

ARTICLE

Open Access



Natural variation in tocopherols, B vitamins, and isoflavones in seeds of 13 Korean conventional soybean varieties

Eun-Ha Kim¹, Jung-Won Jung¹, Oh Suk Yu¹, So-Young Lee¹, Myeong-Ji Kim¹, Sang-Gu Lee¹, Hyoun-Min Park¹, Yongmin Jo², Yongsung Joo³ and Seon-Woo Oh^{1*} 

Abstract

Soybean seeds are excellent sources of tocopherols, B vitamins, and isoflavones, which are well known for their health benefits. This study investigated the influence of environment and genotype on these constituents across 13 Korean soybean varieties cultivated in three locations during the 2017–2019 growing seasons. Statistical analyses, employing both univariate and multivariate methods, revealed significant impacts of genetic and environmental factors on the composition of tocopherols, B vitamins, and isoflavones. Through permutational univariate analysis of variance, the primary contributors to each measured component were identified. Genotype strongly influenced the levels of β - and δ -tocopherols, whereas the interaction between location and year predominantly affected α - and γ -tocopherols. Vitamin B1 content was predominantly determined by genotype, whereas B3 and B6 were influenced by annual variations. Vitamin B2 level was primarily affected by the interplay between environmental and genotypic effects. Genotype had a significant effect on isoflavone components, with the exception of daidzein. Furthermore, early maturing varieties and those with black seed coats exhibited low levels of isoflavone components and total isoflavones, suggesting a relationship between maturity group and seed coat color in isoflavone variation. These findings can be used as reference values for compositional equivalence assessment of genetically modified soybeans.

Keywords Tocopherol, Vitamin B, Isoflavone, Variety, Environment, Soybean

Introduction

Soybean (*Glycine max* (L.) Merr.) is an economically important crop worldwide and is consumed in various forms, such as soybean oil, soybean sprout, soy paste, soymilk, and tofu [1]. Soybean seeds are excellent sources of protein, essential fatty acids, carbohydrate, and vitamins [1]. In addition, soybean seeds contain several high-value health-beneficial secondary metabolites, including isoflavones, phenolic acids, and soyasaponins, which are considered to be the most effective natural antioxidants [2, 3]. Owing to the high economic importance of soybean, new varieties with various traits are continuously developed using genetic engineering technologies and

*Correspondence:

Seon-Woo Oh
ohsw0507@korea.kr

¹Biosafety Division, National Institute of Agricultural Sciences, 370 Nongsanengmyeong-ro, Jeonju, Jeollabuk-do 54874, Republic of Korea

²Department of Statistics and Data Science, College of Natural Science, Dongguk University-Seoul, Seoul 04620, Republic of Korea

³Department of Statistics, College of Natural Science, Dongguk University-Seoul, Seoul 04620, Republic of Korea

conventional breeding strategies and then introduced into the global market [4].

Soybean is the world's largest genetically modified (GM) crop owing to its agronomic, nutritional, and industrial interest and its amenability to genetic transformation. GM soybeans comprise 50% of global biotech crop production [5]. For GM crop biosafety assessment, the compositions of GM crops are compared with those of direct comparators (i.e., near isogenic conventional control) and their conventional comparators that have a history of safe consumption [6, 7]. The range of compositional data of conventional comparators (reference data) are needed to evaluate the composition of GM crops within the natural range of variation [8]. Reference data from comparators grown concurrently in the same field trials of the GM crop, ranges of their compositions reported in the Organization for Economic Co-operation and Development (OECD) consensus documents, the Crop Composition Database (CCDB, www.cropcomposition.org), and peer-reviewed scientific literature can be incorporated into the evaluation of GM crop. Therefore, we started developing a crop composition database (the National Institute of Agricultural Sciences) in South Korea to provide reference data for conventionally commercialized crops such as rice, red pepper, and soybean. Studies were conducted for several years in different regions of South Korea to obtain the ranges of compositional data according to genotype and environmental conditions as recommended in the OECD consensus documents. Our composition data for rice and pepper were deposited in the CCDB (version 9.1) for expansion of crop composition data.

Soybean seeds are a rich source of vitamin E tocopherols, which occur exclusively as tocopherols. Tocopherols are potent chain-breaking antioxidants that protect against lipid peroxidation [9]. Tocopherols can be differentiated into four isoforms (α , β , δ , γ) based on the number and position of the methyl groups attached to its chromanol head. In soybean seeds, γ -tocopherol accounts for approximately 70% of the total tocopherol contents and is the most abundant. Vitamin B, a water-soluble vitamin, is important for the growth and metabolism of living organisms and acts as a co-factor in different metabolic mechanisms. Vitamin B2 comprises riboflavin, flavin adenine dinucleotide (FAD), and flavin mononucleotide (FMN). Riboflavin is a precursor to FAD and FMN. The major components of vitamin B3 are niacin and nicotinamide. Nicotinamide is converted into niacin by nicotinamide deamidase. Vitamin B6 comprises pyridoxal, pyridoxamine, and pyridoxine, and they are converted to pyridoxal 5'-phosphate (PLP), the active form of vitamin B6. PLP plays a pivotal role in amino acid metabolism. In soybean, isoflavones have aglycone structures (daidzein, genistein, and glycitein) and glycosides

such as β -glycosides, acetyl and malonyl glycosides. Malonyl glycosides are the most abundant type of isoflavone, whereas aglycones are present in very low concentrations [10].

Previous studies have demonstrated that the compositions of soybean seed are influenced by genotype, maturity, growing season, locations, and agronomic practices [11–14]. The total amounts and proportions of α -, β -, γ -, and δ -tocopherols are different according to genotype [15, 16]. In soybean seeds, higher levels of α -, γ -, and total tocopherols were observed in early maturing accessions, whereas higher levels of β -tocopherol were obtained from late-maturing accessions [17]. The contents of α -tocopherol increased under condition of warm temperatures or drought stress during seed maturation [18–20]. Few studies have identified the factors that affect B vitamins in soybean seeds. Kim et al. [21] showed that the contents of vitamins B1, B2, B3, B5, B6, and total B vitamins in seeds of 10 black and one yellow soybean varieties varied according to variety. Vedrına-Dragojević et al. [22] determined the contents of vitamins B1, B2, and B3 in four soybean genotypes over the course of 3 years and showed that climatic and genetic factors played a role in vitamin B synthesis. Isoflavone contents in soybean seeds are highly affected by both genetic and environmental factors, such as climate, planting location, crop year, and agricultural management [23–27]. In addition, isoflavone contents, seed coat color, and days of maturity have been reported to have a correlation [28, 29].

We recently published the results of the influence of natural variation according to genotype and growth environment (grown at three locations during the 2017 and 2018 growing seasons) on the proximate, mineral, fatty acid, phytic acid, and trypsin inhibitor contents in 13 Korean commercial soybean varieties widely used for food in South Korea [30]. To expand on our previous results, the contents of vitamin E (tocopherols), B vitamins [B1 (thiamine), B2 (riboflavin), B3 (nicotinic acid), and B6 (pyridoxine)], and isoflavones (aglycones and glycosides) were determined in the same soybean seeds used in our previous study [30], in addition to soybean seeds grown in 2019 at the same locations. Understanding the natural variation in commercialized soybean varieties provides a critical baseline for comparing the characteristics of GM soybeans. We evaluated the natural variations in these components by identifying the effects of genotype, environment, and their interactions. These findings provide reference data for compositional equivalence assessment of GM soybeans.

Materials and methods

Soybean materials and growing conditions

A total of 13 conventional Korean soybean varieties were grown in Suwon (37°27'×50.02'' N, 126°98'×49.59''

E), Iksan (35°94'×40.02'' N, 126°99'×36.60'' E), and Dalseong (35°90'×66.92'' N, 128°44'×76.59'' E) of South Korea during the 2017, 2018, and 2019 growing seasons. Information on the soybean varieties used in this study is presented in Additional file: Table S1. The plots at each site were arranged in a balanced strip design. Each plot consisted of two 10 m long rows with 20 cm seed spacing. Rows were approximately 0.6 m apart, and plots were separated by at least 0.8 m. The soil was a silt clay loam at all sites. Fertilizer was applied prior to planting at a rate of 30-30-32 (N-P-K) kg/ha. Appropriate pesticides were used to control disease and insects. Weeds were removed manually. Seeds were collected from individual plants during the R8 (full maturity) growth stage and then pooled and stored at 4 °C. The planting date and harvesting date were listed in Additional file: Table S2. The monthly precipitation (mm) and average temperature (°C) at the cultivation sites are presented in Table S3 in the Supporting Information. Climate data of each cultivation region were collected from the Korean Meteorological Administration website (<http://weather.go.kr/w/index/do>).

Compositional analysis of vitamin B1 (thiamine) and vitamin B2 (riboflavin)

Vitamins B1 and B2 were determined as described in Arella et al. [31], with slight modifications. A finely ground 0.1 g sample was added to 0.1 M HCl and incubated in a water bath at 100 °C for 30 min. After cooling, the solution was adjusted to pH 4.5 with 2.5 M sodium acetate. A small quantity of distilled water (DW) was added to takadiastase (Sigma-Aldrich, St. Louis, MO, USA), and the solution was incubated for 3 h at 37 °C in a shaking incubator and then diluted to 4 mL with DW. The solution was filtered using a hydrophilic filter (0.45 µm), and the filtrate obtained was used for chromatographic determination of vitamin B2. For analysis of vitamin B1, an aliquot of the filtrate (0.5 mL) was transferred to a new tube containing an alkaline solution (1.5 mL of potassium ferricyanide (1 mL of 1% potassium ferricyanide solution and 24 mL of 3.75 M sodium hydroxide solution). The solution was vortexed, allowed to stand for 1 min, and then passed through a Sep-Pak C₁₈ cartridge (Waters Co. Milford, MA, USA). The cartridge was washed with 0.05 M sodium acetate (5 mL) and then eluted with methanol-water (70:30 v/v) (2 mL). The elute was filtered through a PTFE 0.45 µm syringe filter (Hangahou Anow Microfiltration Co. Ltd), and the filtrate was used for HPLC analysis of vitamin B1 (as thiochrome). HPLC analysis was performed using a 1260 Infinity II Agilent HPLC system (Agilent technologies, Santa Clara, CA, USA) with a C₁₈ column (250 mm × 4.6 mm, 5 µm internal diameter, Waters Co. Milford, MA, USA) isocratically with a mobile phase consisting

of methanol-0.05 M sodium acetate (30:70 v/v) at a flow rate of 1.0 mL/min and temperature of 30 °C. The fluorometric detector was operated at an excitation wavelength of 366 nm and emission wavelength of 522 nm for vitamin B2. Thiamine-HCl and riboflavin standards were purchased from Sigma-Aldrich, St. Louis, MO, USA.

Compositional analysis of vitamin B3 (niacin)

Niacin content was determined using a gas chromatography-time-of-flight mass spectrometry (GC-TOFMS) as described by Lee et al. [32]. Briefly, 0.95 mL of 2.5 M sulfonic acid and 0.05 mL of an internal standard, D4-nicotinic acid (100 ppm in 0.1 N HCl), were added to 0.1 g of samples. After vortexing, samples were autoclaved for 15 min at 121 °C and then cooled down to 20–25 °C. The samples were centrifuged at 13,000 g for 5 min at 4 °C, and 0.2 mL of the supernatant was passed into an HLB PLUS LP extraction cartridge (Waters Co. Milford, MA, USA). After washing the cartridge with 2 mL of DW, niacin was eluted with 70% methanol, and the eluent was dried in a centrifugal concentrator (CVE-2000, Eyla, Tokyo, Japan). For derivatization, 50 µL of *N*-methyl-*N*-(trimethylsilyl) trifluoroacetamide (MSTFA) and 50 µL of pyridine were added and incubated in a thermomixer comfort (Eppendorf, Hamburg, Germany) at 60 °C for 30 min with a 1,200 g mixing frequency. The derivatized samples were analyzed using an Agilent 7890 A gas chromatograph (Agilent, Atlanta, GA, USA) equipped with a 30 m × 0.25 mm i.d. fused silica capillary column coated with 0.25 µm CP-SIL 8 CB low bleed (Varia, Palo Alto, CA, USA). One microliter of each extract was injected into the capillary column at a split ratio of 1:15. Helium was used as a carrier gas at a flow rate of 1.0 mL/min, and the injector temperature was set at 280 °C. The oven temperature was programmed initially at 250 °C for 2 min and then from 250 °C to 290 °C at a rate of 1 mL/min with a final holding time of 8 min. The GC column effluent was analyzed using a Pegasus HT TOF mass spectrometer (LECO, St. Joseph, MI, USA). The temperatures of the source and interface were 250 °C and 290 °C, respectively. The MS spectra were monitored in full scan mode from *m/z* 70 to 600, and the detector voltage was set at 1800 V.

Compositional analysis of vitamin B6 (pyridoxine)

Pyridoxine was determined according to the procedure described by Choi et al. [33] with modifications. Briefly, 2.5 mL of 0.05 M sodium acetate (pH 4.5 using formic acid) was added to 0.1 g of sample, and extraction was performed in a sonication water bath for 30 min at 40 °C. The sample was placed in a shaking incubator for 18 h at 37 °C. DW of 1.5 mL was added to the sample and then vortexed. The sample was centrifuged at 13,000 g for 20 min at 4 °C. The supernatant was transferred into a new tube and then filtered using a PTFE 0.45 µm

syringe filter (Hangahou Anow Microfiltration Co. Ltd). The filtrates obtained were injected into a HPLC (Agilent technologies 1260 Infinity II) system equipped with a C₁₈ column (Waters Symmetry, 5 μm, 4.6×250 mm, Waters Co. Milford, MA, USA). The mobile phase consisted of 0.02 M sodium acetate (pH 3.6 using formic acid, mobile phase A) and acetonitrile (mobile phase B) with a binary gradient elution according to the following program: 0–15 min, 98% A/2% B; 15–20 min, 60% A/40% B; 20–25 min, 60% A/40% B; 25–30 min, 98% A/2% B; 30–42 min, 98% A/2% B at a flow rate of 1.0 mL/min, and column temperature of 30 °C. The fluorometric detector was operated at excitation and emission wavelengths of 292 and 396 nm, respectively. Pyridoxine standard was purchased from Sigma-Aldrich, St. Louis, MO, USA.

Compositional analysis of vitamin E (Tocopherol)

Vitamin E was determined using a GC-TOF-MS according to the procedure of Park et al. [34]. Ethanol containing 0.1% ascorbic acid (w/v) and 0.05 mL of 5α-cholestane (10 μg/mL) as an internal standard was added to the powdered soybean seed sample (0.1 g). After vortexing, the sample was placed in a water bath at 85 °C for 5 min. Thereafter, 120 μL of potassium hydroxide (80%) was added to the sample, and after vortexing, the sample was further incubated in a water bath for 10 min. The samples were immediately placed on ice, and deionized water (1.5 mL) and hexane (1.5 mL) were then added sequentially. After vortexing, the sample was centrifuged (1,200 g, 5 min, 20 °C). The upper layer was transferred to a new tube, and the pellet was re-extracted with hexane. The hexane fraction was then dried using a centrifugal concentrator (CVE-2000; Eyela, Tokyo, Japan). For derivatization, MSTFA (30 μL) and pyridine (30 μL) were added and incubated in a thermomixer comfort (Eppendorf, Hamgurg, Germany) at 85 °C for 5 min with a 1,200 g mixing frequency. The GC-TOF system used is the same as that used for niacin analysis, except the oven temperature, which was programmed from 250 °C to 290 °C at a rate of 10 °C/min with a final holding time of 10 min. The temperatures of source and interface were 250 °C and 290 °C, respectively. The MS spectra were monitored in full scan mode from *m/z* 50 to 800, and the detector voltage was set at 1800 V. The tocopherol standard set was purchased from EMD Millipore Corp. (Billerica, MA, USA).

Compositional analysis of isoflavone

For isoflavone extraction, 1.2 mL of a 75% (v/v) ethanol solution was added to 0.3 g of ground samples and sonicated for 1 h in a sonication water bath at 25 °C. After centrifugation (2,000 g for 10 min at 4 °C), 800 μL of the supernatant was transferred to a new tube and 150 μL of 2 N NaOH was added. The sample was allowed to

stand at ambient temperature for 10 min and then mixed with 50 μL of acetic acid. After filtration through a PTFE 0.45 μm syringe filter (Hangahou Anow Microfiltration Co. Ltd), the isoflavone concentration was analyzed using the HPLC method described in [35] with a slight modification. Thereafter, 0.3 μL of the filtered extraction was applied in the HPLC analysis (Agilent technology 1260 Infinity) equipped with a Cosmosil 2.5 C₁₈-MS-II column (50 mm × 2.0 mm ID, 2.5 μm, Nacalai Tesque, Inc., Japan). A linear HPLC gradient was employed. Solvent A was acetonitrile, and solvent B was 1% trifluoroacetic acid in water according to the program: 0–0.35 min, 90% A/10% B; 0.35–3.96 min, 30% A/70% B; 3.96–4.32 min, 30% A/70% B; 4.32–9 min, 10% A/90% B at a flow rate of 0.58 mL/min, and column temperature at 30 °C. The isoflavone standards (daidzin, daidzein, genistin, genistein, glycitin, and glycitein) were purchased from Sigma-Aldrich, St. Louis, MO, USA.

Statistical analysis

Statistical analysis was performed on the data using SAS Enterprise Guide 7.0 (SAS Institute, 1999). One-way analysis of variance (ANOVA) was conducted to identify the differences in soybean varieties, locations, and cultivation years. Separation of mean was performed using Bonferroni-corrected *t*-tests, and statistically significant differences were determined at a probability level of $p < 0.05$. Principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) was performed with auto-scaled and log-transformed data using SIMCA version 13 (Umetrics, Umeå, Sweden) [36]. The quality of the PLS-DA model was evaluated based on the goodness of fit measured based on R²X (cum) and R²Y (cum) and predictive ability measured based on Q² (cum). To assess whether the PLS-DA models were overfitted, a permutation test was performed with 7-fold cross validation ($n=200$) [37]. The permutational univariate analysis of variance (PERMANOVA) used to define the explanatory power of the variance components of varieties (V), years (Y), locations (L), and their interactions (V×L, V×Y, L×Y, V×L×Y) with compositions using the Plymouth Routines in Multivariate Ecological Research (PRIMER) software package version 7.0 with Add on PERMANOVA (PRIMER-E Ltd, UK) [38, 39]. The test was computed of raw data using 999 permutations at a significant level of 0.01.

Results and discussion

Tocopherol contents

The contents of individual and total tocopherols in 13 Korean soybean varieties across three locations (Suwon, Iksan, and Dalseong) over 3 years (2017, 2018, and 2019) are presented in Table 1. It is known that γ-tocopherol is a major form of seed tocopherols in soybeans, with a

Table 1 Contents of tocopherols and B vitamins in seeds of 13 soybean varieties across three locations for 3 years

Variety	α -Toc	β -Toc	δ -Toc	γ -Toc	Total Toc	B1	B2	B3	B6
CHO	8.53 ± 3.79 ^{abc}	2.78 ± 0.90 ^b	10.79 ± 2.72 ^{bc}	44.60 ± 17.13 ^c	66.69 ± 23.71 ^{bc}	5.49 ± 1.18 ^b	0.79 ± 0.20 ^a	57.67 ± 3.41 ^a	3.35 ± 1.23 ^b
CJ-3	7.55 ± 3.01 ^{abcd}	2.02 ± 0.27 ^{de}	8.95 ± 1.52 ^{de}	58.54 ± 15.46 ^{abc}	77.06 ± 19.11 ^{abc}	4.14 ± 0.67 ^{cdef}	0.96 ± 0.38 ^a	57.03 ± 5.87 ^a	3.68 ± 0.84 ^{ab}
DC	9.89 ± 4.12 ^{ab}	2.68 ± 0.57 ^b	10.40 ± 1.32 ^{bc}	59.44 ± 19.22 ^{ab}	82.42 ± 24.51 ^{ab}	4.37 ± 0.77 ^{cde}	0.94 ± 0.26 ^a	55.18 ± 5.58 ^a	3.15 ± 0.89 ^b
DP-2	8.65 ± 5.57 ^{abc}	1.86 ± 0.58 ^{de}	10.93 ± 2.22 ^{bc}	62.39 ± 23.25 ^{ab}	83.83 ± 29.67 ^{ab}	4.04 ± 0.53 ^{cdef}	0.82 ± 0.19 ^a	58.18 ± 4.63 ^a	3.37 ± 1.0 ^b
DW	4.81 ± 1.84 ^d	1.60 ± 0.23 ^{ef}	14.53 ± 1.59 ^a	68.31 ± 11.18 ^a	89.24 ± 13.55 ^a	3.92 ± 0.64 ^{def}	0.85 ± 0.19 ^a	58.02 ± 8.06 ^a	3.03 ± 0.74 ^b
MS	9.54 ± 3.00 ^{ab}	3.56 ± 0.67 ^a	10.17 ± 0.96 ^{cd}	48.18 ± 9.04 ^{bc}	71.45 ± 11.37 ^{abc}	4.32 ± 0.62 ^{cde}	0.87 ± 0.24 ^a	56.44 ± 5.20 ^a	3.13 ± 0.73 ^b
PSN	5.10 ± 1.94 ^{cd}	1.18 ± 0.16 ^f	10.02 ± 1.35 ^{cd}	62.80 ± 16.02 ^{ab}	79.11 ± 18.79 ^{abc}	3.95 ± 0.52 ^{def}	1.05 ± 0.38 ^a	59.93 ± 4.03 ^a	3.15 ± 0.76 ^b
PW	10.73 ± 4.20 ^a	2.61 ± 0.27 ^{bc}	8.73 ± 1.02 ^{de}	56.68 ± 10.85 ^{abc}	78.76 ± 14.44 ^{abc}	3.77 ± 0.85 ^{ef}	0.98 ± 0.50 ^a	56.39 ± 5.48 ^a	3.42 ± 0.86 ^{ab}
SCJ	10.07 ± 3.67 ^{ab}	1.97 ± 0.38 ^{de}	6.33 ± 1.70 ^f	51.46 ± 15.89 ^{bc}	78.76 ± 20.44 ^{bc}	4.79 ± 0.80 ^{bc}	1.00 ± 0.36 ^a	56.33 ± 5.18 ^a	4.19 ± 1.01 ^a
SO	8.06 ± 3.95 ^{abcd}	2.09 ± 0.53 ^{de}	8.36 ± 1.59 ^e	44.43 ± 14.83 ^c	69.83 ± 20.20 ^c	6.40 ± 1.11 ^a	0.82 ± 0.19 ^a	58.55 ± 7.25 ^a	3.40 ± 0.80 ^{ab}
SP	11.13 ± 6.13 ^a	3.91 ± 0.87 ^a	11.97 ± 1.44 ^b	49.07 ± 15.11 ^{bc}	76.08 ± 20.87 ^{abc}	3.51 ± 0.32 ^f	0.83 ± 0.18 ^a	57.27 ± 4.68 ^a	2.92 ± 0.74 ^b
TK	7.01 ± 2.89 ^{abcd}	2.10 ± 0.48 ^{de}	11.82 ± 1.42 ^b	61.24 ± 12.23 ^{ab}	82.17 ± 15.10 ^{ab}	4.64 ± 0.81 ^{cd}	0.81 ± 0.23 ^a	56.47 ± 3.38 ^a	3.52 ± 0.76 ^{ab}
UR	7.62 ± 3.36 ^{abcd}	2.12 ± 0.53 ^{cd}	11.82 ± 2.13 ^b	57.03 ± 18.29 ^{abc}	78.58 ± 23.52 ^{abc}	4.31 ± 1.20 ^{cde}	0.91 ± 0.29 ^a	59.22 ± 4.40 ^a	2.97 ± 0.63 ^b
P-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.0104	0.078	<0.0001

Data are presented as the mean and standard deviation ($\mu\text{g/g}$ dry weight basis). Means in the same column followed by the same letter(s) are not significantly different at $P < 0.05$ based on the Bonferroni test. Total toc summed up the contents of four tocopherols. Toc, tocopherol

concentration ranging from 60 to 70% in soybean seeds. In contrast, α -, β -, and δ -tocopherols are often lower in concentrations [13]. It has been reported that total amounts and proportions of α -, β -, γ -, and δ -tocopherols are different according to genotypes [14, 15]. In the present study, the total tocopherol content of soybeans ranged from 66.7 (CHO) to 89.2 $\mu\text{g/g}$ (DW) across all the locations and years. The contents of γ - and δ -tocopherols accounted for 65 to 80% and 8 to 12% of the total tocopherols, respectively, whereas those of α - and β -tocopherols accounted for 5 to 15% and 2 to 8%, respectively. The concentration of α - and β -tocopherols ranged from 4.8 $\mu\text{g/g}$ (DW) to 11.1 $\mu\text{g/g}$ (SP) and from 1.2 $\mu\text{g/g}$ (PSN) to 3.9 $\mu\text{g/g}$ (SP), respectively. The concentration of δ - and γ -tocopherols ranged from 6.3 (SCJ) to 14.5 $\mu\text{g/g}$ (DW) and from 44.4 $\mu\text{g/g}$ (SO) to 68.3 $\mu\text{g/g}$ (DW), respectively. In addition, Ghosh et al. [17] used 493 soybean accessions of different origins belonging to seven maturity groups and showed relationship between maturity groups of cultivars and their tocopherol concentrations: higher levels of α -, γ -, and total tocopherols were observed in early maturing accessions, whereas higher levels of β -tocopherol were obtained from late-maturing accessions. However, in our study, no consistent relationship was observed between maturity groups and their tocopherol content (Table 1, Additional file: Tables S4–S6). Further, two early maturity varieties, CHO and SO, showed the lowest γ - and total tocopherols contents (Table 1) across environments. The relationship between tocopherol content and maturity groups most likely did not reflect due to our small sample size.

In addition to the genotypic factor, the yearly difference at the same location was important for individual and total tocopherol contents, with the exception of β -tocopherol at Dalseong (Table 2). Total tocopherol and

four individual tocopherol contents were the lowest in 2017 in Iksan. Overall, these results were observed in all the varieties grown in Iksan in 2017 compared to those grown in the other 2 years (Additional file: Table S4). Some varieties, such as CHO, DC-2, DP, and UR in Iksan in 2017, had lower tocopherol content compared to those in the other 2 years. Tocopherol composition is greatly affected by growth environment, especially during the stages R5–R7 [13, 14]. For instance, higher α -tocopherol concentrations and lower δ - and total tocopherol concentrations were observed in warmer environments than in cooler environments [19]. In addition, soil moisture and irrigation during seed-filling period affect the tocopherol composition of soybean seeds. Drought stress increased α -tocopherol concentrations but decreased its δ - and γ -tocopherols [18]. However, in Iksan during 2017, the average air temperature and precipitation in September and October, when seed filling occurs, were not exceptional compared to those in 2018 and 2017 (Additional file: Table S3). The lower tocopherol concentrations in Iksan in 2017 might be attributable to other conditions such as soil fertility and crop management practices [13].

B vitamins contents

It is known that the contents of B vitamins in crops such as beans and wheat grains are influenced by variety, seed maturity, and cultivation environment [22, 40]. However, studies for the effects of environmental and genetic factors on vitamin B accumulation in soybean seeds were not sufficiently reported. Vedrına-Dragojević et al. [22] reported that thiamine and riboflavin concentrations were markedly different between four soybean varieties and climatic factors in the 3 years. In contrast, niacin content was similar between the cultivars in the same year. Kim et al. [21] showed that the vitamins B1, B2,

Table 2 Contents of tocopherols and B vitamins in seeds of 13 soybean varieties in each environment across varieties

Variety	α -Toc	β -Toc	δ -Toc	γ -Toc	Total Toc	B1	B2	B3	B6
Suwon									
2017	13.13 \pm 4.83 ^a	2.67 \pm 1.01 ^a	10.84 \pm 2.27 ^a	70.38 \pm 12.43 ^a	97.02 \pm 15.05 ^a	5.12 \pm 0.84 ^a	0.71 \pm 0.15 ^b	53.73 \pm 6.34 ^b	2.56 \pm 0.55 ^b
2018	7.45 \pm 2.75 ^b	2.04 \pm 0.53 ^b	11.33 \pm 2.60 ^a	60.89 \pm 10.33 ^b	81.71 \pm 12.31 ^b	4.25 \pm 1.18 ^b	0.83 \pm 0.15 ^a	60.06 \pm 1.91 ^a	3.74 \pm 0.52 ^a
2019	7.02 \pm 2.07 ^b	2.09 \pm 0.78 ^b	9.28 \pm 2.05 ^b	46.46 \pm 9.62 ^c	64.85 \pm 10.23 ^c	4.13 \pm 0.90 ^b	0.88 \pm 0.17 ^a	58.97 \pm 4.35 ^a	3.53 \pm 0.39 ^a
P-value	<0.0001	0.0009	0.0004	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Iksan									
2017	2.76 \pm 1.93 ^c	1.77 \pm 0.94 ^b	8.91 \pm 3.21 ^b	29.65 \pm 17.30 ^c	43.08 \pm 21.94 ^c	4.36 \pm 0.94 ^b	0.73 \pm 0.17 ^c	51.16 \pm 3.05 ^c	2.25 \pm 0.36 ^b
2018	11.53 \pm 3.02 ^a	2.50 \pm 0.69 ^a	11.91 \pm 2.55 ^a	74.66 \pm 9.24 ^a	100.60 \pm 10.00 ^a	3.61 \pm 0.47 ^c	0.94 \pm 0.22 ^b	61.69 \pm 2.63 ^a	3.73 \pm 0.52 ^a
2019	10.00 \pm 2.65 ^b	2.35 \pm 0.86 ^a	9.07 \pm 2.00 ^b	55.53 \pm 11.22 ^b	76.96 \pm 11.88 ^b	5.38 \pm 1.48 ^a	1.37 \pm 0.51 ^a	57.06 \pm 4.12 ^b	3.66 \pm 0.75 ^a
P-value	<0.0001	0.0004	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Dalseong									
2017	5.19 \pm 3.66 ^b	2.40 \pm 1.05 ^a	9.89 \pm 1.79 ^b	48.73 \pm 10.32 ^b	66.22 \pm 12.96 ^b	4.36 \pm 0.78 ^b	0.85 \pm 0.21 ^a	56.39 \pm 5.37 ^a	2.25 \pm 0.36 ^a
2018	8.48 \pm 2.43 ^a	2.44 \pm 0.72 ^a	11.01 \pm 2.08 ^{ab}	56.58 \pm 11.03 ^a	78.52 \pm 10.93 ^a	3.61 \pm 0.79 ^a	0.87 \pm 0.19 ^a	62.72 \pm 2.51 ^b	3.87 \pm 0.49 ^b
2019	9.70 \pm 2.49 ^a	2.83 \pm 0.97 ^a	11.09 \pm 2.47 ^a	58.46 \pm 10.95 ^a	82.08 \pm 11.26 ^a	5.38 \pm 0.85 ^b	0.85 \pm 0.14 ^a	58.97 \pm 4.35 ^a	4.36 \pm 0.61 ^c
P-value	<0.0001	0.0858	0.0243	0.0002	<0.0001	<0.0001	0.7311	<0.0001	<0.0001

Data are presented as the mean and standard deviation ($\mu\text{g/g}$ dry weight basis). Means in the same column followed by the same letter(s) are not significantly different at $P < 0.05$ based on the Bonferroni test. Total toc summed up the contents of four tocopherols. Toc, tocopherol

Table 3 Contents of isoflavones in seeds of 13 soybean varieties across three locations for 3 years

Variety	Daidzein	Daidzin	Genistein	Genistin	Glycitein	Glycitin	Total isoflavone
CHO	8.2 \pm 1.7 ^{ab}	565.4 \pm 142.0 ^{de}	5.0 \pm 0.8 ^{bcde}	883.2 \pm 171.3 ^{ef}	3.2 \pm 0.7 ^c	88.1 \pm 21.5 ^{bc}	1553.0 \pm 313.5 ^{ef}
CJ-3	11.7 \pm 5.0 ^{ab}	546.8 \pm 153.1 ^e	3.7 \pm 0.7 ^{efg}	1022.2 \pm 197.4 ^{de}	3.3 \pm 2.0 ^c	46.7 \pm 28.1 ^{ef}	1634.452 \pm 372.3 ^{ef}
DC	15.4 \pm 23.4 ^{ab}	768.8 \pm 201.4 ^{cd}	5.3 \pm 1.3 ^{bcd}	1261.2 \pm 371.9 ^{cd}	2.2 \pm 1.5 ^{cd}	88.1 \pm 22.8 ^{bc}	2140.9 \pm 513.7 ^{cd}
DP-2	11.6 \pm 3.3 ^{ab}	1156.0 \pm 243.5 ^{ab}	5.6 \pm 1.3 ^{abc}	1645.8 \pm 368.7 ^a	2.60 \pm 1.0 ^{de}	168.1 \pm 47.4 ^a	2989.8 \pm 579.0 ^a
DW	13.7 \pm 14.6 ^{ab}	887.9 \pm 207.4 ^c	4.2 \pm 1.5 ^{cdef}	1239.5 \pm 247.4 ^{cd}	3.9 \pm 0.5 ^c	111.5 \pm 21.1 ^b	2260.6 \pm 459.9 ^{cd}
MS	13.9 \pm 16.2 ^{ab}	846.9 \pm 171.6 ^c	3.3 \pm 1.2 ^{fg}	932.8 \pm 155.8 ^{ef}	1.0 \pm 1.0 ^d	52.6 \pm 30.6 ^{def}	1850.6 \pm 335.5 ^{de}
PSN	17.0 \pm 15.2 ^a	755.1 \pm 198.6 ^{cde}	3.9 \pm 1.2 ^{defg}	1074.3 \pm 280.5 ^{de}	3.1 \pm 3.8 ^c	81.1 \pm 22.2 ^c	1934.4 \pm 480.0 ^{de}
PW	16.5 \pm 9.2 ^a	1148.8 \pm 260.3 ^{ab}	6.7 \pm 2.9 ^a	1462.3 \pm 310.1 ^{abc}	6.2 \pm 4.6 ^b	148.9 \pm 53.8 ^a	2789.4 \pm 601.0 ^{ab}
SCJ	7.7 \pm 8.4 ^{ab}	562.3 \pm 128.2 ^{de}	2.6 \pm 1.0 ^f	894.5 \pm 237.9 ^{ef}	10.1 \pm 1.5 ^a	76.2 \pm 25.1 ^{cd}	1553.3 \pm 365.4 ^{ef}
SO	5.6 \pm 3.3 ^b	280.5 \pm 66.8 ^f	3.7 \pm 0.8 ^{efg}	584.5 \pm 151.0 ^g	2.6 \pm 0.5 ^{cd}	36.4 \pm 5.3 ^f	913.3 \pm 299.3 ^g
SP	10.4 \pm 3.4 ^{ab}	965.8 \pm 179.6 ^{bc}	6.4 \pm 2.4 ^{ab}	1381.7 \pm 375.7 ^{bc}	7.2 \pm 3.3 ^b	113.8 \pm 23.5 ^b	2485.3 \pm 543.8 ^{bc}
TK	7.8 \pm 5.6 ^{ab}	586.8 \pm 520.4 ^{de}	3.2 \pm 1.6 ^{fg}	688.6 \pm 223.7 ^{fg}	2.3 \pm 1.4 ^{cd}	70.5 \pm 19.4 ^{cde}	1258.3 \pm 410.8 ^f
UR	14.0 \pm 8.4 ^{ab}	1356.9 \pm 283.0 ^a	5.5 \pm 1.1 ^{abc}	1615.3 \pm 293.5 ^{ab}	2.7 \pm 0.8 ^{cd}	89.8 \pm 26.2 ^{bc}	3084.2 \pm 568.9 ^a
P-value	0.0006	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

Data are presented as the mean and standard deviation ($\mu\text{g/g}$ dry weight basis). Means in the same column followed by the same letter(s) are not significantly different at $P < 0.05$ based on the Bonferroni test

B3, B5, and B6 contents in 10 black soybean seeds were affected by genotypic factors alongside the differences in cotyledon color. In the present study, the differences in vitamins B1, B2, and B6 contents between the varieties were significant, whereas vitamin B3 content was not significantly different (Table 1). The contents of vitamins B1, B2, B3, and B6 ranged from 3.5 (SP) to 6.4 (SO), 0.8 (CHO) to 1.1 (PSN), 55.2 (DC) to 60.0 (PSN), and 2.9 (SP) to 4.2 $\mu\text{g/g}$ (SCJ), respectively. The year effect of each location across varieties was observed in all four B vitamins, with the exception of vitamin B2 in Dalseong (Table 2). In Suwon, vitamin B1 was the highest in 2017, whereas vitamins B2, B3, and B6 were the lowest in 2017. For the Iksan-grown samples, the contents of vitamins B1 and B2 were the highest in 2019, whereas those of vitamins B3 and B6 were the highest in 2018. In Dalseong,

the contents of vitamins B1 and B6 were the highest in 2019, whereas that of vitamin B3 was the highest in 2018.

Isoflavone contents

Soybean seeds contain 12 isoflavone components, which are composed of aglycones (i.e., daidzein, genistein, and glycitein), β -glucosides (i.e., daidzin, genistin, and glycitin), malonyl- β -glucosides (i.e., 6''-O-malonyldaidzin, 6''-O-malonylgenistin, 6''-O-malonylglycitin), and acetyl- β -glucosides (i.e., 6''-O-acetyldaidzin, 6''-O-acetylgenistin, 6''-O-acetylglycitin). In this study, only aglycones and β -glucosides were identified. Six individual and total isoflavone contents in the 13 Korean soybean varieties across three locations (Suwon, Iksan, and Dalseong) over 3 years are presented in Table 3. The results for each of the nine environments are presented in Additional file:

Tables S7–S9. Consistent with previous studies [26, 27], significant differences in total and individual isoflavone concentrations were observed according to genotype, site, and year. The contents of daidzein, genistein, and glycitein ranged from 5.6 (SO) to 17.0 (PSN), 2.6 (SCJ) to 6.7 (PW), and 1.0 (MS) to 10.1 (SCJ), respectively. The levels of daidzin, genistin, and glycitin ranged from 280.5 (SO) to 1356.9 (UR), 584.5 (SO) to 1645.8 (DP-2), and 36.4 (SO) to 168.1 $\mu\text{g/g}$ (DP-2), respectively. The total isoflavone contents in soybeans ranged from 913.3 (SO) to 3084.2 $\mu\text{g/g}$ (UR) across the environments.

In addition, the effect of year at each growth location on isoflavone biosynthesis was significant (Table 4). The contents of daidzin, genistin, and total isoflavone were significantly higher in 2018 than in other years in all locations (Table 4). The content of glycitein was not different according to year in each location, whereas that of glycitin was the lowest in 2017 compared to other years at Iksan and Dalseong. In previous studies, lower isoflavone concentrations were generally observed in early maturing cultivars than in late-maturing soybean cultivars [28, 29]. In our study, the varieties CHO and SO, which belong to the early maturing ecotype, and SCJ and CJ-3, which have a black seed coat, tended to have lower isoflavone concentrations (Table 3). It is known that soybean seed isoflavone concentration and composition are influenced by temperature during seed development [19, 25]. When the R6 growth stage plants grown under intermediate night/daytime temperatures of 18/28°C were subjected to either intermediate (18/28°C), low (13/23°C), or high (23/33°C) temperature conditions, the decrease in temperature significantly increased the isoflavone concentrations [25]. Lower isoflavone content in early maturing varieties might be attributable to higher temperatures during seed maturation than in mid-late- and late-maturing varieties

(Table 1). Some studies have investigated the relationship between isoflavone content and seed coat color; for instance, high total isoflavone content in black soybeans [29, 41]. In contrast, Lee et al. [42] found less isoflavones in black soybeans than in yellow soybeans. These authors [27, 29, 42] suggested that there is no consensus regarding the relationship between isoflavone content and seed coat color.

Despite significant environmental effects on isoflavone concentration, varieties with consistently high and low isoflavone concentrations across environments were observed by Seguin et al. [26], who carried out an investigation on 20 cultivars grown in replicated trials at two sites in Montreal, Canada, in 2002/2003. Similarly, in our study, the ranking of some varieties with the highest and lowest total isoflavone concentrations was relatively stable across locations and years. DP-2, PW, SP, and UR consistently had the highest total isoflavone concentrations at each location per year (Additional file: Tables S7–S9). In contrast, SO and TG consistently had the lowest total isoflavone concentrations in each of the nine locations. These results revealed the existence of genetic differences in total isoflavone concentrations.

Chemometric analyses

Principal component analysis (PCA) and partial least squares-discriminant analysis (PLS-DA) have been used as the most common chemometric tools for extracting information from any multivariate data of a biological system [43]. When unsupervised PCA analysis was performed to visualize for the separation among varieties, locations, and years of compositional data, the t1 and t2 accounted for 26% and 18% of the total variance, respectively (Fig. 1). Each point represents a particular sample. Large variances among samples of the same variety were

Table 4 Contents of isoflavones in seeds of 13 soybean varieties in each environment across varieties

Variety	Daidzein	Daidzin	Genistein	Genistin	Glycitein	Glycitin	Total
Suwon							
2017	7.1 ± 2.1 ^b	648.9 ± 267.0 ^b	3.2 ± 1.3 ^b	957.1 ± 287.8 ^b	3.3 ± 2.6 ^a	86.5 ± 37.1 ^a	1706.1 ± 581.8 ^b
2018	9.2 ± 4.2 ^a	1027.0 ± 427.6 ^a	5.0 ± 2.3 ^a	1453.9 ± 522.6 ^a	2.9 ± 2.6 ^a	108.9 ± 47.3 ^a	2606.8 ± 950.4 ^a
2019	6.3 ± 2.7 ^b	830.0 ± 346.1 ^b	3.5 ± 1.3 ^b	1144.6 ± 376.3 ^b	3.6 ± 2.7 ^a	107.2 ± 51.4 ^a	2095.6 ± 756.1 ^b
P-value	0.0017	< 0.0001	0.0003	< 0.0001	0.5698	0.0587	< 0.0001
Iksan							
2017	31.8 ± 20.4 ^a	681.3 ± 268.3 ^b	6.2 ± 1.3 ^a	1034.8 ± 327.2 ^b	3.4 ± 2.4 ^a	54.5 ± 30.4 ^b	1812.0 ± 607.2 ^b
2018	13.5 ± 13.5 ^b	1107.9 ± 360.8 ^a	5.3 ± 1.5 ^b	1544.7 ± 366.4 ^a	3.7 ± 2.7 ^a	97.9 ± 51.6 ^a	2772.9 ± 752.5 ^a
2019	10.7 ± 6.7 ^b	742.6 ± 268.5 ^b	4.3 ± 1.5 ^c	998.0 ± 311.2 ^b	4.0 ± 2.8 ^a	80.7 ± 45.3 ^a	1857.4 ± 589.7 ^b
P-value	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.5251	0.0001	< 0.0001
Dalseong							
2017	10.1 ± 3.8 ^a	697.1 ± 314.8 ^{ab}	5.4 ± 2.9 ^a	976.6 ± 336.4 ^{ab}	5.6 ± 5.0 ^a	66.0 ± 26.9 ^b	1760.9 ± 656.7 ^{ab}
2018	9.0 ± 2.5 ^{ab}	799.8 ± 322.3 ^a	4.4 ± 1.2 ^{ab}	1141.9 ± 379.3 ^a	4.3 ± 3.6 ^a	94.0 ± 38.3 ^b	2053.4 ± 698.5 ^a
2019	8.1 ± 1.9 ^b	685.1 ± 451.8 ^b	3.7 ± 1.2 ^b	915.6 ± 309.5 ^b	4.1 ± 3.2 ^a	115.4 ± 51.7 ^a	1662.0 ± 581.5 ^b
P-value	0.0109	0.0245	0.0007	0.0126	0.1974	< 0.0001	0.024

Data are presented as the mean and standard deviation ($\mu\text{g/g}$ dry weight basis). Means in the same column followed by the same letter(s) are not significantly different at $P < 0.05$ based on the Bonferroni test

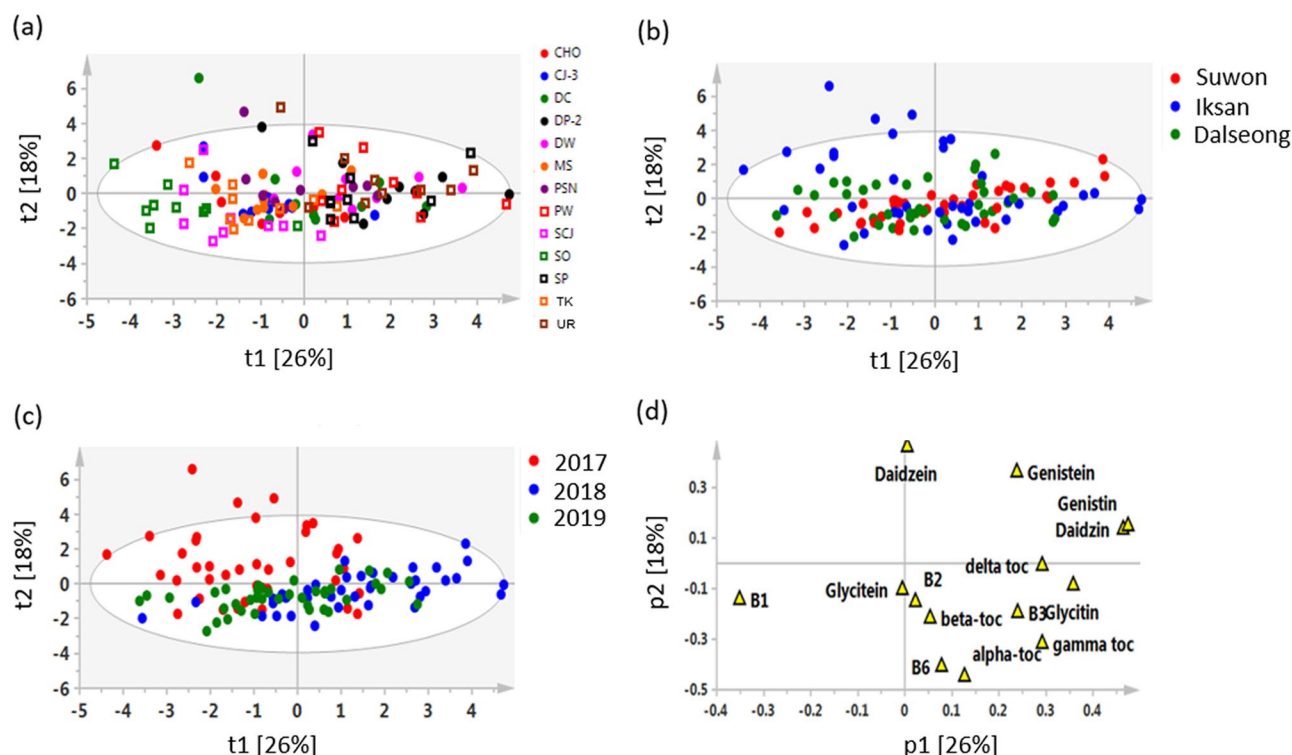


Fig. 1 Score and loading plots of principal components 1 and 2 of the principal component analysis (PCA) generated from tocopherols, B vitamins, and isoflavones. PCA score plots colored according to variety (a), location (b), and year (c)

clearly observed on the PCA score plot (Fig. 1a), indicating that there are considerable environmental effects on the composition. Notably, SO and SP were separated on the PC1. The loading plot of the corresponding PCA indicated that discrimination of SO and SP could be attributed in part to differences in levels vitamin B1, daidzin, and genistin (Fig. 1d). The PCA was in agreement with the levels of vitamin B1, daidzin and genistin in SO and SP (Tables 1 and 3). There was no clear separation among the three locations, with exception of that some data of the Iksan-grown samples were differentiated from the Suwon- and Dalseong-grown samples (Fig. 1b). Year 2017 could be separated from the year 2018 and 2019 on the t2 (Fig. 1c).

Since PCA is an unsupervised method, not taking into account varieties, locations, or years in the definition of the components, PLS-DA, a supervised classification method was utilized (Fig. 2). The R^2 and Q^2 parameters of the PLS-DA model were used to measure the goodness of fit and predictive ability of the model, respectively. These values ranged from a minimum of zero to a maximum of one. The model fits and predicts better if their values are close to 1.0, and a model with $Q^2 > 0.5$ is considered to have good predictive capacity [36]. The score plot of PLS-DA according to variety ($R^2X=0.578$, $R^2Y=0.267$, $Q^2=0.183$) showed some differences among the varieties on t1 and t2, although some varieties were

not significantly different (Fig. 2a). The two components t1 and t2 accounted for 25.2% and 10.6% of the total model variance, respectively. As already indicated by the PCA, Variable Importance in Projection (VIP) score was obtained from the PLS-DA and then used to identify potential metabolites for discrimination. Variables with a VIP score > 1 were considered more important for classification. β -tocopherol, glycitein, δ -tocopherol, vitamin B1, and glycitin contributed to the separation of varieties. The score plot of PLS-DA by location ($R^2X=0.564$, $R^2Y=0.297$, and $Q^2=0.09$) showed that there were no apparent differences among the three locations (Fig. 2b). As Daidzein, vitamin B2, and genistein contributed to the differences between the locations (Fig. 2b). The score plot of PLS-DA by year ($R^2X=0.583$, $R^2Y=0.588$, $Q^2=0.502$) showed some differences among the 3 years (Fig. 2c). Data from soybeans grown in 2017 were different from data from 2018 to 2019. Vitamins B6 and B3 were the most significant components in the PLS-DA model for the separation of data according to cultivation year (Fig. 2c).

The parameters Q^2 of the PLS-DA for variety (Fig. 2a) and location (Fig. 2b) were 0.183 and 0.09, respectively, suggesting the poor predictive abilities of the models. The Q^2 value of the PLS-DA for year (Fig. 2c) was > 0.5 , indicating a good predictability of the model. Q^2 values strongly depend on the properties of a dataset, such as

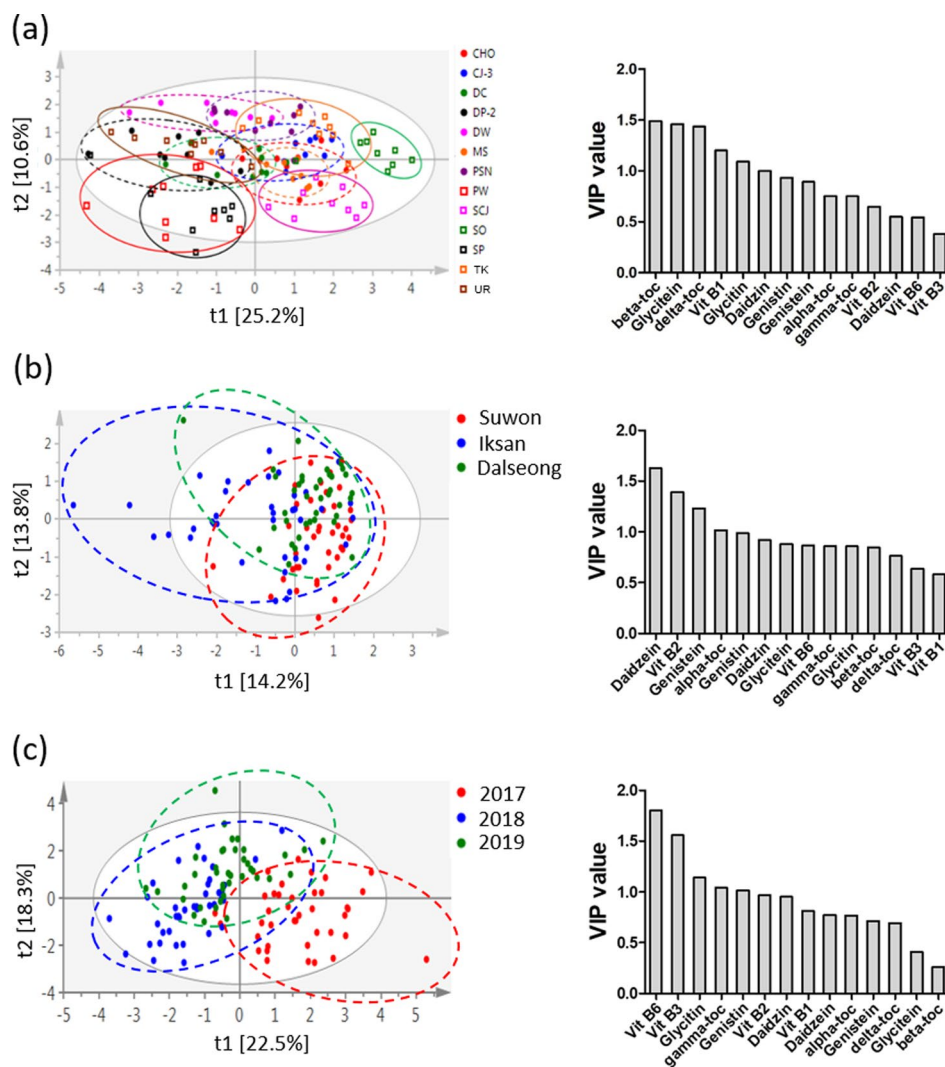


Fig. 2 Latent structure discrimination analysis (PLS-DA) score plots and variable importance in the projection (VIP) score plots of tocopherol, B vitamins, and isoflavone. Compositional data were subjected to PLS-DA according to variety (a), location (b), and year (c)

the number of observations. It has been shown that models with poor predictability are frequently validated by a permutation test when the predictive ability of the original model is greater than that of any model with random permutations of y variables [37, 44]. Results of the permutation tests for the three models are shown in Additional file: Fig. S1. All the permuted R^2 and Q^2 values were smaller than the original values of their models, and permuted Q^2 value, the intercept on the y -axis was negative. This suggests that the models were acceptable.

Analysis of variance using permutational univariate analysis (PERMANOVA)

PERMANOVA, a nonparametric analysis of variance, can partition variation directly among individual terms in multifactorial ANOVA model [45]. In this study, a PERMANOVA based on 999 permutations using the Euclidean distance and partitioning was done using Type

III sum of squares for each variable (genotype, location, year, and respective interactions) to determine the contributions of variables to the component composition (Table 5). The component of variation in PERMANOVA (COV) is a value that indicates the degree of influence of each factor. A higher COV indicates a greater influence of a specific factor or interaction effect [38]. Table 5 summarizes the PERMANOVA results of Pseudo- F , p (perm), and COV values.

Tocopherol concentrations are mainly influenced mainly by genetic and environmental factors such as temperature during seed filling and soil moisture [13, 18]. The relative contribution of these factors of variation to tocopherol composition is quite contradictory in previous studies; Whent et al. (2009) reported that α - (57%), γ - (70%), δ - (43%), and total tocopherol (69%) contents were most affected by genotype. The second most important source of variation for individual tocopherol was

Table 5 Results of the PERMANOVA for tocopherols, B vitamins, and isoflavone contents in seeds of 13 soybean cultivars grown at three locations for 3 consecutive years

Parameters	V (Variety) df = 12		L (Location) df = 2		Y (Years) df = 2		V x L df = 24		V x Y df = 24		L x Y df = 4		V x L x Y df = 48	
	Pseudo-F	COV	Pseudo-F	COV	Pseudo-F	COV	Pseudo-F	COV	Pseudo-F	COV	Pseudo-F	COV	Pseudo-F	COV
α-Toc	47.81**	34.78	65.28**	11.02	292.02**	49.90	4.09**	6.87	7.52**	14.54	385.19**	197.62	4.48**	23.30
β-Toc	246.1**	59.85	78.94**	4.39	28.17**	1.53	25.27**	3.81	25.27**	10.45	130.37**	21.87	9.11**	17.83
δ-Toc	218.1**	26.62	74.51**	6.82	178.8**	5.03	7.50**	2.39	12.99**	4.41	74.47**	6.24	11.98**	3.31
γ-Toc	34.85**	13.60	74.51**	6.82	201.74**	18.62	8.03**	8.48	6.63**	6.79	277.33**	76.91	9.85**	32.03
B1	157.5**	15.31	5.72*	0.11	292.86**	6.59	7.65**	1.95	13.08**	3.55	123.27**	8.28	9.15**	7.17
B2	6.32**	3.17	35.86**	4.79	78.40**	10.63	6.15**	9.20	2.44**	2.58	31.95**	12.75	4.71**	19.86
B3	3.30**	0.24	4.29*	0.08	125.6**	2.99	2.47**	0.46	2.84**	0.57	14.66*	0.98	1.54*	2.80
B6	16.30**	5.81	10.50**	0.83	561.54**	49.13	2.93**	2.20	3.62**	2.98	17.06**	4.22	2.81**	10.25
daidzein	1186.9**	63.51	4651.4**	57.48	2096.0**	25.89	323.72**	51.85	336.03**	53.83	1732.1**	64.19	256.97**	123.38
daidzin	10.976**	112.02	2368.8**	5.58	9260.1**	21.81	157.24**	4.78	228.12**	6.95	1055.8**	7.43	108.04**	9.83
genistein	640.16**	53.15	978.55**	18.76	624.69**	11.97	71.44**	17.57	70.45**	17.33	328.2**	18.26	65.06**	47.95
genistin	388.47**	61.42	131.07**	4.76	533.84**	19.49	6.68**	2.70	12.26**	5.50	51.47**	5.54	32.83**	9.52
glycitein	285.74**	236.66	97.18**	18.29	18.30**	3.29	61.69**	150.05	17.01**	39.59	25.0**	13.69	20.72**	146.25
glycitin	8897.7**	118.74	5017.4**	15.45	9395.4**	28.93	503.68**	20.13	559.93**	22.38	1157.7**	10.69	275.74**	33.00

Pseudo-F, F value by permutation; COV, the component of variation; Toc, tocopherol; *P(perm) < 0.05; **P(perm) < 0.001; the most attributable variances for each component concentration were in bold

the environment, followed by genotype-by-environment interactions. However, Carrera et al. [13] reported that the environment accounted for most of the total variation in the concentration of α- (84%), γ- (38%), δ- (84%), and total tocopherol (41%). In our study, the L×Y effect was the most significant factor for α- (COV=197.62) and γ- (COV=76.91) tocopherols, whereas the V effect was for β- (COV=59.85) and δ- (COV=26.62) tocopherols (Table 5). Notably, tocopherol isomers with the same benzoquinol structures present in the same biosynthetic pathway are influenced by the same variance factors. The different results between studies may be due to differences in growing locations, evaluation years, and genotypes.

To date, studies on the genotypic and/or environmental factors affecting the concentrations of B vitamins are few. Our results revealed that vitamin B1 was attributable to the V effect (COV=15.31), followed by the L×Y effect (COV=8.28). Vitamin B3 (COV=2.99), and B6 (COV=49.13) contents were attributable to the year effect, whereas the vitamin B2 content was mainly affected by the V×L×Y effect (COV=19.86), followed by the V×L effect (COV=12.75) (Table 5). With regard to the isoflavone content, the V effect was the most significant for daidzin (COV=11.02), genistein (COV=53.15), genistin (COV=61.42), glycitein (COV=236.66), and glycitin (COV=118.74). However, the daidzein content was influenced by the V×L×Y effect (COV=123.38) rather than the V effect alone (COV=63.51). Our results are in agreement with those of Hoeck et al. [46] and Zhang et al. [28], who previously reported that genetic factors play the most important role in isoflavone accumulation rather than environmental factors, such as site and year or interaction effect between genetic and environmental factors [28]. These results were further supported by the identified variables that contributed to the discriminations caused by variety, location, and years in PLS-DA model: β-tocopherol, glycitein, and δ-tocopherol for variety; daidzein, vitamin B2, and genistein for location; vitamin B6 and B3 for year (Fig. 2). This shows that our evaluation of the major variables determining the contents of these compositions is reliable.

This study investigated the impact of genotypic and environmental factors on the tocopherol, B vitamin, and isoflavone contents in 13 Korean soybean varieties, with differing seed coat color, maturity durations, and food usage. Our findings revealed significant effects of both genotypic and environmental variables on these seed constituents. Utilizing the PLS-DA model, we observed a greater influence of the cultivation year on the measured components compared to variety and location. PERMANOVA analysis highlighted genetic factors as the primary sources of variations in β- and δ-tocopherols, vitamin B1, daidzin, genistein, genistin, glycitein, and

glycitin accumulation. Additionally, location and year interactions significantly impacted α - and γ -tocopherols. Thus, optimizing the growing environment becomes crucial for enhancing α - and γ -tocopherols in soybean seeds. Furthermore, cultivation year was a key determinant of vitamin B3 and B6 contents, whereas daidzein and vitamin B2 contents were influenced by genotype-environmental interactions. Notably, isoflavone accumulation was found to be lower in early maturing varieties compared to late maturing varieties. These findings contribute to a better understanding of the factors governing seed composition in soybeans, and to expand the compositional dataset of commercial soybeans for the safety assessment of genetically modified soybeans.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13765-024-00896-5>.

Supplementary Material 1

Acknowledgements

EHK was supported by a postdoctoral training program in RDA of South Korea.

Author contributions

Conceptualization: E-HK and S-WO. Methodology and validation: J-WJ, OSY, S-YL, MJK, H-MP, and S-GL. Data analysis: E-HK, J-WJ, YJ, and YJ. Writing and editing: E-HK, S-WO. Supervision: S-WO. All authors have read and approved the final manuscript.

Funding

This study was supported by the National Academy of Agricultural Science (Code PJ0160972) from the Rural Development Administration of the Republic of Korea.

Data availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Declarations

Competing interests

The authors declare no competing interests.

Received: 3 October 2023 / Accepted: 9 April 2024

Published online: 27 May 2024

References

1. Kumar V, Rani A, Chauhan GS (2010) In: Singh G (ed) Nutritional value of soybean. The soybean: Botany, production and uses. CAB International, Wallingford, UK, pp 375–403
2. Slavin M, Cheng Z, Luther M, Kenworthy W, Yu L (2009) Antioxidant properties and phenolic, isoflavone, tocopherol and carotenoid composition of Maryland-grown soybean lines with altered fatty acid profiles. Food Chem 114:20–27
3. Kim I-S, Kim C-H, Yang W-S (2021) Physiologically active molecules and functional properties of soybeans in human health—a current perspective. Int J Mol Sci 22:4054
4. Jiang G-L, Rajcan I, Zhang Y-M, Han T, Mian R (2023) Editorial: soybean molecular breeding and genetics. Front Plant Sci 14:1157632zd
5. ISAAA (2021) Pocket K No. 16: Biotech Crop Highlight in 2019. International service for the acquisition of agri-biotech applications. Updated May 2021 <http://www.isaaa.org>
6. Codex (2003) Guideline for the Conduct of Food Safety Assessment of Foods Derived from Recombinant –DNA Plants (CAC/GL 45-2003), Geneva
7. Kok EJ, Keijer J, Kleter GA, Kuiper HA (2008) Comparative safety assessment of plant-derived foods. Regul Toxicol Pharm 50:98–113
8. Harrigan GG, Glenn KC, Ridley WP (2010) Assessing the natural variability in crop composition. Regul Toxicol Pharm 58:513–520
9. Rizvi S, Raza ST, Ahmed F, Ahmad A, Abbas S, Mahdi F (2014) The role of vitamin E in human health and some disease. Sultan Qaboos Univ Med J 14:e157–e165
10. Kudou S, Fleury Y, Welti D, Magnolato D, Uehida T, Kitamura K, Okubo K (1991) Malonyl isoflavone glycosides in soybean seeds [*Glycine max* (L.) Merrill]. Agric Biol Chem 55:2227–2233
11. Carrera C, Martínez MJ, Dardanelli J, Balzarini M (2011) Environmental variation and correlation of seed components in nontransgenic soybeans: protein, oil, unsaturated fatty acids, tocopherols, and isoflavones. Crop Sci 51:800–809
12. Bellaloui N, Bruns A, Aggas HK, Mengistu A, Fisher DK, Reddy KN (2015) Agricultural practices altered soybean seed protein, oil, fatty acids, sugars, and minerals in the Midsouth USA 31:1
13. Carrera CS, Seguin P (2016) Factors affecting tocopherol concentrations in soybean seeds. J Agric Food Chem 64:9465–9474
14. Dolde D, Vlahakis C, Hazebroek J (1999) Tocopherols in breeding lines and effects of planting location, fatty acid composition, and temperature during development. J Am Oil Chem Soc 76:349–355
15. Ujiie A, Yamada T, Fujimoto K, Endo Y, Kitamura K (2005) Identification of soybean varieties with high α -tocopherol content. Breed Sci 55:123–125
16. Rani A, Kumar V, Verma SK, Shakya AK, Chauhan GS (2007) Tocopherol content and profile of soybean: genotypic variability and correlation studies. J Amer Oil Chem Soc 84:377–383
17. Ghosh S, Zhang S, Azam M, Gebregziabher BS, Abdelghany AM, Shaibu AS, Qi J, Feng Y, Agyenim-Boateng KG, Liu Y, Feng H, Li Y, Li J, Li B, Sun J (2022) Natural variation of seed tocopherol composition in diverse world soybean accessions from maturity group 0 to VI grown in China. Plants 11:206
18. Britz SJ, Kremer DF (2002) Warm temperatures or drought during seed maturation increase free α -tocopherol in seeds of soybean (*Glycine Max* [L.] Merr.). J Agric Food Chem 50:6058–6063
19. Chennupati P, Seguin P, Liu W (2011) Effects of high temperature stress at different stages on soybean isoflavone and tocopherol concentrations. J Agric Food Chem 59:13081–13088
20. Britz SJ, Kremer DF, Kenworthy WJ (2017) Tocopherols in soybean seeds: genetic variation and environmental effects in field-grown crops. J Am Oil Chem Soc 85:931–936
21. Kim G-P, Lee J, Ahn K-G, Hwang Y-S, Choi Y, Chun J, Chang W-S, Choung M-G (2014) Differential responses of B vitamins in black soybean seeds. Food Chem 153:101–108
22. Vedrına-Dragojević I, Šebečić B, Balint L (1989) Variability of thiamine, riboflavin, and niacin content in soybean seed. Die Nahrung 33:1017–1019
23. Lee SJ, Yan W, Ahn JK, Chung IM (2003) Effects of year, site, genotype and their interactions on various soybean isoflavones. Field Crops Res 81:181–192
24. Wu D, Li D, Zhao X, Zhan Y, Teng W, Qiu L, Zheng H, Li W, Han Y (2020) Identification of a candidate gene associated with isoflavone content in soybean seeds using genome-wide association and linkage mapping. Plant J 104:950–963
25. Lozovaya VV, Lygin AV, Ulanov AV, Nelson RL, Daydé J, Widholm JM (2005) Effect of temperature and soil moisture status during seed development on soybean seed isoflavone concentration and composition. Crop Sci 45:1934–1940
26. Seguin P, Zheng W, Smith DL, Deng W (2004) Isoflavone content of soybean cultivars grown in eastern Canada. J Sci Food Agric 84:1327–1332
27. Azam M, Zhang S, Abdelghany AM, Shaibu AS, Feng Y, Li Y, Tian Y, Hong H, Li B, Sun J (2020) Seed isoflavone profiling of 1168 soybean accessions from major growing ecoregions in China. Food Res 130:108957
28. Zhang J, Ge Y, Han F, Li B, Yan S, Sun J, Wang L (2014) Isoflavone content of soybean cultivars from maturity group 0 to VI grown in northern and southern China. J Am Oil Chem Soc 91:1019–1028
29. Choi Y-M, Yoon H, Lee S, Ko H-C, Shin M-J, Lee M-C, Oh S, Desta KT (2020) Comparison of isoflavone composition and content in seeds of soybean (*Glycine max* (L.) Merrill) germplasms with different seed coat colors and days to maturity. Korean J Plant Res 33:558–577

30. Kim E-H, Oh S-W, Lee S-Y, Park H-Y, Kang Y-Y, Lee G-M, Baek D-Y, Kang H-J, Park S-Y, Ryu T-H, Chung Y-S, Lee S-G (2021) Composition of the seed nutritional composition between conventional varieties and transgenic soybean overexpressing *Physaria* FAD3-1. *J Sci Food Agric* 101:2601–2613
31. Arella F, Lahély S, Bourguignon JB, Hasselmann C (1996) Liquid chromatographic determination of vitamins B1 and B2 in foods. A collaborative study. *Food Chem* 56:81–86
32. Lee S-Y, Kim E-H, Baek D-Y, Lee G-M, Park S-Y, Lee S-G, Ryu T-H, Kang H-J, Oh S-W (2020) Genotypic and environmental effects on the variation of vitamin contents in red pepper (*Capsicum annuum* L.) varieties. *J Korean Soc Food Sci Nutr* 49:244–252
33. Choi S-R, Song E-J, Song Y-E, Choi M-K, Han H-A, Lee I-S, Shin S-H, Lee K-K, Choi Y-M, Kim H-R (2017) Determination of vitamin B6 content using HPLC in agricultural products cultivated in local areas in Korea. *Korean J Food Nutr* 30:710–718
34. Park S-Y, Lee SM, Lee J-H, Ko H-S, Kwon SJ, Suh S-C, Shin K-S, Kim JK (2012) Compositional comparative analysis between insect-resistant rice (*Oryza sativa* L.) with a synthetic cry1Ac gene and its non-transgenic counterpart. *Plant Biotechnol Rep* 6:29–37
35. Zdziebło AP, Reuter WM (2015) Analysis of isoflavones in soy products by UHPLC with UV detection. Application Note: liquid chromatography. PerkinElmer (resources.perkinelmer.com)
36. Eriksson L, Johansson E, Kettaneh-Wold N, Trygg J, Wikström C, Wold S (2006) Multi- and megavariate data analysis. Part I: basic principles and application. Umetrics, Sweden
37. Westerhuis J, Hoefsloot HJ, Smit S, Vis D, Smilde A, van Velzen EJ, van Duynhoven JM, van Dorsten F (2008) Assessment of PLS-DA Cross validation. *Metabolomics* 4:81–89
38. Anderson MJ, Gorley RN, Clarke KR (2008) PERMANOVA + for PRIMER: guide to software and statistical methods. PRIMER-E, Plymouth Marine Laboratory, Plymouth, U.K., 214
39. Clark KR, Gorley RN (2015) PRIMER v7: user Manual/Tutorial. PRIMER-E, Plymouth, U.K., p 296
40. Shewry PR, Schaik FV, Ravel C, Charmet G, Rakszegi M, Bedi Z, Ward JL (2011) Genotype and environment effects on the contents of vitamins B1, B2, B3, and B6 in wheat grain. *J Agric Food Chem* 59:10564–10571
41. Kim JA, Hong SB, Jung WS, Yu CY, Ma KH, Gwang JG, Chung IM (2007) Comparison of isoflavones composition in seed, embryo, cotyledon and seed coat of cooked-with-rice and vegetable soybean (*Glycine max* L.) varieties. *Food Chem* 102:738–744
42. Lee SJ, Seguin P, Kim JJ, Moon HI, Ro HM, Kim EH, Seo SH, Kang EY, Ahn JK, Chung IM (2010) Isoflavones in Korean soybeans differing in seed coat and cotyledon color. *J Food Compos Anal* 23:160–165
43. Ramadan Z, Jacobs D, Grigorov M, Kochhar S (2006) Metabolic profiling using principal component analysis, discriminant partial least squares, and genetic algorithms. *Talanta* 68:1683–1691
44. Triba MN, Le Moyec L, Amathieu R, Goossens C, Bouchemal N, Nahon P, Rutledge DN, Savain P (2015) PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters. *Mol Biosyst* 11:13
45. Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecol* 26:32–46
46. Hoeck JA, Fehr WR, Murphy PA, Welke GA (2000) Influence of genotype and environment on isoflavone contents of soybean. *Crop Sci* 40:48–51

Publisher's Note

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