

# Determination of lipophilic metabolites for species discrimination and quality assessment of nine leafy vegetables

Tae Jin Kim<sup>1</sup> · Kyoung Bok Lee<sup>1</sup> · Seung-A Baek<sup>1</sup> ·  
Jaehyuk Choi<sup>1</sup> · Sun-Hwa Ha<sup>2</sup> · Sun-Hyung Lim<sup>3</sup> ·  
Soo-Yun Park<sup>3</sup> · Yunsoo Yeo<sup>3</sup> · Sang Un Park<sup>4</sup> · Jae Kwang Kim<sup>1</sup>

Received: 8 July 2015 / Accepted: 1 September 2015 / Published online: 21 September 2015  
© The Korean Society for Applied Biological Chemistry 2015

**Abstract** The lipophilic compounds in nine vegetables consumed in Korea were characterized for diversity in phytochemical content. We also analyzed the relationships among these compounds in terms of their contents. The profiles of 18 lipophilic compounds in the leaves were subjected to data-mining processes, including principal component analysis (PCA), Pearson's correlation analysis, hierarchical clustering analysis (HCA), and partial least squares discriminant analysis (PLS-DA). These species could be distinguished by means of the PCA results. HCA of these phytochemicals resulted in clusters derived from closely related biochemical pathways. PLS-DA showed significant separation among extracts from the following four families: Amaranthaceae, Asteraceae, Brassicaceae, and Malvaceae. The major metabolites that facilitated differentiation in the PLS-DA model were campesterol,  $\beta$ -sitosterol,  $\beta$ -amyirin, stigmasterol, cholesterol, octacosanol (c28),  $\alpha$ -amyirin, and hexacosanol (c26). Chard contained high levels of triacontanol (c30), which was positively

correlated with the c28 content ( $r = 0.746$ ,  $p < 0.0001$ ). Stigmasterol,  $\alpha$ -amyirin, and  $\beta$ -amyirin contents were higher in Asteraceae species including chicon, lettuce, and oak than other plant families. These results demonstrate the utility of metabolic profiling, combined with multivariate analysis, for discrimination of vegetable species as well as evaluation of food quality.

**Keywords** Multivariate analysis · Phytosterol · Policosanol · Tocopherol · Vegetable

## Introduction

Lipophilic compounds such as tocopherols, phytosterols, and policosanols are becoming more popular because of their benefits to human health. Dietary tocopherols, especially  $\alpha$ -tocopherol, can prevent oxidative stress in vivo and are beneficial for overall health (Haffner 2000). Phytosterols have also been shown to reduce total serum cholesterol levels, reduce low-density lipoprotein (LDL) cholesterol levels, and increase high-density lipoprotein (HDL) cholesterol levels in the blood (Arruzazabala et al. 2002). In addition, consuming 5–20 mg/day policosanols has been shown to lower total cholesterol by 17–21 %, and to lower LDL by 21–29 % and increase HDL by 8–15 % (Arruzazabala et al. 1993; Gouni-Berthold and Berthold 2002). Previously, we determined the phytochemical diversity among varieties of perilla (Kim et al. 2012d). However, to date, few studies have reported the contents of lipophilic compounds in leafy vegetables.

Most vegetables contain low-to-moderate levels of vitamin E. Due to their abundance in our diets, vegetables are an important and consistent source of vitamin E.

✉ Sang Un Park  
supark@cnu.ac.kr

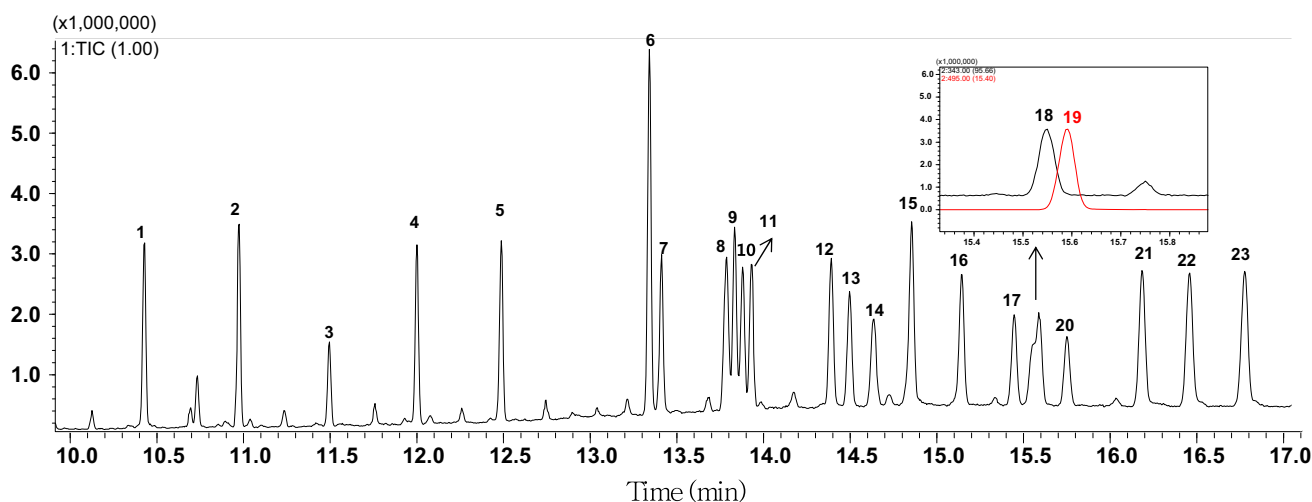
✉ Jae Kwang Kim  
kjkpj@inu.ac.kr

<sup>1</sup> Division of Life Sciences, College of Life Sciences and Bioengineering, Incheon National University, Incheon 406-772, Republic of Korea

<sup>2</sup> Graduate School of Biotechnology, Crop Biotech Institute, Kyung Hee University, Youngin 446-701, Republic of Korea

<sup>3</sup> National Academy of Agricultural Science, Rural Development Administration, Wanju-gun, Jeollabuk-do 565-851, Republic of Korea

<sup>4</sup> Department of Crop Science, Chungnam National University, Daejeon 305-764, Republic of Korea



**Fig. 1** Representative total ion chromatogram of a standard mixture obtained in scanning mode as TMS derivatives separated on Rtx-5MS column. The upper traces were recorded in SIM mode. *Peak 1* c20 (eicosanol); *2* c21 (heneicosanol); *3* c22 (docosanol); *4* c23 (tricosanol); *5* c24 (tetracosanol); *6* 5 $\alpha$ -cholestane (internal standard);

*7* c26 (hexacosanol); *8*  $\beta$ -tocopherol; *9*  $\gamma$ -tocopherol; *10* c27 (heptacosanol); *11*  $\delta$ -tocotrienol; *12* c28 (octacosanol); *13*  $\gamma$ -tocotrienol; *14*  $\alpha$ -tocopherol; *15* cholesterol; *16* brassicasterol; *17*  $\alpha$ -tocotrienol; *18* campesterol; *19* c30 (triacontanol); *20* stigmasterol; *21*  $\beta$ -sitosterol; *22*  $\beta$ -amyirin; *23*  $\alpha$ -amyirin

**Table 1** Gas chromatographic and mass spectrometric data of trimethylsilyl derivatives of lipophilic compounds

Compound <sup>a</sup>	RT <sup>b</sup>	RRT <sup>c</sup>	[M] <sup>+</sup>	Other characteristic ions <sup>d</sup> [m/z (%)]
c20	10.43	0.782	370 (0)	<b>355 (100)</b>
c21	10.98	0.823	384 (0)	<b>369 (100)</b>
c22	11.50	0.862	398 (0)	<b>383 (100)</b>
c23	12.02	0.901	411 (0)	<b>397 (100)</b>
c24	12.49	0.936	426 (0)	<b>411 (100)</b>
c26	13.43	1.006	454 (0)	<b>439 (100)</b>
$\beta$ -Tocopherol	13.79	1.034	489 (100)	263 (13) <b>222 (87)</b>
$\gamma$ -Tocopherol	13.84	1.037	488 (60)	263 (10) <b>223 (100)</b>
c27	13.89	1.041	468 (0)	<b>453 (100)</b>
$\delta$ -tocotrienol	13.94	1.045	469 (45)	264 (14) 249 (43) <b>209 (100)</b>
c28	14.40	1.079	482 (0)	<b>467 (100)</b>
$\gamma$ -Tocotrienol	14.50	1.087	483 (43)	261 (21) <b>223 (100)</b>
$\alpha$ -Tocopherol	14.65	1.098	503 (90)	277 (8) <b>237 (100)</b>
Cholesterol	14.87	1.114	458 (0)	368 (47) <b>329 (53)</b> 255 (17) 247 (15) 129 (100)
Brassicasterol	15.15	1.136	470 (23)	380 (31) 341 (19) 255 (45) <b>129 (100)</b>
$\alpha$ -Tocotrienol	15.45	1.158	497 (50)	275 (19) <b>237 (100)</b>
Campesterol	15.56	1.166	473 (0)	382 (38) 367 (23) <b>343 (50)</b> 255 (18) 129 (100)
c30	15.59	1.169	510 (0)	<b>496 (100)</b> , 495 (60)
Stigmasterol	15.76	1.181	484 (23)	<b>394 (38)</b> 355 (15) 255 (52) 129 (100)
$\beta$ -Sitosterol	16.20	1.214	487 (17)	396 (42) <b>357 (44)</b> 255 (17) 129 (100)
$\beta$ -Amyrin	16.47	1.234	499 (0)	<b>218 (100)</b> 203 (42) 189 (26)
$\alpha$ -Amyrin	16.79	1.259	499 (0)	218 (100) 203 (18) <b>189 (32)</b>

<sup>a</sup> c20, eicosanol; c21, heneicosanol; c22, docosanol; c23, tricosanol; c24, tetracosanol; c26, hexacosanol; c27, heptacosanol; c28, octacosanol; c30, triacontanol

<sup>b</sup> Retention time (min)

<sup>c</sup> Relative retention times (retention time of analyte/retention time of 5 $\alpha$ -cholestane)

<sup>d</sup> Ions in boldface indicate the specific mass ion used for quantification

Vegetable oils are known to be the best natural sources of dietary plant sterols (Piironen et al. 2000). Validated food composition data are needed to further investigate the effects of natural plant sterol levels. The policosanols composition of various materials, such as beeswax, rice bran, wheat germ, sugar cane, sorghum grain, perilla seed, and wheat grain, has been analyzed (Hwang et al. 2004; Lin et al. 2004; Irmak and Dunford 2005; Adhikari et al. 2006; Kim et al. 2012d). However, to the best of our knowledge, there is little published data on the variation in policosanols composition of leafy vegetables.

Plant surfaces exposed to the atmosphere usually have a layer of wax that preserves plant water balance, minimizes mechanical damage to cells, and resists fungal and insect attacks (Irmak and Dunford 2005). Sterols are integral membrane components that serve to regulate the fluidity and permeability of membranes, and affect various membrane functions such as simple diffusion and active transport across the membrane. Sterols also modulate the activities of membrane-associated proteins, including enzymes, receptors, and signal transduction components (Piironen et al. 2000). Correlations between the concentrations of various metabolites can be computed to understand metabolic associations (Park et al. 2014). These correlations are the net result of direct enzymatic conversions and indirect cellular regulation of transcriptional or biochemical processes. Thus, additional information can be obtained from functional genomic studies (Steuer et al. 2003).

Plant identification and differentiation at the species and individual genotype levels are of major importance for plant scientists (Goodacre et al. 2007). Differences in species or the origin thereof may result in different metabolic phenotypes, as plant metabolite composition is influenced by species or origin. Metabolic profiling facilitates classification of samples of diverse biological status, origin, or quality, using chemometrics such as principal component analysis (PCA) and partial least squares discriminate analysis (PLS-DA) (Kim et al. 2013). Previously, the multivariate tools of metabolome were used to classify herb plants (Kim et al. 2012a), to predict the flower colors of *Catharanthus roseus* (Pan et al. 2012) and the quality of green tea (Pongsuwan et al. 2007), and to identify several unknown vegetable oils (Rusak et al. 2003). However, no metabolic profiling studies have been conducted on the leafy vegetables. Thus, we used PCA and PLS-DA to discriminate according to species and family, and compare the quality of leafy vegetable samples, often consumed in Korea.

The most commonly used technique for determining tocopherol content is normal-phase HPLC with ultraviolet or fluorescence detection. Gas chromatography (GC) is the most frequently used method of analyzing sterols and policosanols (Adhikari et al. 2006). Furthermore, GC–mass

spectrometry (GC–MS) can be used to simultaneously quantify lipophilic compounds such as fatty acids, tocopherols, phytosterols, and policosanols (Lechner et al. 1999; Shepherd et al. 2007). This study evaluated the lipophilic composition such as policosanols, sterols, and tocopherols in nine leafy vegetables, and to identify vegetable species by lipophilic profiling using GC–MS and multivariate analysis.

## Materials and methods

### Samples and chemicals

Mustard, oak, Chinese mallow, chicon, kale, lettuce, tatsoi, amaranth, and chard were purchased at local supermarkets. Each sample was collected in triplicate. The samples were freeze-dried at  $-70\text{ }^{\circ}\text{C}$  for at least 72 h and pulverized into a fine powder using a mortar and pestle. The milled powders were kept at  $-70\text{ }^{\circ}\text{C}$  prior to extraction. Eicosanol

**Table 2** Linearity and limit of detection for the assay of 22 lipophilic compounds as TMS derivatives

Compound <sup>a</sup>	Calibration curve		Linearity <sup>b</sup>	LOD <sup>c</sup> ( $\mu\text{g}$ )
	Slope	Intercept		
c20	5.5302	−0.5562	0.9952	0.010
c21	5.7924	−0.1396	0.9993	0.005
c22	2.2903	−0.1618	0.9982	0.005
c23	4.3386	+0.0681	0.9989	0.010
c24	5.0162	−0.8048	0.9972	0.005
c26	1.9141	−0.1801	0.9992	0.010
$\beta$ -tocopherol	5.1054	−0.4289	0.9996	0.020
$\gamma$ -Tocopherol	5.3583	−0.3956	0.9993	0.010
c27	1.8125	−0.1955	0.9992	0.020
$\delta$ -Tocotrienol	2.2631	−0.2756	0.9945	0.020
c28	1.2644	−0.3246	0.9912	0.010
$\gamma$ -Tocotrienol	3.0151	−0.2622	0.9992	0.020
$\alpha$ -Tocopherol	3.5579	−0.3415	0.9991	0.020
Cholesterol	1.0998	−0.0318	0.9997	0.010
Brassicasterol	1.5648	−0.3041	0.9948	0.020
$\alpha$ -Tocotrienol	3.0971	−0.3566	0.9988	0.020
Campesterol	0.3768	−0.0593	0.9914	0.020
c30	0.4382	−0.0304	0.9915	0.020
Stigmasterol	0.2779	−0.0214	0.9997	0.020
$\beta$ -Sitosterol	0.9351	−0.1361	0.9963	0.020
$\beta$ -Amyrin	5.4311	−0.5127	0.9967	0.020
$\alpha$ -Amyrin	1.5760	−0.1655	0.9958	0.010

<sup>a</sup> c20, eicosanol; c21, heneicosanol; c22, docosanol; c23, tricosanol; c24, tetracosanol; c26, hexacosanol; c27, heptacosanol; c28, octacosanol; c30, triacontanol

<sup>b</sup> Linearity, Correlation coefficient

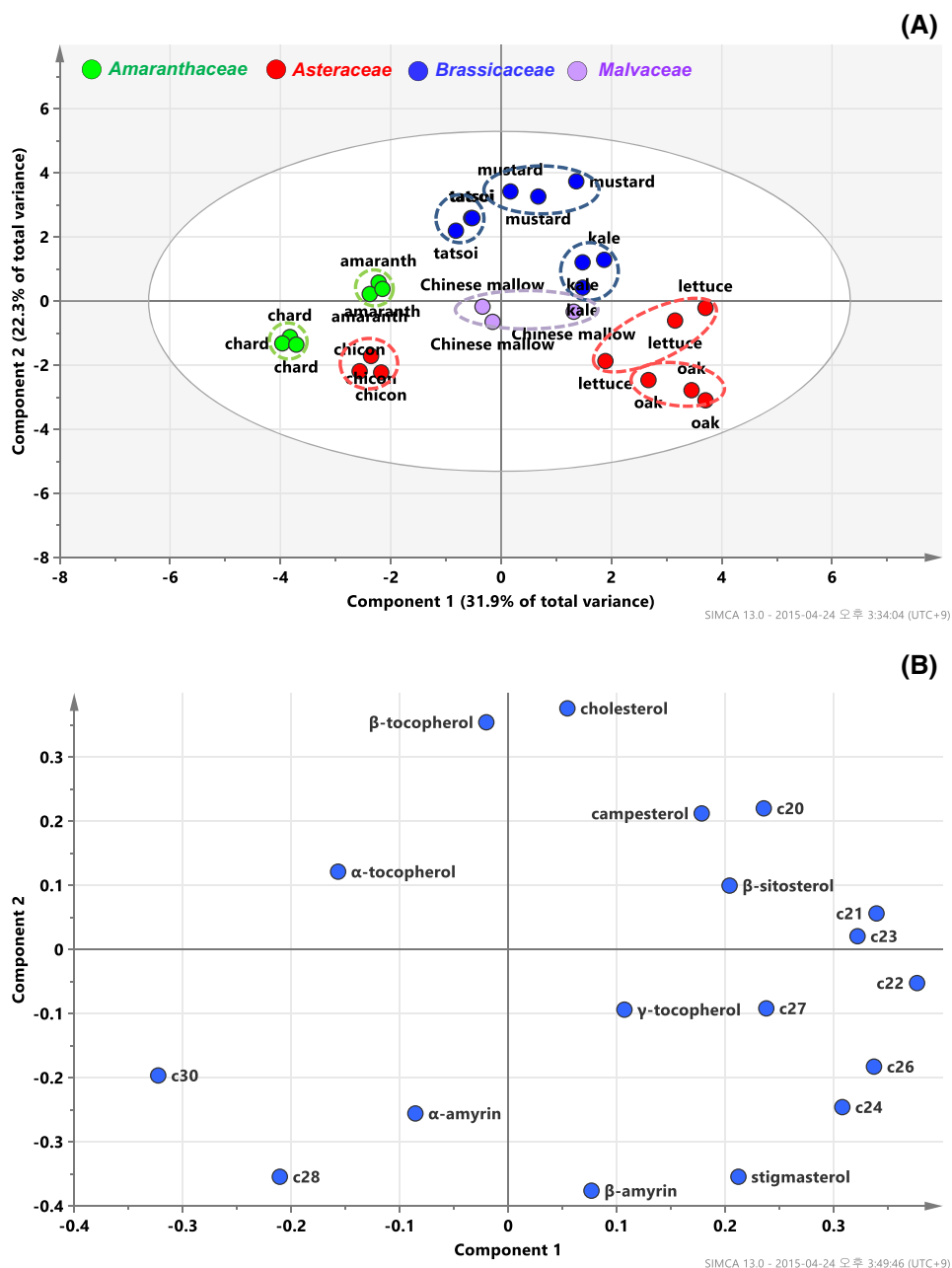
<sup>c</sup> LOD limit of detection (signal-to-noise ratio < 3)

(c20), heneicosanol (c21), tricosanol (c23), tetracosanol (c24), hexacosanol (c26), heptacosanol (c27), octacosanol (c28), triacontanol (c30), campesterol, stigmasterol,  $\beta$ -sitosterol,  $5\alpha$ -cholestane, pyridine, ascorbic acid, and *N*-methyl-*N*-trimethylsilyltrifluoroacetamide (MSTFA) were obtained from Sigma (USA). Docosanol (c22) was obtained from Aldrich. Tocopherols were purchased from Merck (Germany). All other chemicals used in this study were reagent grade, unless stated otherwise.

### Preparation of stock solution and standard solutions

Standard stock solutions of  $\alpha$ -tocopherol,  $\beta$ -tocopherol,  $\gamma$ -tocopherol,  $\alpha$ -tocotrienol,  $\beta$ -tocotrienol,  $\gamma$ -tocotrienol,  $\delta$ -tocotrienol, cholesterol, brassicasterol, campesterol, stigmasterol,  $\beta$ -sitosterol, and the internal standard (IS) used in this study were prepared at 100  $\mu\text{g}/\text{mL}$  in *n*-hexane. Standard stock solutions of c20, c21, c22, c23, c24, c26, c27, c28, c30,  $\beta$ -amyryn, and  $\alpha$ -amyryn were prepared at 100  $\mu\text{g}/$

**Fig. 2** PCA score plots (A) and loading plots (B) derived from lipophilic metabolites of nine vegetable species. PC1 and PC2 accounted for >54.2 % of the total variance. The *ellipse* represents the Hotelling T2 with 95 % confidence in the score plot. The *loading plots* represent the original variables in the space of the PCs. They reveal the magnitude and direction of correlation of the original variables with the first two PCs. c20, eicosanol; c21, heneicosanol; c22, docosanol; c23, tricosanol; c24, tetracosanol; c26, hexacosanol; c27, heptacosanol; c28, octacosanol; c30, triacontanol



mL in chloroform. Calibration samples, ranging from 0.25 to 5.00  $\mu\text{g}$ , were prepared by mixing appropriate aliquots of each stock solution. All standard solutions were stored at  $-20\text{ }^{\circ}\text{C}$ , with the exception of stock solutions, which were stored at  $-70\text{ }^{\circ}\text{C}$ .

### Extraction and analysis of samples

Extraction of tocopherol, phytosterol, and policosanol was performed according to the method developed by our group, with a slight modification (Kim et al. 2012d). Lipophilic metabolite was extracted from the powdered samples (0.05 g) by the addition of 3 mL of ethanol containing 0.1 % ascorbic acid (w/v) and 0.05 mL of  $5\alpha$ -cholestane (10  $\mu\text{g}/\text{mL}$ ) as an IS, mixed by vortexing for 20 s, and placed in a water bath at  $85\text{ }^{\circ}\text{C}$  for 5 min. After removal from the water bath, 120  $\mu\text{L}$  of potassium hydroxide (80 %, w/v) was added, and the samples were vortexed for 20 s and returned to the water bath for 10 min. After saponification, the samples were immediately placed on ice, and deionized water (1.5 mL) was added. Hexane (1.5 mL) was then added to each sample, followed by vortex-mixing for 20 s and centrifugation ( $1200\times g$ , 5 min). The upper layer was pipetted into a separate tube, and the pellet was re-extracted using hexane. The hexane fraction was dried in a centrifugal concentrator (CC-105, TOMY, Japan). For derivatization, 30  $\mu\text{L}$  of MSTFA and 30  $\mu\text{L}$  of pyridine were added, and the mixture was incubated at  $60\text{ }^{\circ}\text{C}$  for 30 min at a mixing frequency of 1200 rpm using a Thermomixer Comfort (model 5355; Eppendorf AG, Germany).

Samples (1.0  $\mu\text{L}$  each) were injected with a split 10:1 ratio and gas chromatography was performed using an Rtx-5MS column (30-m length, 0.25-mm diameter and 0.25- $\mu\text{m}$  thickness). The GCMS-QP2010 Ultra system, with autosampler AOC-20i from Shimadzu (Japan), was used to separate the lipophilic compounds. The injection, interface, and ion source temperatures were set at 290, 280, and  $230\text{ }^{\circ}\text{C}$ , respectively. Helium gas was used as a carrier, with a column flow rate of 1.0 mL/min. The temperature was standardized to 2 min isothermal heating at  $150\text{ }^{\circ}\text{C}$ , followed by an increase with a ramping rate of  $15\text{ }^{\circ}\text{C}/\text{min}$  up to  $320\text{ }^{\circ}\text{C}$ , with a hold time of 10 min. Mass spectra data were recorded at 2.00–23.33 min runtime. The chromatograms and mass spectra were evaluated using the Labsolutions GCMSsolution software version 4.11 (Shimadzu).

### Statistical analysis

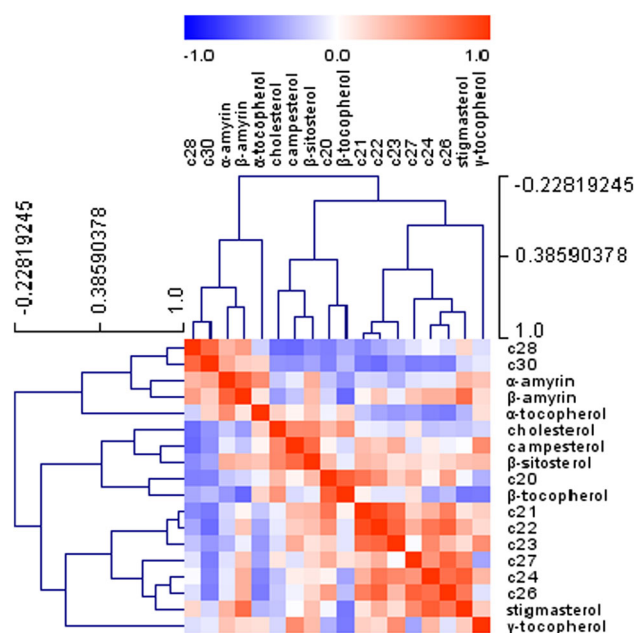
All analyses were performed at least in triplicate. Data acquired from GC–MS were subjected to PCA and PLS-DA (SIMCA-P version 12.0; Umetrics, Sweden) to evaluate the relationships among groups. The PCA and PLS-DA

output consisted of score plots to visualize the contrast among samples and loading plots to explain the cluster separation. Pearson's correlation analysis was performed using the SAS 9.2 software package (SAS Institute, USA). HCA and heat map visualization of the correlation coefficients were performed using the software Multi-Experiment Viewer version 4.9.0 (<http://www.tm4.org/mev/>).

## Results and discussion

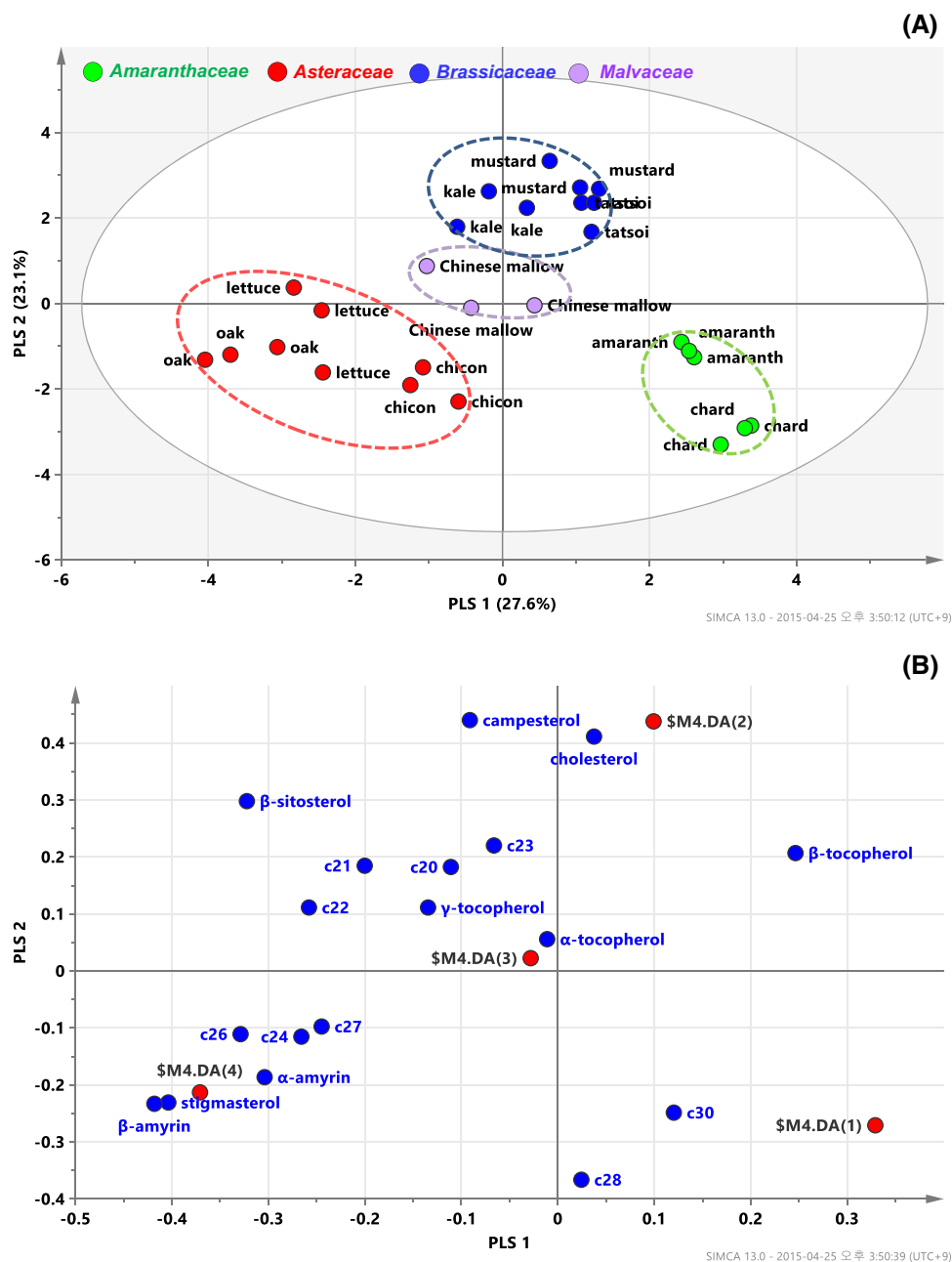
### Lipophilic metabolites profiling of samples

In this study, lipophilic compounds such as policosanol, phytosterol, and tocopherol in vegetables were measured by GC–MS (Fig. 1). Previously, we analyzed 14 lipophilic metabolites using GC–MS (Kim et al. 2012b). Under the GC–MS conditions described here, the peaks of 22 lipophilic compounds were detected within 17 min, resulting in high-throughput analysis of almost 75/24 h. Quantification was performed using selected ions, as described in Table 1. Precursor and product ions in the mass spectra of tocopherol and phytosterol were in agreement with those in the literature (Van Pelt et al. 1998; Schummer et al. 2009). The



**Fig. 3** Correlation matrix and cluster analysis of results of the 18 lipophilic metabolites from nine vegetable species. Each square indicates a Pearson's correlation coefficient for a pair of compounds. The value for the correlation coefficient is represented by the intensity of the blue or red color, as indicated on the color scale. Hierarchical clusters are presented as a cluster tree. c20, eicosanol; c21, heneicosanol; c22, docosanol; c23, tricosanol; c24, tetracosanol; c26, hexacosanol; c27, heptacosanol; c28, octacosanol; c30, triacontanol

**Fig. 4** PLS-DA score plots (A) and loading plots (B) derived from lipophilic metabolites of nine vegetable species. PC1 and PC2 accounted for >50.7 % of the total variance. The ellipse represents the Hotelling T2 with 95 % confidence in the score plot. The loading plot defines the orientation of the PLS planes with the original variables (metabolites and family variables). They reveal the magnitude and direction of correlations of the original variables with the first two PLS components. Family variables: \$M8.DA(1), Amaranthaceae; \$M8.DA(2), Brassicaceae; \$M8.DA(3), Malvaceae; \$M8.DA(4), Asteraceae. c20, eicosanol; c21, heneicosanol; c22, docosanol; c23, tricosanol; c24, tetracosanol; c26, hexacosanol; c27, heptacosanol; c28, octacosanol; c30, triacontanol



mass spectra of policosanols TMS derivatives showed that the molecular ion  $[M-15]^+$  was generally dominant. TMS-derivates of sterols showed mainly the fragments  $[M-90]^+$ ,  $[M-129]^+$  and  $m/z$  129.

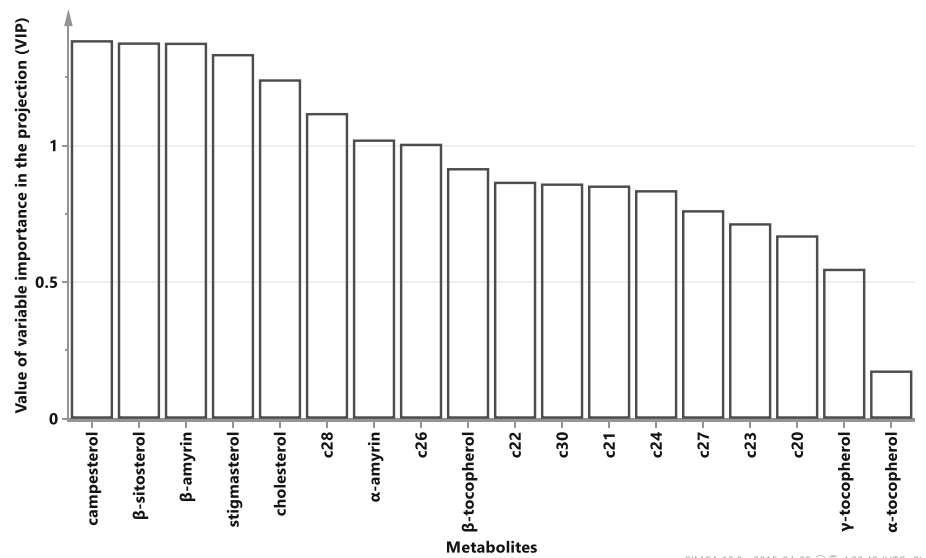
Quantification was performed by means of three-point calibration curves, for which the concentrations of a mixture of 22 authentic standards ranged from 0.25 to 5.0  $\mu\text{g}$ , whereas the amount of the IS was constant at 0.5  $\mu\text{g}$ . The calibration curves of 22 lipophilic compounds measured under optimized saponification conditions were linear ( $r^2 = 0.9912\text{--}0.9997$ ), and LODs varied within the range of 0.005–0.020  $\mu\text{g}$  (Table 2). These results confirmed the suitability of the present method for measurement of

tocopherols, tocotrienols, policosanols, and sterols in unknown samples.

Tocopherols are lipophilic antioxidants with important functions in plants and humans (Haffner 2000). Phytosterols are bioactive lipophilic compounds capable of reducing serum cholesterol (Agren et al. 2001). Policosanols have various beneficial physiological effects, including reducing platelet aggregation, endothelial damage, and foam cell formation, and improving the exercise performance of coronary heart disease patients (Irmak and Dunford 2005). Thus, we analyzed the contents of these metabolites in nine vegetables: mustard leaf, oak leaf, Chinese mallow, chicon, kale, red lettuce, tatsoi, amaranth, and chard. In the present



**Fig. 5** The influence of variables used to create a family predictor for a vegetable sample. VIP indicates the relative influence of each metabolite to the grouping. Metabolites with higher VIP values are more influential. c20, eicosanol; c21, heneicosanol; c22, docosanol; c23, tricosanol; c24, tetracosanol; c26, hexacosanol; c27, heptacosanol; c28, octacosanol; c30, triacontanol



SIMCA 13.0 - 2015-04-25 오후 4:02:48 (UTC+9)

study, the alkaline saponification method was used to measure lipophilic components. The quantitative results include free sterols and their esters, but not steryl glycosides. In total, 18 metabolites, including three tocopherols, four phytosterols, nine policosanols, β-amyrin, and α-amyrin, were detected in the samples.

#### Classification of vegetable species by principal component analysis and correlation between metabolites in vegetables by hierarchical clustering analysis and Pearson's correlation analysis

Food component analysis traditionally involves classifying food constituents into very broad categories, such as proteins, fats, carbohydrates, fiber, vitamins, trace elements, solids, and ash. However, the advent of metabolomics has allowed metabolite profiling combined with chemometrics to direct quality control strategies for improving and optimizing the specific balance of the components of fresh food (Kim et al. 2012c). PCA is an important tool, especially in preliminary steps of a multivariate analysis, to obtain an overview of, and find patterns in, complex experimental data. The data obtained for the lipophilic metabolites were subjected to PCA to outline the lipophilic profile differences among the vegetable samples. PCA transforms the original variables, using an orthogonal linear transformation, to a new set of uncorrelated variables known as principal components (PCs). Representation of the PCs scores and loadings in a bi-dimensional plot can provide an overview of the data, pointing out patterns hidden in the data set and identifying correlations among variables. Each point in the score plots represents an individual sample, and samples exhibiting similar variances are clustered together.

The PCA results clearly demonstrated the absence of large variances among samples of the same species (Fig. 2). Furthermore, the results showed apparent separation among three families, the exception being Asteraceae. The first PC, accounting for 31.9 % of the total variance, resolved the measured metabolites profiles of Amaranthaceae and other families such as Brassicaceae and Malvaceae. The corresponding loading plots identify the metabolites responsible for separation on the score plots. The significant metabolites for PC1 were policosanols such as c22, c21, c26, and c23. In addition, in PC1, the corresponding loading was negative for c28 and c30, suggesting that the c28 and c30 contents of chard were higher than those of other vegetable species.

To examine detailed relationships among the concentrations of the 18 metabolites in vegetable, we performed Pearson's correlation analysis and HCA on data for the accessions (Fig. 3). There were strong correlations between metabolites that participate in closely related pathways, demonstrating the robustness of the present experimental system. In our studies, significant positive relationships were observed between β-amyrin and α-amyrin ( $r = 0.7740$ ,  $P < 0.0001$ ). Likewise, there were significant correlations among sterols such as cholesterol, campesterol, and β-sitosterol. These results suggest that the reasonable score range of the components could be used for sample selection according to the correlation between the variables and these two components.

#### Classification of the vegetable species by PLS-DA

To optimize the separation among Amaranthaceae, Asteraceae, Brassicaceae, and Malvaceae, we then utilized PLS-DA to visualize the metabolite differences. PLS-DA is a

**Table 3** Composition and content ( $\mu\text{g/g}$ ) of lipophilic compounds in nine vegetables

Compound <sup>a</sup>	Amaranthaceae			Asteraceae			Brassicaceae			Malvaceae	
	Amaranth	Chard	Chicon	Lettuce	Oak	Kale	Mustard	Tatsoi	Chinese mallow		
<i>Policosanols</i>											
c20	3.65 ± 0.18	3.29 ± 0.20	3.66 ± 0.25	13.14 ± 3.94	5.43 ± 0.26	7.03 ± 0.60	15.00 ± 1.58	4.24 ± 0.30	3.30 ± 0.23		
c21	0.71 ± 0.01	0.58 ± 0.04	0.75 ± 0.02	1.36 ± 0.25	1.08 ± 0.03	1.43 ± 0.09	1.04 ± 0.16	0.76 ± 0.07	0.71 ± 0.05		
c22	19.19 ± 0.74	6.12 ± 0.39	20.76 ± 2.94	147.70 ± 50.18	184.85 ± 14.55	150.59 ± 5.35	84.50 ± 0.52	44.30 ± 2.61	24.63 ± 3.77		
c23	1.13 ± 0.14	ND <sup>b</sup>	0.68 ± 0.13	1.55 ± 0.39	3.02 ± 0.25	3.87 ± 0.42	1.91 ± 0.26	0.69 ± 0.16	1.83 ± 0.81		
c24	37.86 ± 3.26	16.39 ± 1.07	20.41 ± 6.31	62.69 ± 8.98	123.73 ± 7.28	45.22 ± 1.49	28.26 ± 2.27	15.55 ± 0.91	73.53 ± 14.46		
c26	81.95 ± 7.70	133.18 ± 31.83	65.46 ± 36.36	1671.19 ± 102.07	1283.59 ± 139.97	364.73 ± 7.14	304.13 ± 49.07	211.42 ± 8.88	365.03 ± 133.32		
c27	15.93 ± 3.90	13.73 ± 3.80	10.68 ± 1.44	45.96 ± 13.10	38.82 ± 11.74	2.44 ± 0.13	26.88 ± 8.14	22.70 ± 6.41	22.80 ± 3.62		
c28	293.54 ± 87.83	682.74 ± 99.18	472.99 ± 151.25	233.08 ± 144.70	387.81 ± 82.92	194.77 ± 57.32	133.69 ± 40.79	102.73 ± 28.38	305.77 ± 114.7		
c30	279.47 ± 29.52	1035.95 ± 103.11	837.15 ± 188.00	43.65 ± 2.63	74.40 ± 14.38	270.17 ± 56.49	67.89 ± 15.91	121.67 ± 40.16	211.12 ± 35.09		
<i>Sterols and terpenoids</i>											
$\alpha$ -Amyrin	4.02 ± 1.56	11.22 ± 2.06	1596.14 ± 61.18	236.26 ± 140.34	302.15 ± 86.57	12.81 ± 1.31	4.34 ± 0.92	3.76 ± 0.44	179.10 ± 303.78		
$\beta$ -Amyrin	5.64 ± 0.81	ND	290.35 ± 12.95	140.76 ± 56.05	266.87 ± 19.83	2.33 ± 0.14	2.67 ± 0.12	2.13 ± 0.06	26.89 ± 32.55		
Campesterol	83.59 ± 5.21	48.16 ± 6.87	342.57 ± 58.99	265.31 ± 18.34	344.57 ± 17.51	696.53 ± 27.22	420.81 ± 26.81	690.23 ± 105.18	606.51 ± 109.25		
Cholesterol	2.85 ± 0.40	3.75 ± 0.85	10.10 ± 2.82	6.40 ± 0.57	10.29 ± 0.61	9.25 ± 1.90	44.00 ± 3.10	41.93 ± 3.42	21.10 ± 2.76		
$\beta$ -Sitosterol	15.52 ± 1.45	44.66 ± 13.36	836.81 ± 93.71	630.36 ± 59.31	588.05 ± 28.08	608.23 ± 110.61	759.08 ± 57.53	812.18 ± 37.71	662.94 ± 209.05		
Stigmasterol	10.05 ± 1.04	121.30 ± 18.16	462.40 ± 97.47	552.72 ± 17.72	868.18 ± 34.23	17.71 ± 2.63	76.06 ± 21.52	47.58 ± 12.87	595.43 ± 176.90		
$\alpha$ -Tocopherol	193.99 ± 27.27	117.49 ± 22.41	250.52 ± 42.62	118.57 ± 13.05	125.46 ± 19.69	149.94 ± 14.91	194.02 ± 22.36	178.67 ± 23.62	125.39 ± 17.02		
$\beta$ -Tocopherol	3.05 ± 0.14	2.73 ± 0.24	2.22 ± 0.06	ND	2.23 ± 0.06	2.47 ± 0.34	4.09 ± 0.09	2.67 ± 0.16	2.35 ± 0.14		
$\gamma$ -Tocopherol	18.28 ± 0.41	3.57 ± 0.34	29.58 ± 5.14	9.19 ± 2.19	24.84 ± 1.29	40.5 ± 3.18	10.97 ± 0.6	11.90 ± 2.52	35.53 ± 4.86		

Each value is the mean of three replications  $\pm$  standard deviation<sup>a</sup> c20, eicosanol; c21, heneicosanol; c22, docosanol; c23, tricosanol; c24, tetracosanol; c26, hexacosanol; c27, heptacosanol; c28, octacosanol; c30, triacontanol<sup>b</sup> ND not detected



projection method that separates groups of observations by rotating the PCAs to obtain the maximum separation among classes. In this study, classes were the families. Good separation was obtained among the score plots in PC1 and PC2 of PLS-DA (Fig. 4). We used an internal validation method for model validation. The quality of the model was described by  $R^2$  and  $Q^2$  values.  $R^2$  is defined as the proportion of variance in the data explained by the model and indicates the goodness of fit, and  $Q^2$  is defined as the proportion of variance in the data predictable by the model and indicates the predictability. Usually,  $Q^2 > 0.9$  indicates a strong predictive ability of the model,  $Q^2 > 0.5$  indicates a general predictive ability of the model, and  $Q^2$  smaller than 0.5 indicates a poor ability of the model (Eriksson et al. 2001). The predictive model had an  $R^2$  of 0.614 and  $Q^2$  of 0.547. According to Eriksson et al. (2001),  $Q^2 > 0.5$  indicates a good predictive model. In the first component of the PLS-DA, the loading plot indicated that c30 levels were higher in amaranth and chard than in other species. The loading plot also indicated that high levels of campesterol were present in the Brassicaceae samples, and high levels of  $\beta$ -amyirin and stigmasterol were present in the Asteraceae samples. The contribution of variables in the projection could be explained using variable importance in the projection (VIP) values (Fig. 5). Variables with  $VIP > 1$  have the greatest influence on the model. Campesterol was found to be the most significant component in the PLS-DA model for the vegetable family.

### Lipophilic metabolite contents in leafy vegetables

To date, few studies have reported comprehensive measurements of lipophilic compounds in vegetables. The contents of policosanols, sterols, and terpenoids in leafy vegetables commonly consumed in Korea are shown in Table 3. In general, c26 and c28 were the most abundant policosanols in vegetable samples. The c28 level was higher in chard (0.68 mg/g). Due to its cytoprotective effects, c28 is one of the most abundant alcohols in policosanols and is consumed as an alternative to aspirin for patients suffering from gastric irritation (Taylor et al. 2003). One study involving tail-suspended rats showed that c28 could counteract some effects of simulated weightlessness on rats. Chard might thus benefit astronauts during space travels (Bai 1997). In this study, three foods belonging to the Asteraceae species contained higher amounts of  $\alpha$ -amyirin and  $\beta$ -amyirin. The relative  $\alpha$ -amyirin and  $\beta$ -amyirin contents could potentially serve as markers to distinguish Asteraceae from other families.

In summary, this study demonstrated the diversity of lipophilic compounds in nine leafy vegetables and identified correlations among their contents. PCA and PLS-DA were performed to identify differences among the plant species

and families. PCA facilitated visualization of complex data and identification of significant differences between species. PLS-DA enabled discrimination of vegetable samples from different families. Metabolites—such as campesterol,  $\beta$ -sistosterol,  $\beta$ -amyirin, stigmasterol, cholesterol, c28,  $\alpha$ -amyirin, and c26—are useful to identify specific plant families. The HCA results indicated strong correlations among metabolites in closely related pathways.

One means of enhancing the intake of phytochemicals would be to increase their content in fresh vegetables by utilizing crop production practices such as selection of species and cultivars. These vegetables could serve as fresh market products or as raw materials for functional foods and supplements. Reliable discrimination of biomarkers of vegetables is essential for the consumer and producer. This study indicates that chard vegetable contains relatively high levels of c30. Furthermore, the Asteraceae species, including chicon, lettuce, and oak, contain relatively high levels of  $\alpha$ -amyirin,  $\beta$ -amyirin, and stigmasterol. Therefore, these metabolites could be the chemotaxonomic markers to discriminate the Asteraceae species from other leafy vegetables. The potential of metabolite profiling combined with chemometrics as a powerful tool for assessing food quality was demonstrated here.

**Acknowledgments** This work was supported by a grant from the Next-Generation BioGreen 21 Program (SSAC, PJ011094042015), Rural Development Administration, Republic of Korea.

### References

- Adhikari P, Hwang KT, Park JN, Kim CK (2006) Policosanols content and composition in perilla seeds. *J Agric Food Chem* 54:5359–5362
- Agren JJ, Tvrzicka E, Nenonen MT, Helve T, Hänninen O (2001) Divergent changes in serum sterols during a strict uncooked vegan diet in patients with rheumatoid arthritis. *Br J Nutr* 85:137–139
- Arruzazabala ML, Carbajal D, Mas R, Garcia M, Fraga V (1993) Effects of policosanols on platelet aggregation in rats. *Thromb Res* 69:321–327
- Arruzazabala ML, Molina V, Mas R, Fernández L, Carbajal D, Valdés S, Castaño G (2002) Antiplatelet effects of policosanols (20 and 40 mg/day) in healthy volunteers and dyslipidaemic patients. *Clin Exp Pharmacol Physiol* 29:891–897
- Bai S (1997) Effects of octacosanol in food physiological parameters in tail-suspended rats. *Space Med Med Eng* 10:450–452
- Eriksson L, Johansson E, Kettaneh-Wold N, Wold S (2001) Multi- and megavariate data analysis principles and applications. Umetrics AB, Umeå
- Goodacre R, Roberts L, Ellis DI, Thorogood D, Reader SM, Ougham H, King I (2007) From phenotype to genotype: whole tissue profiling for plant breeding. *Metabolomics* 3:489–501
- Gouni-Berthold I, Berthold HK (2002) Policosanols: clinical pharmacology and therapeutic significance of a new lipid-lowering agent. *Am Heart J* 143:356–365
- Haffner SM (2000) Clinical relevance of the oxidative stress concept. *Metabolism* 49:30–34

- Hwang KT, Weller CL, Cuppett SL, Hanna MA (2004) Policosanol contents and composition of grain sorghum kernels and dried distillers grains. *Cereal Chem* 81:345–349
- Irmak S, Dunford NT (2005) Policosanol contents and compositions of wheat varieties. *J Agric Food Chem* 53:5583–5586
- Kim HE, Choi YH, Choi KH, Park JS, Kim HS, Jeon JH, Heu MS, Shin DS, Lee JH (2012a) Metabolic classification of herb plants by NMR-based metabolomics. *J Korean Magn Reson Soc* 16:91–102
- Kim JK, Ha SH, Park SY, Lee SM, Kim HJ, Lim SH, Suh SC, Kim DH, Cho HS (2012b) Determination of lipophilic compounds in genetically modified rice using gas chromatography time-of-flight-mass spectrometry. *J Food Compos Anal* 25:31–38
- Kim JK, Park SY, Jung JY, Ha SH, Lim SH, Lee SM, Woo HJ, Park SU, Suh SC (2012c) The policosanols content and composition of Korean rice (*Oryza sativa* L.) cultivars. *Cereal Chem* 89:151–154
- Kim JK, Park SY, Na JK, Seong ES, Yu CY (2012d) Metabolite profiling based on lipophilic compounds for quality assessment of perilla (*Perilla frutescens*) cultivars. *J Agric Food Chem* 60:2257–2263
- Kim YB, Park SY, Thwe AA, Seo JM, Suzuki T, Kim SJ, Kim JK, Park SU (2013) Metabolomic analysis and differential expression of anthocyanin biosynthetic genes in white- and red-flowered buckwheat cultivars (*Fagopyrum esculentum*). *J Agric Food Chem* 61:10525–10533
- Lechner M, Reiter B, Lorbeer E (1999) Determination of tocopherols and sterols in vegetable oils by solid-phase extraction and subsequent capillary gas chromatographic analysis. *J Chromatogr A* 857:231–238
- Lin Y, Rudrum M, van der Wielen RP, Trautwein EA, McNeill G, Sierksma A, Meijer GW (2004) Wheat germ policosanols failed to lower plasma cholesterol in subjects with normal to mildly elevated cholesterol concentrations. *Metabolism* 53:1309–1314
- Pan O, Dai Y, Nuringtyas TR, Rianika RM, Schulte AE, Verpoorte R, Choi YH (2012) Investigation of the chemomarkers correlated with flower color in different organs of *Catharanthus roseus* using NMR-based metabolomics. *Phytochem* 25:66–74
- Park SY, Choi SR, Lim SH, Yeo Y, Kweon SJ, Bae YS, Kim KW, Im KH, Ahn SK, Ha SH, Park SU, Kim JK (2014) Identification and quantification of carotenoids in paprika fruits and cabbage, kale, and lettuce leaves. *J Korean Soc Appl Biol Chem* 57:355–358
- Piironen V, Lindsay DG, Miettinen TA, Toivo J, Lampi AM (2000) Plant sterols: biosynthesis, biological function and their importance to human nutrition. *J Agric Food Chem* 80:939–966
- Pongsuwan W, Fukusaki E, Bamba T, Yonetani T, Yamahara T, Kobayashi A (2007) Prediction of Japanese green tea ranking by gas chromatography/mass spectrometry-based hydrophilic metabolite fingerprinting. *J Agric Food Chem* 55:231–236
- Rusak DA, Brown LM, Martin SD (2003) Classification of vegetable oils by principal component analysis of FTIR spectra. *J Chem Educ* 80:541–543
- Schummer C, Delhomme O, Appenzeller BMR, Wennig R, Millet M (2009) Comparison of MTBSTFA and BSTFA in derivatization reactions of polar compounds prior to GC/MS analysis. *Talanta* 77:1473–1482
- Shepherd T, Dobson G, Verrall SR, Conner S, Griffiths DW, McNicol JW, Davies HV, Stewart D (2007) Potato metabolomics by GC-MS: what are the limiting factors? *Metabolomics* 3:475–488
- Steuer R, Kurths J, Fiehn O, Weckwerth W (2003) Interpreting correlations in metabolomic networks. *Biochem Soc Trans* 31:1476–1478
- Taylor JC, Rapport L, Lockwood GB (2003) Octacosanol in human health. *Nutrition* 19:192–195
- Van Pelt CK, Haggarty P, Brenna T (1998) Quantitative subfemtomole analysis of  $\alpha$ -tocopherol and deuterated isotopomers in plasma using tabletop GC/MS/MS. *J Agric Food Chem* 70:4369–4375