Prev Nutr Food Sci
Vol 17, p 184~191 (2012)
http://dx.doi.org/10.3746/pnf.2012.17.3.184

Chemometric Approach to Fatty Acid Profiles in Soybean Cultivars by Principal Component Analysis (PCA)

Eui-Cheol Shin[1], Chung Eun Hwang[1], Byong Won Lee[2], Hyun Tae Kim[2], Jong Min Ko[2], In Youl Baek[2], Yang-Bong Lee[3], Jin Sang Choi[1], Eun Ju Cho[4], Weon Taek Seo[1], and Kye Man Cho[1†]

[1]Department of Food Science & Institute of Fusion Biotechnology, Gyeongnam National University of Science and Technology, Gyeongnam 660-758, Korea
[2]Department of Functional Crop, National Institute of Crop Science (NICS), Rural Development Administration (RDA), Gyeongnam 627-803, Korea
[3]Department of Food Science and Technology, Pukyong National University, Busan 608-737, Korea
[4]Department of Food Science and Nutrition, Pusan National University, Busan 609-735, Korea

Abstract

The purpose of this study was to investigate the fatty acid profiles in 18 soybean cultivars grown in Korea. A total of eleven fatty acids were identified in the sample set, which was comprised of myristic (C14:0), palmitic (C16:0), palmitoleic (C16:1, ω7), stearic (C18:0), oleic (C18:1, ω9), linoleic (C18:2, ω6), linolenic (C18:3, ω3), arachidic (C20:0), gondoic (C20:1, ω9), behenic (C22:0), and lignoceric (C24:0) acids by gas-liquid chromatography with flame ionization detector (GC-FID). Based on their color, yellow-, black-, brown-, and green-colored cultivars were denoted. Correlation coefficients (r) between the nine major fatty acids identified (two trace fatty acids, myristic and palmitoleic, were not included in the study) were generated and revealed an inverse association between oleic and linoleic acids (r=-0.94, $p<0.05$), while stearic acid was positively correlated to arachidic acid (r=0.72, $p<0.05$). Principal component analysis (PCA) of the fatty acid data yielded four significant principal components (PCs; i.e., eigenvalues>1), which together account for 81.49% of the total variance in the data set; with PC1 contributing 28.16% of the total. Eigen analysis of the correlation matrix loadings of the four significant PCs revealed that PC1 was mainly contributed to by oleic, linoleic, and gondoic acids, PC2 by stearic, linolenic and arachidic acids, PC3 by behenic and lignoceric acids, and PC4 by palmitic acid. The score plots generated between PC1-PC2 and PC3-PC4 segregated soybean cultivars based on fatty acid composition.

Key words: soybean, cultivar, fatty acid, linoleic acid, PCA

INTRODUCTION

Soybean is one of the major oilseed crops of the world by providing an important nutrient source to the world's population. Soybeans have received considerable attention because their lipid profile is well balanced in the types of unsaturated fatty acids (1,2). Well-balanced lipids in food sources can provide energy to the body, as well as provide essential fatty acids, such as linoleic acid. However, despite linoleic acids' nutritional and functional properties, the presence of two isolated double bonds makes it susceptible to oxidation or degradation during processing. To improve nutritional attributes and oxidative stability of such oils in oilseed crops, plant breeders have successfully modified the fatty acid profile in oilseed crops, creating new lines with an increased content of oleic acid at the expense of linoleic acid (2,3). Liu and White (4) denoted the oxidative stability of soybean oils with altered fatty acid compositions and they found linoleic levels correlated highly with the peroxide value (r=0.95, p=0.01).

Chemometrics is the combination study of mathematical, statistical and other logic-based approaches to efficiently manage and interpret chemically-derived data (5). Principal component analysis (PCA) is a popular multivariate modeling and analysis method commonly used in chemometric studies. PCA provides a reduced number of variables in a large data set and defines a limited number of principal components that describe independent variation structures in the data. When more than three variables have been measured, visualization of the data by various plotting systems is then possible (6). Therefore, PCA shows relationships among groups of variables in a data set and shows relationships that exist between objects.

PCA has been used effectively to define relationships that exist in fatty acid characterization studies of food lipids due to its ability to manage and interpret large

†Corresponding author. E-mail: kmcho@gntech.ac.kr
Phone: +82-55-751-3272, Fax: +82-55-751-3279

data sets (7-9). For example, almond cultivars have been classified in several studies using chemometric techniques (10). García-López et al. (10) used cluster analysis of the major fatty acid composition in almond lipids to classify the cultivars into three groups. In a later study, the three groups of almond cultivars were further subdivided by including minor fatty acids in the multivariate analysis (11). In recent study, Sathe et al. (12) applied cluster analysis to show that fatty acid composition of California grown almonds is influenced by cultivar, location of growth and crop year.

Although soybean oil has been included in some chemometric studies comparing vegetable oils, soybean cultivars have yet to be extensively classified using multivariate techniques (13-15). This study proposes and describes an analytical method to detect, identify and quantify fatty acid profiles in soybeans, applying GC with FID detection associated with PCA, a powerful chemometric approach. We conducted a study which re-examined and updated compositional information of soybeans, including the levels of key bio-actives, grown in Korea. To ensure the most accurate data, an intensive sampling program was designed and implemented by the industry to provide cultivars in current production by accepted agricultural practices employed in Korea. Findings of the research will be used for the nutrient database of various soybean cultivars. This paper reports on the employment of PCA, a multivariate statistical method, to discriminate and classify field-grown soybean cultivars based upon their fatty acid profiles. Although soybean lipids have been included in some chemometric studies comparing vegetable oil fatty acid compositions, this paper provides more and advanced information concerning the application of PCA to the fatty acid profiles of a distinct sample set of only the most predominant commercial soybeans. Utilizing PCA will effectively reduce the number of variables needed to classify soybean cultivars, and, in this manner, will permit soybean researchers (e.g., breeders, geneticists) to more easily develop significant relationships between important soybean characteristics.

MATERIALS AND METHODS

Materials

Soybean samples (n=18), comprising 9 yellow (*Saedanbaek*, *Daewon*, *Daepung*, *Neulchan*, *Taekwang*, *Sunyu*, *Whanggeum*, *Daemang*, & *Pungsannamul*), 6 black (*Ilpumblack*, *Black #5*, *Chungja*, *Chungja #2*, *Chungja #3*, & *Dawon*), 2 brown (*Galmi* & *Galchae*), and 1 green, (*Nokchae*) were harvested in 2011 and were provided by the National Institute of Crop Science (NICS) of the Rural Development Administration (RDA) in Miryang, Korea. After their arrival at Gyeongnam National University of Science and Technology, the soybeans were packaged in labeled vacuum pouches to prevent their degradation. The vacuum-packaged soybean samples were stored at -40°C until analyzed.

Chemicals

HPLC-grade methanol (CH_3OH), chloroform ($CHCl_3$), isooctane and ACS-grade boron trifluoride in methanol (BF_3), anhydrous sodium sulfate, and sodium chloride were purchased from Fisher Scientific Company LLC (Suwanee, GA, USA). Undecanoic acid (98% purity), and a variety of lipid standard mixture (37 FAMEs) were acquired from Sigma Chemical Company (St. Louis, MO, USA).

Lipid extraction

Raw soybeans were removed from the freezer and allowed to reach room temperature. Total lipids were extracted according to the classical Bligh-Dyer method (16) with slight modifications. For each sample, 20 g of soybeans were finely ground in a commercial coffee mill. Five grams of the ground sample were accurately weighed and transferred to a 250-mL Erlenmeyer flask. Twenty milliliters of deionized H_2O, 50 mL of CH_3OH, 25 mL of $CHCl_3$, and ~10 mg of hydroquinone (as antioxidant) were added, and the contents were blended at 5,400 rpm for 2 min using a Polytron®-type homogenizer (Pro Scientific INC, Monroe, CT, USA). After the initial blending, an additional 25 mL of $CHCl_3$ were added, and homogenization was repeated for 1 min. Next, 25 mL of deionized H_2O and 35 mL of $CHCl_3$ were added; the mixture was again blended for 1 min. The slurry was filtered through Whatman No. 1 filter paper (GE Healthcare Life Sci., Piscataway, NJ, USA) using a Büchner funnel under slight vacuum. Approximately 1 g of NaCl crystals was added to the filtrate to facilitate phase separation. The filtrate was quantitatively transferred to a 250-mL separatory funnel and allowed to stand overnight to permit complete separation of the layers. The $CHCl_3$ phase was passed through Whatman No. 1 filter paper containing anhydrous Na_2SO_4 to eliminate moisture and collected in a 100-mL round bottom flask. The $CHCl_3$ was removed under vacuum at <40°C using a Rotavapor/Heating Bath (Model R-210 and B-491, respectively; Büchi Corporation, New Castle, DE, USA). The resultant lipid was transferred *via* a Pasteur pipette to an amber-colored vial. The round bottom flask was rinsed with a small portion of $CHCl_3$ to ensure a quantitative transfer of the extracted lipids. Residual $CHCl_3$ was removed from the vial using a stream of nitrogen. Samples were stored under a N_2-headspace at -80°C until further analyzed.

Fatty acid methylation

The extracted soybean lipids were used for fatty acid analysis. Fatty acid methyl esters (FAMEs) were prepared according to the Ngeh-Nawainbi's method with slight modifications (17). Briefly, extracted lipids (~25 mg) were transferred to a Reacti-vial™ small reaction vial (5 mL, Thermo Fisher Scientific, Rockford, IL, USA), and the mass was accurately weighed. Undecanoic acid was employed as the internal standard (IS) for this work. One milliliter of IS (*i.e.*, undecanoic acid 1 mg/mL in isooctane) was added to each Reacti-vial™. Soybean lipids were mixed with 0.5 N NaOH in methanol and flushed with a stream of nitrogen gas. The mixtures were heated at 100°C for 5 min. After cooling to room temperature, 2 mL of the 14% boron trifluoride in methanol (BF_3/MeOH), along with a Reacti-vial™ magnetic stirrer, were added to each vial, which was tightly capped, vortexed for 1 min, and placed in Reacti-Block™ B-1 aluminum block within a Reacti-Therm III™ Heating/Stirring Module (Thermo Fisher Scientific, Rockford, IL, USA) at 100°C for 30 min. After derivatization, samples were removed and allowed to cool to room temperature. Next, 5 mL of saturated NaCl solution was added to each reaction vial; the solution was vortexed for 30 s and fatty acid methyl esters (FAMEs) were extracted with 1.5 mL of isooctane. The isooctane layers were combined in a test tube. The isooctane layer was transferred to 2-mL wide-opening crimp top vials. Vials were capped with 11-mm silver aluminum caps, clear PTFE/red rubber septa, and then crimped with a crimper.

Gas chromatographic analysis

An Agilent Technologies 6890N Network gas chromatograph system with FID detector (Agilent Technologeis Inc., Wilmington, DE, USA) was employed for fatty acid analysis profiling. Operating conditions were as follows: the column was a SP-2560 capillary column (100 m×0.25 mm *i.d.*, 0.25-μm film thickness; Sigma-Aldrich Co., St. Louis, MO, USA); ultra-high purity nitrogen was the carrier gas at a flow rate of 1 mL/min and analyses were performed in constant flow mode; a split liner with glass wool was installed in the injector; the injector temperature was set at 220°C for split injection at a split ratio of 10:1; the FID temperature was set at 240°C; ultra-high purity hydrogen and scientific-grade air were the fuel gases for the FID and set at 40 mL/min and 450 mL/min, respectively; the initial oven temperature was set at 140°C and held for 5 min before ramping up at 4°C/min to 230°C; this temperature was maintained for an additional 35 min. Analyses were preformed in triplicate.

Identification of fatty acids

A Supelco 37 FAME reference standard was used to identify and quantity individual FAMEs from soybean samples. A relative response factor was calculated for each FAME using methyl undecanoate as an IS. Each FAME has a different response to the FID, depending on chain length, saturation, and *cis/trans* configuration (17):

$$R_i = (Ps_i \times Ws_{C11:0}) \ / \ (Ps_{C11:0} \times Ws_{is})$$

where R_i = relative response factor for fatty acid *i*; Ps_i = peak area of individual FAME *i* in FAMEs standard solution; $Ws_{C11:0}$ = mg of C11:0 FAME in injected FAMEs standard solution; $Ps_{C11:0}$ = peak area of C11:0 FAME in FAMEs standard solution; and Ws_{is} = mg of individual FAMEs *i* in injected FAMEs standard solution.

Oil characteristics

The ratio of oleic to linoleic (O/L) acid, iodine value (IV), the ratio of unsaturated to saturated fatty acids (U/S), and the percentage of saturation (% saturation) were calculated from GC fatty acid determinations according to the following formulae (18):

O/L ratio = % oleic acid / % linoleic acid

IV = (0.8601 × % oleic acid) + (1.7321 × % linoleic acid) + (2.616 × % linolenic acid) + (0.7854 × % gondoic acid)

U/S = (% oleic + linoleic + linolenic + gondoic acids) / (% palmitic + stearic + arachidic + behenic + lignoceric acids)

% Saturation = (% palmitic + stearic + arachidic + behenic + lignoceric acids) / (% palmitic + stearic + oleic + linoleic + linolenic + arachidic + gondoic + behenic + lignoceric acids)

Principal component analysis (PCA)

PCA was used for the classification and discrimination of soybeans. PCA is a multi-variate statistical method that entails data reconstruction and reduction. PCA generates a set of new orthogonal axes or variables known as principal components (PCs), which are different from, but representative of, the original variables. The data sets presented on the orthogonal axes are uncorrelated with one another, and express much of the total variability in the data set through comparison of only a few PCs (19). The maximal amount of variance in the data set and its direction are often explained by the first PC (*i.e.,* PC1). Each PC is defined by a vector known as the eigenvector of the variance-covariance matrix. The variance along the vector is known as the eigenvalue. Each eigenvalue, the amount of variance that is explained by a given component, was used for the determination of variances of the major PCs. The loadings (or scores) corresponding to the PCs were calculated from the correlation matrix (20). The arithmetic value of each PC is determined by followed equation (21):

$$PC = a_1 \frac{(x_1 - \bar{x}_1)}{SD_1} + \cdots$$

where x_1 are measurements of the original variables, $\bar{x}_1$ are mean values for the corresponding variables, SD_1 are standard deviations for the corresponding variables, and a_1 are loadings of the linear transformation.

Each loading of variables was used for the contribution of the original variable to the PC. Variance reduction was achieved by neglecting the unimportant directions in which samples' variance are insignificant. The important variables are along several significant directions, and the number of these directions approximates the dimensionality of the sample set. For a visualization of the data discrimination, PCA plots mapped variables and samples through dimensional spaces determined by PCs with eigenvalues greater than 1.0, based on Kaiser's rule (22). The plots depicted the identification and correlation of variables and objectives. Analysis of fatty acid composition in soybean samples was performed in triplicate. Results are expressed as the mean ± standard deviation (SD).

RESULTS AND DISCUSSION

Fatty acid profiles

A representative GC chromatogram of the FAMEs from soybean lipid extracts is depicted in Fig. 1. A total of 11 fatty acids containing 2 trace (>0.1% weight) fatty acids were identified in this study by retention time mapping with external standards and quantified relative to an internal standard (*i.e.*, methyl undecanoate, the FAME resulting from the derivatization of C11:0). Response factors of individual fatty acids were shown in Table 1 for the method validation. Table 2 reports the fatty acid compositions of lipid extracts from 18 soybean cultivars. Palmitic acid (C16:0) ranged from 9.18% to 13.02%; stearic acid (C18:0), 2.80% to 4.74%; oleic acid (C18:1, ω9), 16.52% to 33.80%; linoleic acid (C18:2, ω6), 44.72% to 60.34%; linolenic acid (C18:3, ω3), 5.62 % to 9.54%; arachidic acid (C20:0), 0.26% to 0.48%; gondoic acid (C20:1, ω9), 0.22% to 0.41%; behenic acid (C22:0), 0.27% to 0.59%; and lignoceric acid (C24:0), 0.00% to 0.39%. These results correspond with those from other studies and show that the sum of oleic and linoleic acids accounts for almost 80% of the total fatty acids detected in soybean samples (1,3,23). The oleic to linoleic acid (O/L) ratio is a quality index employed for the determination of genetic soybean characteristics and it ranges from 0.29 to 0.82. In other lipid characteristics, IV, a degree of unsaturation, ranged from 123.71 to 139.78; U/S, 4.60 to 5.91; % saturation, 14.39 to 17.69. Pearson correlation coefficients between fatty acid variables are given in Table 3. The correlation between oleic acid and linoleic acid was strong but negative (r=-0.94, $p<0.05$); that is, an increase in one fatty acid leads to a corresponding decrease in the other. This negative relationship originates from the biochemical pathways of soybean development: in the soybean germplasm, palmitoyl CoA is elongated to stearoyl CoA followed by desaturation forming oleic acid. Then, the action of oleoyl-phosphatidylcholine desaturase (a Δ^{12}-fatty acid desaturase) synthesizes linoleic acid from oleic acid (24). In addition, stearic acid is positively correlated to arachidic acid (r=0.72, $p<0.05$), while oleic acid is negatively correlated to linolenic acid (r=-0.50, $p<0.05$).

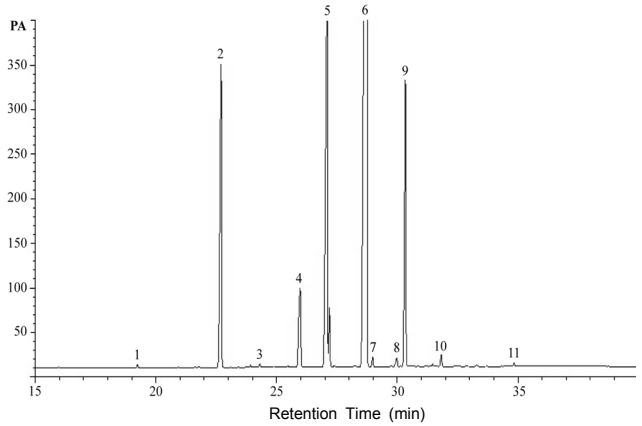

Fig. 1. Representative GC chromatogram of fatty acid methyl esters (FAMEs) from the lipid extract of soybean cultivars (peak 1, myristic acid (C14:0); 2, palmitic acid (C16:0); 3, palmitoleic acid (C16:1, ω7); 4, stearic acid (C18:0); 5, oleic acid (C18:1, ω9); 6, linoleic acid (C18:2, ω6); 7, arachidic acid (C20:0); 8, gondoic acid (C20:1, ω9); 9, linolenic acid (C18:3, ω3); 10, behenic acid (C22:0); and 11, lignoceric acid (C24:0).

Table 1. Response factor of individual fatty acids

Fatty acid	Response factor
C14:0	1.18
C16:0	1.28
C16:1 ω7	1.20
C18:0	1.36
C18:1 ω9	1.36
C18:2 ω6	1.35
C18:3 ω3	1.42
C20:0	1.35
C20:1 ω9	1.39
C22:0	1.46
C24:0	1.46

Principal component of analysis (PCA)

The data matrix of the predominant 9 fatty acids in the soybean samples analyzed was subjected to PCA to decrease the number of descriptors associated with the data set, while still explaining the maximum amount of variability present in the data. Table 4 shows the most significant PCs generated from the soybean fatty acid

Table 2. Fatty acid composition and chemical characteristics of lipids extracted from soybeans[1)]

Sample	Fatty acid composition (%weight)											O/L[2)]	IV[3)]	U/S[4)]	%Sat.[5)]
	C14:0	C16:0	C16:1 ω7	C18:0	C18:1ω9	C18:2 ω6	C18:3 ω3	C20:0	C20:1 ω9	C22:0	C24:0				
Yellow															
Saedanbaek	tr[6)]	11.62±2.51	tr	3.31±0.75	23.23±2.10	51.41±7.54	8.92±0.99	0.39±0.07	0.41±0.07	0.59±0.11	0.17±0.02	0.52±0.04	132.60±9.32	5.24±1.03	16.13±1.47
Daewon	tr	10.48±1.78	tr	3.22±0.54	21.23±3.11	56.42±6.99	7.63±1.25	0.27±0.05	0.29±0.05	0.37±0.07	0.09±0.02	0.38±0.05	136.02±6.89	5.91±0.48	14.45±1.77
Daepung	tr	10.64±2.22	tr	3.55±0.96	16.52±2.85	60.34±8.02	8.01±1.56	0.29±0.09	0.27±0.04	0.27±0.06	0.07±0.01	0.29±0.07	139.78±7.24	5.66±0.78	14.87±0.96
Neulchan	tr	10.16±1.27	tr	3.54±0.52	22.46±2.66	55.83±7.00	7.04±1.33	0.32±0.10	0.30±0.02	0.28±0.04	0.12±0.03	0.41±0.06	134.55±5.87	5.87±0.69	14.39±0.78
Taekwang	tr	10.14±3.31	tr	3.62±0.45	33.80±2.14	44.72±6.74	6.64±2.01	0.34±0.08	0.39±0.07	0.48±0.02	0.00±0.00	0.82±0.07	124.05±6.94	5.85±0.68	14.45±0.94
Sunyu	tr	10.94±2.45	tr	2.80±0.33	27.91±1.99	48.88±5.98	8.27±1.77	0.28±0.07	0.31±0.08	0.37±0.05	0.21±0.04	0.57±0.05	130.59±7.02	5.76±0.72	14.59±1.32
Whanggeum	tr	13.02±2.27	tr	3.09±0.62	20.64±2.17	54.04±6.39	8.06±1.41	0.26±0.04	0.29±0.04	0.39±0.03	0.19±0.03	0.38±0.02	132.68±5.55	4.86±0.84	17.04±1.11
Daemang	tr	11.79±2.95	tr	3.92±0.44	26.27±2.49	49.14±5.74	7.92±0.86	0.28±0.08	0.28±0.08	0.32±0.04	0.12±0.04	0.49±0.03	128.57±6.96	5.06±1.01	16.38±1.07
Pungsannamul	tr	11.13±2.65	tr	3.62±0.68	25.72±2.11	49.03±5.03	9.54±1.29	0.29±0.04	0.27±0.05	0.35±0.02	0.13±0.03	0.51±0.04	132.05±7.23	5.45±1.11	15.49±0.99
Mean±SD		11.10±2.54		3.41±0.42	24.21±2.84	52.17±5.88	8.02±1.84	0.30±0.04	0.34±0.08	0.38±0.06	0.12±0.02	0.49±0.05	132.32±7.58	5.54±0.98	15.31±1.15
Black															
Ilpumblack	tr	10.46±1.75	tr	4.61±0.55	29.90±3.33	48.01±6.02	5.62±0.48	0.48±0.05	0.22±0.07	0.49±0.02	0.19±0.02	0.59±0.04	123.71±6.18	5.05±1.03	16.27±1.02
Black #5	tr	11.72±2.51	tr	3.64±0.61	29.32±1.93	47.13±5.61	7.21±0.74	0.28±0.04	0.27±0.04	0.38±0.03	0.13±0.03	0.57±0.06	125.88±7.25	5.16±0.95	16.09±1.32
Chungja	tr	10.63±1.44	tr	3.44±0.28	25.24±2.03	52.32±6.11	7.33±1.28	0.26±0.05	0.26±0.06	0.41±0.05	0.18±0.05	0.48±0.04	131.56±5.00	5.66±0.89	14.85±0.96
Chungja #2	tr	10.59±2.01	tr	3.31±0.55	22.54±2.41	53.43±7.03	8.66±1.54	0.27±0.06	0.29±0.07	0.49±0.02	0.39±0.06	0.39±0.02	134.78±5.36	5.58±0.68	15.06±0.82
Chungja #3	tr	12.12±1.98	tr	4.74±0.39	17.35±2.01	55.77±7.15	8.82±1.11	0.36±0.07	0.27±0.09	0.37±0.04	0.11±0.04	0.31±0.03	134.91±6.14	4.60±0.74	17.69±1.41
Dawon	tr	11.54±1.74	tr	4.59±0.42	20.44±1.07	54.66±5.84	7.64±1.56	0.37±0.07	0.28±0.08	0.36±0.03	0.09±0.02	0.43±0.04	132.38±6.53	4.91±0.71	17.01±1.27
Mean±SD		11.18±2.77		4.03±0.52	24.14±2.25	51.92±6.08	7.53±1.42	0.34±0.06	0.27±0.07	0.41±0.04	0.18±0.03	0.46±0.05	130.53±7.02	5.16±0.92	16.16±1.11
Brown															
Galmi	tr	10.42±1.65	tr	3.32±0.44	21.44±1.00	55.73±6.13	8.03±1.96	0.32±0.06	0.32±0.08	0.41±0.04	0.18±0.04	0.36±0.04	136.01±7.00	5.78±0.84	14.57±0.97
Galchae	tr	9.18±1.38	tr	4.36±0.50	24.41±1.84	53.00±5.51	7.69±0.94	0.43±0.04	0.33±0.05	0.39±0.07	0.17±0.03	0.47±0.02	133.16±6.52	5.76±0.94	14.55±1.24
Mean±SD		9.80±1.88		3.94±0.47	22.92±1.54	54.42±5.88	7.88±1.11	0.38±0.05	0.33±0.06	0.40±0.06	0.18±0.04	0.42±0.03	134.59±6.79	5.77±0.90	14.56±1.17
Green															
Nokchae	tr	11.91±2.03	tr	3.61±0.68	27.51±2.09	48.41±4.03	7.28±1.48	0.42±0.04	0.31±0.07	0.43±0.08	0.19±0.04	0.57±0.06	126.76±3.58	5.06±0.18	16.47±0.34

[1)]Data represents the mean±standard deviation (SD) of each sample assayed in triplicate.
[2)]O/L: the ratio of oleic to linoleic acids. [3)]IV: iodine value. [4)]U/S: the ratio of unsaturated to saturated fatty acids.
[5)]%Sat.: the percent of saturated to total fatty acids. [6)]tr: trace (<0.1%).

Table 3. Pearson correlation coefficients between percentage levels of nine fatty acids from lipid extracts of soybeans grown in Korea (n=18)

	C16:0	C18:0	C18:1 ω9	C18:2 ω6	C18:3 ω3	C20:0	C20:1 ω9	C22:0	C24:0
C16:0	1								
C18:0	-0.05	1							
C18:1 ω9	-0.19	-0.08	1						
C18:2 ω6	-0.07	0.02	-0.94*	1					
C18:3 ω3	0.31	-0.27	-0.50*	0.29	1				
C20:0	0.02	0.72*	0.07	-0.13	-0.32	1			
C20:1 ω9	0.03	-0.35	0.09	-0.14	0.33	-0.32	1		
C22:0	-0.02	-0.03	0.33	-0.39	0.02	0.41	0.35	1	
C24:0	-0.03	-0.24	-0.13	0.10	0.22	0.13	-0.26	0.35	1

*Symbol corresponds the significant values at the level of significance α=0.05 (two-tailed test).

Table 4. Eigen analysis of the correlation matrix loadings of the significant principal components (PCs)

Fatty acid[1)]	PC1	PC2	PC3	PC4
C16:0	-0.14[2)]	-0.06	0.20	0.59
C18:0	0.22	0.53	0.02	0.35
C18:1 ω9	0.53	-0.30	-0.16	-0.10
C18:2 ω6	-0.50	0.32	0.03	-0.12
C18:3 ω3	0.34	0.42	0.37	0.20
C20:0	-0.06	-0.51	0.03	0.34
C20:1 ω9	-0.42	-0.21	0.33	0.23
C22:0	0.30	-0.23	0.57	0.04
C24:0	-0.06	0.00	0.60	-0.54
Eigenvalue	2.53	2.07	1.54	1.18
Variance (%)	28.16	23.04	17.15	13.14
Cumulative (%)	28.16	51.20	68.35	81.49

[1)]Nine fatty acid from extracts of soybeans grown in Korea. [2)]+, positive correlation; -, negative correlation.

data and their statistical loadings in the current study. A new set of nine orthogonal variables (PCs) was generated by PCA. The first principal component (*i.e.*, PC1) had the highest eigenvalue of 2.53, and accounted for 28.16% of the variability in the data set. The second, third, and fourth PCs (PC2, PC3, and PC4) had eigenvalues of 2.07, 1.54, and 1.18, and accounted for 23.04%, 17.15% and 13.14% of the variance in the data, respectively (Note: only eigenvalues of >1.0 are considered significant descriptors of data variance, according to Kaiser's rule). The remaining five generated PCs (PC5 to PC9) yielded progressively smaller eigenvalues and did not explain significant variability in the data (<19% total). Therefore, according to Kaiser's rule, only the first four PCs were used for further study. These numbers represent significant contributions of individual fatty acid variables to the total variability explained by the generated PCs. PC1 describes 28.16% of the variance in the data set, and its loadings indicate that it has high contributions from oleic (0.53), linoleic (-0.50), and gondoic (-0.42) acid variables. PC2 showed positive loadings for stearic (0.53) and linolenic (0.42) acids and a negative loading for arachidic acid (0.42) and PC3 was most described by behenic (0.57) and lignoceric (0.60) acids. PC4 shows a positive loading for palmitic acid (0.59). Specific patterns of correlation between the variables tested can be visualized by comparing loading plots between the PCs, as shown in the four planes (I, II, III, and IV) on the PC1-PC2 and PC3-PC4 axes in Fig. 2A and 2B. The objective of a loading projection is to visualize the position of the variables with respect to one another in two-dimensional space and their corresponding correlations. Variables closest to one another and far

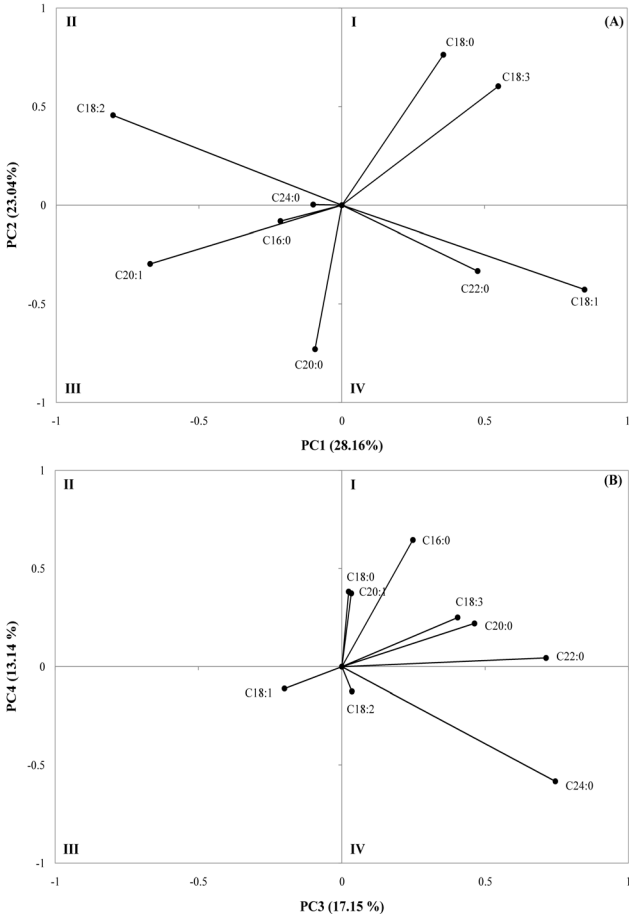

Fig. 2. Loading plots of PC1-PC2 and PC3-PC4 for fatty acid profiles in soybeans. (A) PC1-PC2 and (B) PC3-PC4.

from the plot origin are positively correlated, while variables opposite one another on the plot are negatively correlated. The two loading plots generated from the data of Table 4 can explain the relationships between two variables by their angle from the center. The correlation coefficient between two variables is defined as the cosine of the angle between their respective vectors on the plot. The cosine of 180° (*i.e.*, the angle between oleic acid and linoleic acid on the PC1-PC2 plot) is 1; therefore, they are negatively correlated. Based on this mathematical rule, uncorrelated variables occur at right angles to one another because the cosine of the angle between them is cosine 90° = 0, or not correlated. Similarly, the cosine of 0° is 1, which denotes a positive correlation between the variables (25). Fig. 3 (A and B) depicts the score plots of fatty acids in soybean cultivars (n=18) generated from comparing PC1-PC2 and PC3-PC4, and revealed four district sample groupings in the four planes (I, II, III, and IV) on the PC1−PC2 and PC3-PC4 axes. For Fig. 3(A), plane I, with the vectors of stearic and linolenic acids showing a positive correlation (see Fig. 2A-I), contains samples from the *Ilpumblack*, *Galchae*

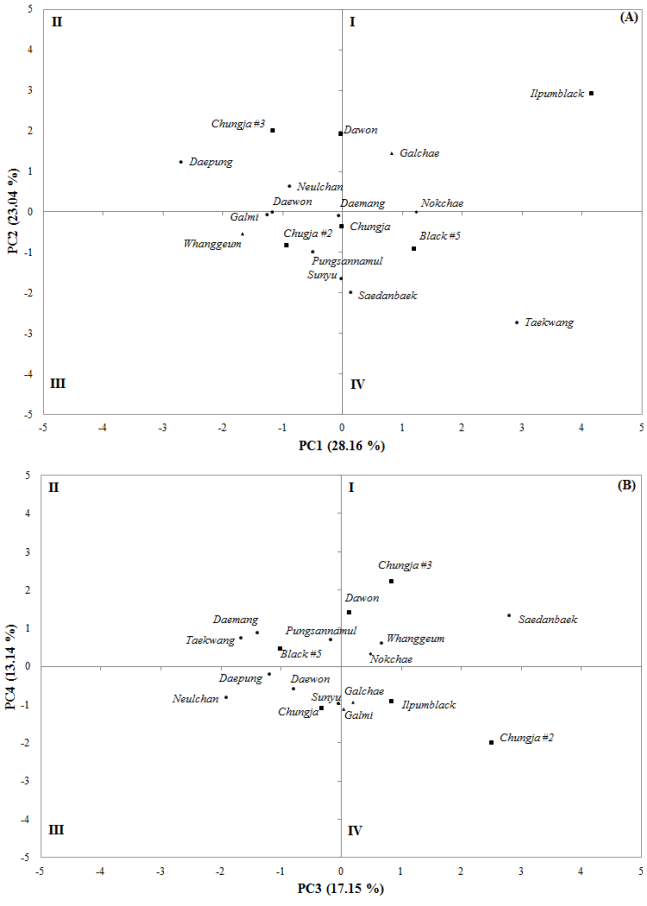

Fig. 3. Score plots of PC1-PC2 and PC3-PC4 for fatty acid profiles in soybeans. (A) PC1-PC2 and (B) PC3-PC4. Cultivar symbols: (●) Yellow; (■) Black; (▲) Brown; (◆) Green.

cultivars. Plane II is especially descriptive of a negative contribution of PC1 and a positive contribution of PC2. It predominately contains *Neulchan*, *Daepung*, and *Chungja #3* cultivars, which possess high linoleic acid content, but lower oleic acid content relative to other cultivars in the sample set (Fig. 2A-II and Fig. 3A-II). These relationships are apparent in Table 4, which denotes the negative vector loadings of arachidic acid to PC2, and the positive vector loadings of stearic acid to PC2. Plane III is especially described by a negative contribution from PC1 and PC2. This plane comprises mostly *Galmi*, *Whanggeum*, *Chungja #2*, and *Punsannamul* samples. Observations of plane III reveal the lower quantities of oleic acid compared to other samples in the sample set (Fig. 2A-III and Fig. 3A-III). Plane IV is indicative of a positive contribution of PC1 and a negative contribution of PC2. It contains mostly *Black #5*, *Saedanbaek*, and *Taekwang* cultivars (Fig. 2A-IV and Fig. 3A-IV). Observations of plane IV show that it contains higher levels of oleic and behenic acids content relative to the other samples in the sample set. For Fig. 3(B), plane I demonstrates higher levels of palmitic acid and is comprised of *Chungja #3*, *Saedanbaek*, *Whanggeum*, and *Nokchae* samples (Fig. 2B-I and Fig. 3B-I); plane II is observed to contain lower levels of lignoceric acid and is comprised of *Daemang*, *Taekwang*, *Black #5* and *Pungsannamul* cultivars (Fig. 2B-II and Fig. 3B-II). Plane III reveals higher levels of oleic acid in the cultivars *Neulchan*, *Chungja*, *Daewon*, and *Daepung* (Fig. 2B-III and Fig. 3B-III). Lastly, analysis of plane IV indicates higher levels of lignoceric acid in the *Chungja #2*, *Ilpumblack*, and *Galchae* samples (Fig. 2B-IV and Fig. 3B-IV). Mohamed and Rangappa (23) analyzed fatty acid profiles in seventeen soybeans and noted that the content of oleic acid was inversely related to the concentrations of linoleic (r=-0.82) and linolenic (r=-0.66) acids. Other investigations reported fatty acid profiles in soybean paste during fermentation using PCA as a tool, however, they didn't classify soybeans by cultivars (26).

This research specifically focused on Korean field-grown soybean cultivars and showed how the contributions of individual fatty acids related to the generated PCs; individual fatty acid contributions to the total variability of a PC are often omitted from other investigations, but are invaluable (23). Eigen analysis of the correlation matrix loadings of the four significant PCs revealed that PC1 was mainly contributed to by oleic, linoleic, and gondoic acids, PC2 by stearic, linolenic, and arachdic acids, PC3 by behenic and lignoceric acids and PC4 by palmitic acid. When the loading and score plots for soybean cultivars were projected as PC1-PC2 and PC3-PC4 groupings, there was an evident reduction in the number

of variables necessary for the discrimination of soybean cultivars. PCs 1 through 4 together were found to be explanatory of more than 81.49% of the total variability in the data set. Statistical examinations need to be conducted routinely with data generated by the most up-to-date scientific methods and technologies to maintain the highest degree of validity. Furthermore, this study clearly indicates that the combination of experimental GC fatty acid data along with a chemometric approach (PCA, in this case) can be successfully employed by soybean researchers to give more information on variation in soybean cultivars than is capable with the experimental data alone.

ACKNOWLEDGMENTS

This study was financial supported by "Cooperative Research Program for Agricultural Science & Technology Development (PJ00846101)", Rural Development Administration (RDA), Republic of Korea.

REFERENCES

1. Akpinar N, Akpinar MA, Türkoğlu Ş. 2001. Total lipid content and fatty acid composition of the seeds of some *Vicia* L. species. *Food Chem* 74: 449-453.
2. Patil AG, Oak MD, Taware SP, Tamhankar SA, Rao VS. 2010. Nondestructive estimation of fatty acid composition in soybean [*Glycine max* (L.) Merrill] seeds using Near-Infrared Transmittance Spectroscopy. *Food Chem* 120: 1210-1217.
3. Liu WH, Stephen Inbaraj B, Chen BH. 2007. Analysis and formation of *trans* fatty acids in hydrogenated soybean oil during heating. *Food Chem* 104: 1740-1749.
4. Liu HR, White PJ. 1992. Oxidative stability of soybean oils with altered fatty acid compositions. *J Am Oil Chem Soc* 69: 528-532.
5. Haswell SJ. 1992. *Practical guide to chemometrics*. Marcel Dekker Inc., New York, NY, USA. p 1-3.
6. Kamal-Eldin A, Andersson R. 1997. A multivariate study of the correlation between tocopherol content and fatty acid composition in vegetable oils. *J Am Oil Chem Soc* 74: 375-380.
7. Brodnjak-Vončina D, Kodba ZC, Novič M. 2005. Multivariate data analysis in classification of vegetable oils characterized by the content of fatty acids. *Chemometr Intell Lab Syst* 75: 31-43.
8. Matos LC, Cunha SC, Amaral JS, Pereira JA, Andrade PB, Seabra RM, Oliveira BPP. 2007. Chemometric characterization of three varietal olive oils (*Cvs. Cobrançosa*, *Madural* and *Verdeal Transmontana*) extracted from olives with different maturation indices. *Food Chem* 102: 406-414.
9. Kadegowda AKG, Piperova LS, Erdman RA. 2008. Principal component and multivariate analysis of milk long-chain fatty acid composition during diet induced milk fat depression. *J Dairy Sci* 91: 749-759.
10. García-López C, Grané-Teruel N, Berenguer-Navarro V, García-García JE, Martín-Carratalá ML. 1996. Major fatty acid composition of 19 almond cultivars of different origins. A chemometric approach. *J Agric Food Chem* 44: 1751-1755.
11. Martín-Carratalá ML, García-López C, Berenguer-Navarro V, Grané-Teruel N. 1998. New contribution to the chemometric characterization of almond cultivars on the basis of their fatty acid profiles. *J Agric Food Chem* 46: 963-967.
12. Sathe SK, Seeram NP, Kshirsagar HH, Heber D, Lapsley KA. 2008. Fatty acid composition of California grown almonds. *J Food Sci* 73: 607-614.
13. Zagonel GF, Peralta-Zamora P, Ramos LP. 2004. Multivariate monitoring of soybean oil ethanolysis by FTIR. *Talanta* 63: 1021-1025.
14. Brandão LF, Braga JW, Suarez PA. 2012. Determination of vegetable oils and fats adulterants in diesel oil by high performance liquid chromatography and multivariate methods. *J Chromatogr A* 1225: 150-157.
15. Luna AS, da Silva AP, Pinho JSA, Ferré J, Boqué R. 2012. Rapid characterization of transgenic and non-transgenic soybean oils by chemometric methods using NIR spectroscopy. *Spectrochim Acta A Mol Biomol Spectrosc* [Epub ahead of print].
16. Bligh EG, Dyer WJ. 1959. A rapid method of total lipid extraction and purification. *Can J Biochem Physiol* 37: 911-917.
17. Ngeh-Ngwainbi J, Lin J, Chandler A. 1997. Determination of total, saturated, unsaturated, and monounsaturated fats in cereal products by acid hydrolysis and capillary gas chromatography: Collaborative study. *J AOAC Int* 80: 359-372.
18. Hashim IB, Koehler PE, Eitenmiller RR, Kvien CK. 1993. Fatty acid composition and tocopherol content of drought stressed Florunner peanuts. *Peanut Sci* 20: 21-24.
19. Sola-Larrañaga C, Navarro-Blasco I. 2009. Chemometric analysis of minerals and trace elements in raw cow milk from the community of Navarra, Spain. *Food Chem* 112: 189-196.
20. Massart DL, Vandeginste BGM, Deming SN, Michotte Y, Kaufman L. 1988. *Chemometrics: A textbook*. Elsevier Science Publishing Company Inc., NY, USA. p 339-370.
21. Tsimidou M, Macrae R, Wilson I. 1987. Authentication of virgin olive oils using principal component analysis of triglyceride and fatty acid profiles: Part 1－Classification of Greek olive oils. *Food Chem* 25: 227-239.
22. Kaiser HF. 1960. The application of electronic computers to factor analysis. *Educ Psychol Meas* 20: 141-151.
23. Mohamed AI, Rangappa M. 1992. Nutrient composition and anti-nutritional factors in vegetable soybean: II. Oil, fatty acids, sterols, and lipoxygenase activity. *Food Chem* 44: 277-282.
24. Patel M, Jung S, Moore K, Powell G, Ainsworth C, Abbott A. 2004. High oleate peanut mutants result from a MITE insertion into the FAD2 gene. *Theor Appli Genet* 108: 1492-1502.
25. López A, Montaño A, García P, Garrido A. 2006. Fatty acid profile of table olives and its multivariate characterization using unsupervised (PCA) and supervised (DA) chemometrics. *J Agric Food Chem* 54: 6747-6753.
26. Namgung HJ, Park HJ, Cho IH, Choi HK, Kwon DY, Shin SM, Kim YS. 2010. Metabolite profiling of doenjang, fermented soybean paste, during fermentation. *J Sci Food Agric* 90: 1926-1935.

(Received July 9, 2012; Accepted September 6, 2012)