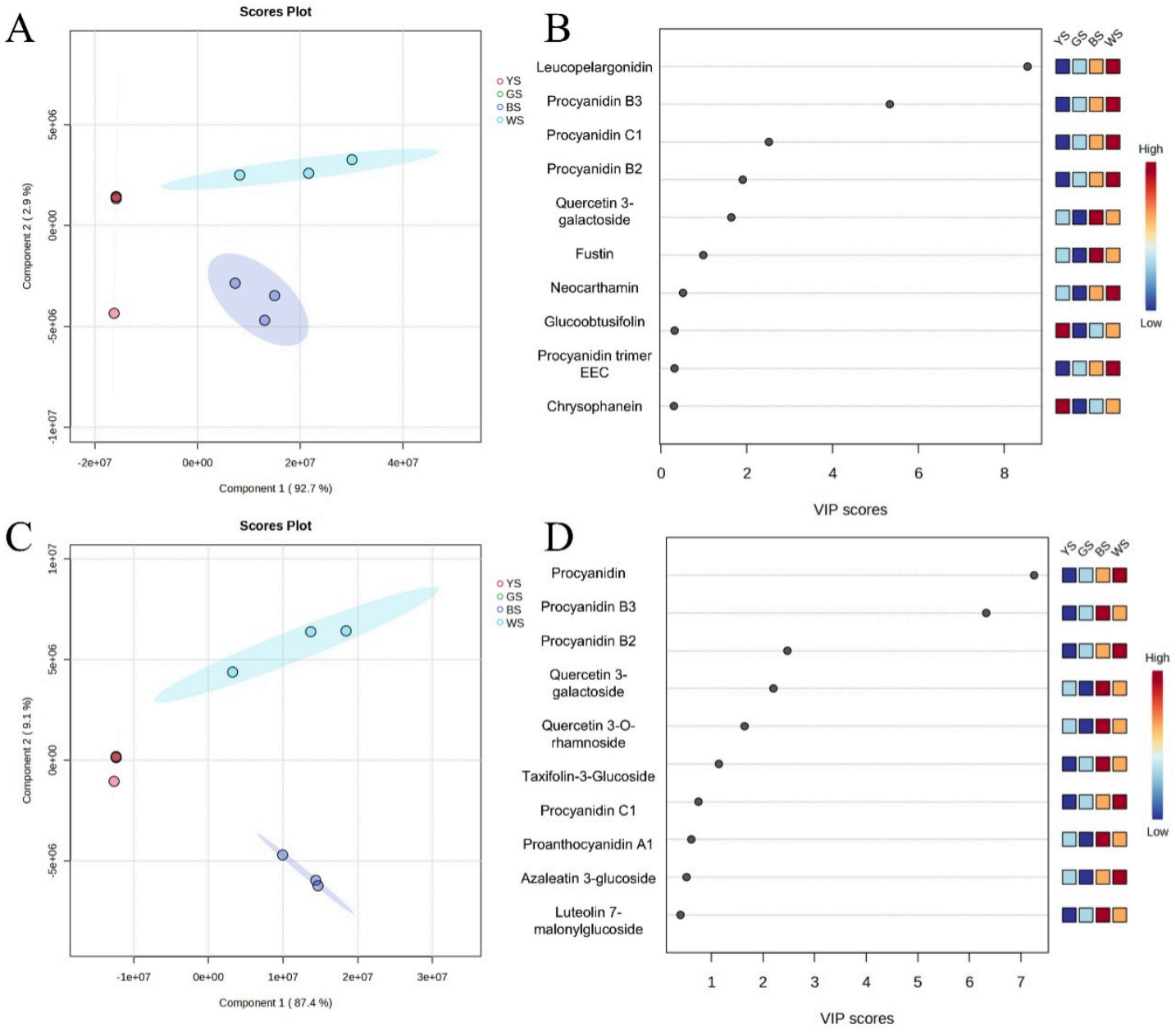


Figure 2

(A, C) PLS-DA and (B, D) VIP analysis of significantly different pigment-related metabolites in GS, YS, BS, and WS. Results are presented for (A, B) the seed coats and (C, D) the seed embryos.

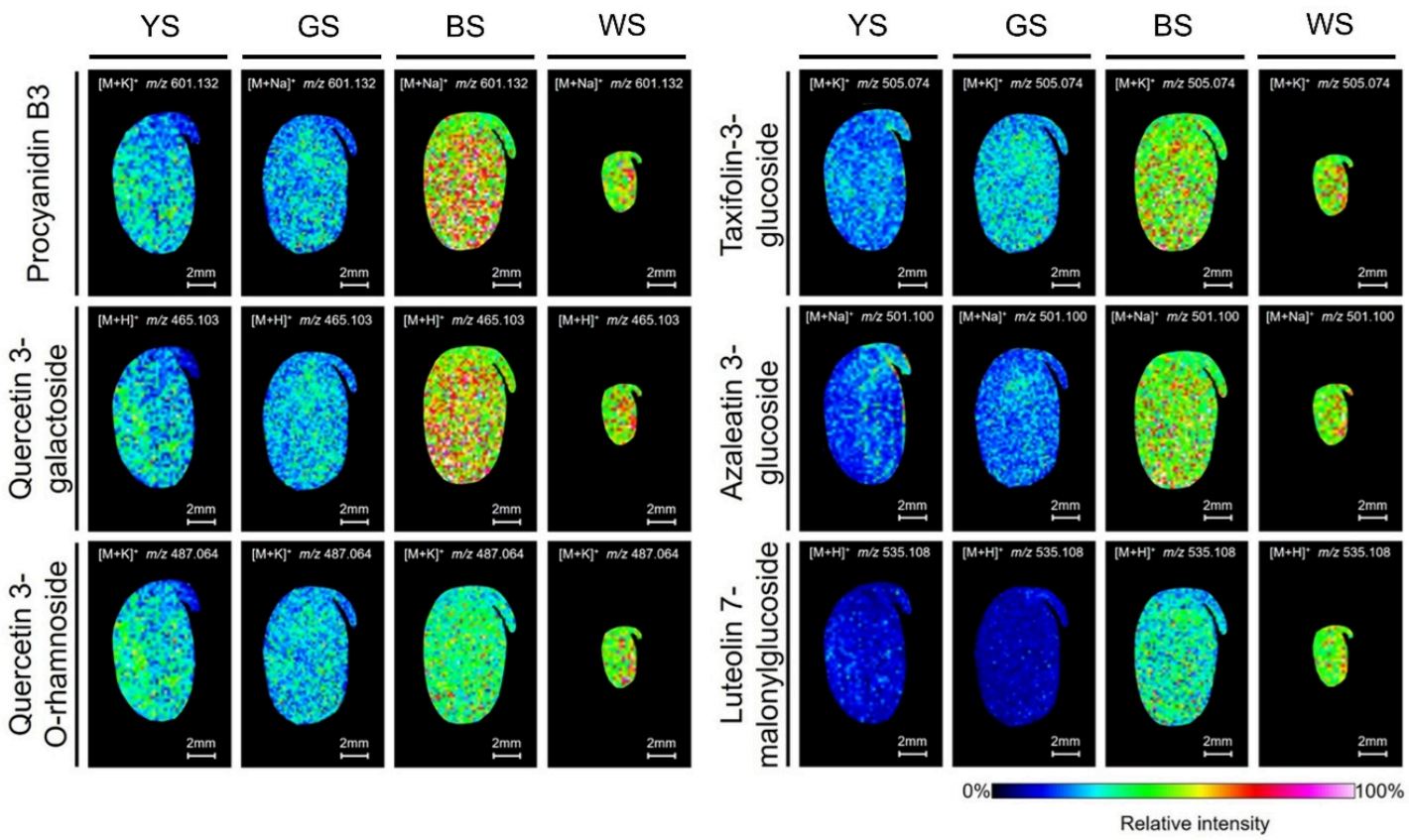


Figure 3

Spatial visualization of differential metabolite distribution in the YS, GS, BS, and WS embryos using MALDI-MSI.

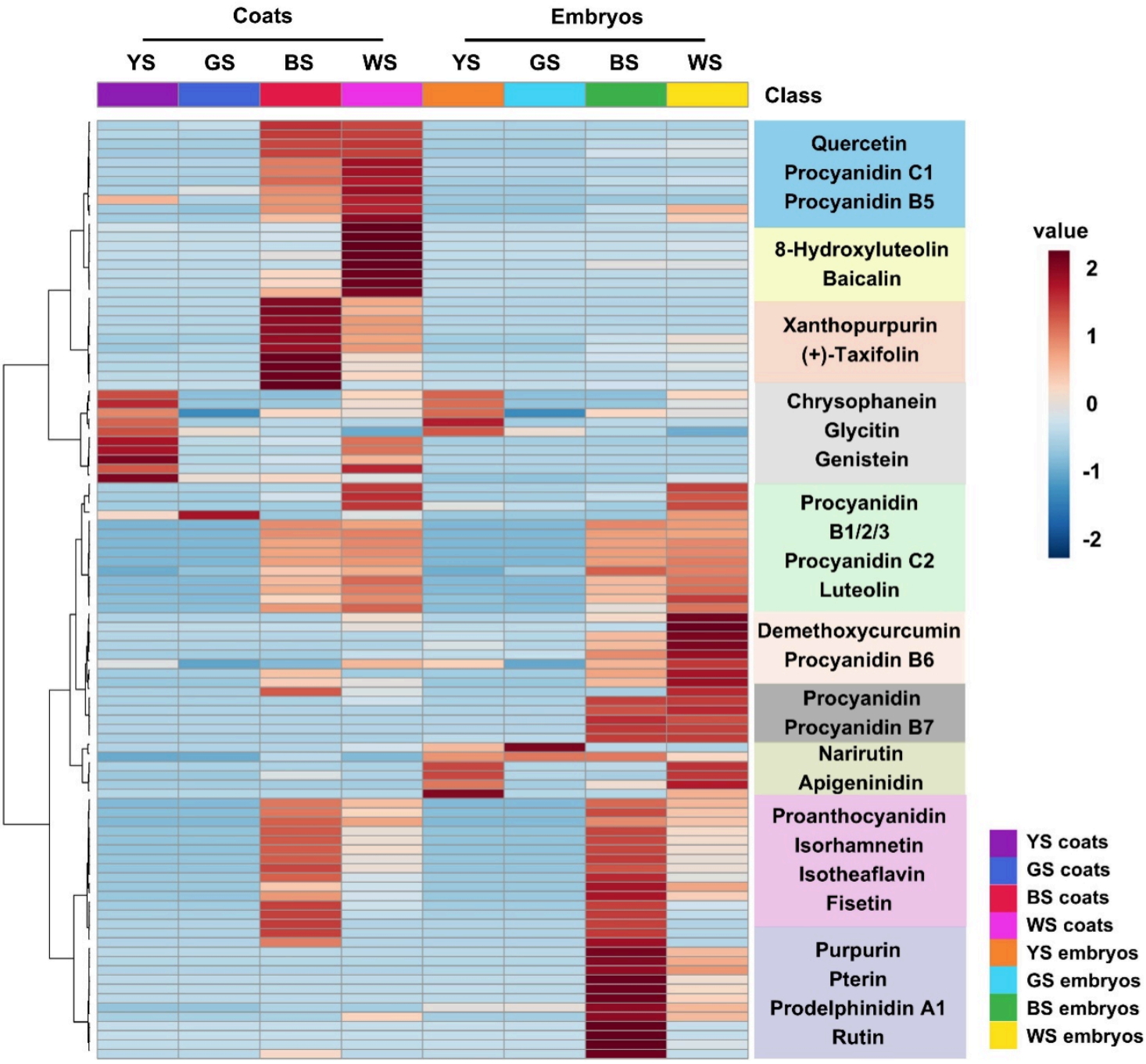


Figure 4

Pearson correlation-based hierarchical clustering heatmap of differential pigment-related metabolites in the seed coats and embryos of BS, GS, YS, and WS.

Supplementary Files

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

- Supportinginformation.docx