ARTICLE

Phytochemical profiles of Brassicaceae vegetables and their multivariate characterization using chemometrics

Kyoung Bok Lee[1] · Ye Jin Kim[1] · Hyo Jin Kim[1] ·
Jaehyuk Choi[1] · Jae Kwang Kim[1]

Abstract Twenty-eight metabolites were extracted from nine Brassicaceae of Korean origin (broccoli, Brussels sprouts, cabbage, Chinese cabbage, kale, kohlrabi, pak choi, radish sprouts, and red cabbage) and analyzed using gas chromatography–mass spectrometry and high-performance liquid chromatography. Principal components analysis (PCA), orthogonal projection to latent structure-discriminant analysis (OPLS-DA), Pearson's correlation analysis, hierarchical clustering analysis (HCA), and batch learning self-organizing map analysis (BL-SOM) were used to visualize metabolite pattern differences among Brassicaceae samples. The PCA score plots from the metabolic data sets provided a clear distinction between *Brassica* species and radish sprouts (genus *Raphanus* L.). Additionally, *B. oleracea* L. varieties were differentiated from *B. rapa* L. varieties by PCA and OPLS-DA score plots. HCA and BL-SOM of these metabolites clustered metabolites that are metabolically related. This study demonstrates that plants' characterization by multivariate statistical analysis using metabolic profiling allows distinguishing their phenotypes and identifying desired characteristics.

Keywords Batch learning self-organizing map analysis · Bioinformatics · Metabolic profiling · Metabolomics · Principal components analysis

✉ Jaehyuk Choi
    jaehyukc@inu.ac.kr

✉ Jae Kwang Kim
    kjkpj@inu.ac.kr

[1] Division of Life Sciences and Bio-Resource and
    Environmental Center, Incheon National University, Incheon,
    Republic of Korea

Introduction

Brassicaceae vegetables are widely recognized for their contribution to human nutrition and health benefits [1], as they contain many health promoting and potentially protective phytochemicals, including tocopherols, carotenes, plant sterols, and policosanols [2–4]. Although extensively consumed in Korea, few studies have considered the phytochemicals of Brassicaceae vegetables [5, 6]. However, studies on the qualitative and quantitative distribution of primary phytochemicals in Brassicaceae vegetables may help breeders to develop a germplasm with a high level of these phytochemicals.

Comprehensive chemical analysis seems to be the most reliable method for estimating food quality. Recently, with the development of various analytical techniques, many chemical data were generated for food identification and food quality determination. Plants' identification has been performed based on genetic and metabolome analyses. A study used simple sequence repeats (SSRs) to generate genetic maps in two *Brassica* spp. populations (*Brassica oleracea* L. and *B. rapa* L.) [7], but, to our knowledge, few investigations have compared metabolite profiles of Brassicaceae species to identify biomarkers for their discrimination. In fact, there is only one metabolic classification of Amaranthaceae, Asteraceae, Brassicaceae, and Malvaceae [8]. According to the previously mentioned paper by Kim et al. [8], the plants have been classified by metabolic components, including policosanol, phytosterol, amyrin, and tocopherol, but no studies have been done with Brassicaceae vegetables.

Chemometric pattern recognition techniques have been used for the discrimination of the geographical origin and variety of plants [9–11]. The application of multivariate

techniques to biological studies produces weighted combinations of the original variables that allow grouping them, which is often not evident via the classical univariate analytical approaches [12]. Group classification was initially performed using unsupervised principal component analysis (PCA)-based approaches. In previous studies, PCA and hierarchical clustering analysis (HCA) confirmed differences between rice and Chinese cabbage in terms of metabolite contents [10, 11]. Orthogonal projection to latent structure-discriminant analysis (OPLS-DA) has been used to maximize differences between samples and to identify the variables responsible for their differentiation, and it has provided a clear discrimination between *Gladiolus* sp. genotypes [13]. In addition, the relationship between two metabolites can be obtained by Pearson correlation analysis, and the correlations can be examined to acquire information on metabolic associations. In our previous study, Pearson's correlation analysis was used to identify metabolic links in rice seeds [14]. A batch learning self-organizing map (BL-SOM) was used in another study to analyze differences in metabolite levels between *Arabidopsis thaliana* cells cultured under salt stress over time [9]. The resulting map was used to interpret the metabolic networks. In addition, BL-SOM has been used to study transcriptome and metabolome data in plants [15]. However, there are few published data on the discrimination of Brassicaceae vegetables using the chemometric pattern recognition techniques.

This study evaluated lipophilic compounds in nine Brassicaceae species of Korean origin. Twenty-eight compounds, including policosanol, phytosterol, amyrin, carotenoids, and tocopherol, were analyzed and classified from broccoli, red cabbage, cabbage, Brussels sprouts, Chinese cabbage, kale, kohlrabi, pak choi, and radish sprouts. PCA, OPLS-DA, Pearson's correlation analysis, HCA, and BL-SOM were used in the present study to visualize chemical differences between the nine species analyzed. Through this approach, we were able to discriminate the Brassicaceae vegetables according to their species. We emphasize that this multivariate statistical analysis using metabolic profiling is a powerful tool for the assessment of quality and for the discrimination of species.

Materials and methods

Samples and chemicals

The edible parts of nine Brassicaceae were purchased from supermarket in the Incheon city, Korea in 2015, and included leaves (Brussels sprouts, cabbage, Chinese cabbage, kale, pak choi, and red cabbage), flowers (broccoli), sprouts (radish sprouts), and roots (kohlrabi) (three

Fig. 1 Representative total ion chromatograms (TIC) (**A**) of lipophilic standards and TIC (**B**) and selected ion monitoring (SIM) chromatogram (**C**) of lipophilic metabolites extracted from radish sprouts as trimethylsilyl derivatives. The selected compounds in radish sprouts are displayed in (**C**). Inverted triangle represents peak of target compound. 1 C20, eicosanol; 2 C21, heneicosanol; 3 C22, docosanol; 4 C23, tricosanol; 5 C24, tetracosanol; 6 δ-tocopherol; 7 5α-cholestane (internal standard); 8 C26, hexacosanol; 9 β-tocopherol; 10 γ-tocopherol; 11 C27, heptacosanol; 12 δ-tocotrienol; 13 C28, octacosanol; 14 γ-tocotrienol; 15 α-tocopherol; 16 cholesterol; 17 brassicasterol; 18 α-tocotrienol; 19 campesterol; 20 C30, triacontanol; 21 stigmasterol; 22 β-sitosterol; 23 β-amyrin; 24 α-amyrin

biological replicates). Each sample was divided into two portions, one for gas chromatography-mass spectrometry (GC–MS) analysis and one for high-performance liquid chromatography (HPLC) analysis. Each sample was freeze-dried at − 70 °C for 72 h and then crushed using a mortar and pestle. The resulting powder was stored at − 80 °C until extraction. Pyridine, ascorbic acid, and *N*-methyl-*N*-trimethylsilyl trifluoroacetamide (MSTFA) were obtained from Sigma-Aldrich (St. Louis, MO, USA). All the other chemicals used in this study were reagent grade, unless otherwise stated.

Lipophilic compound extraction and analysis

Nineteen types of lipophilic compounds were extracted as in the previously described method [8]. Briefly, each powdered sample (0.05 g) was added to 3 mL 0.1% ascorbic acid in ethanol (w/v) and 0.005 mL 5α-cholestane (internal standard; 100 μg mL^{-1}). After saponification using 120 μL 80% potassium hydroxide (w/v), samples were immediately put on ice for 5 min, and distilled water and hexane (1.5 mL of each) were added. After re-extraction using hexane, the hexane layer was then concentrated using a centrifugal concentrator (CC-105, TOMY, Tokyo, Japan), and mixed with MSTFA and pyridine (30 μL of each) for 30 min, at 60 °C, and under 1200×*g* using a thermomixer (model 5355, Eppendorf AG, Hamburg, Germany). This mixture was analyzed by GC–MS in a GCMS-QP2010 Ultra system equipped with an auto sampler AOC-20i (both Shimadzu, Kyoto, Japan) prepared with the Rtx-5MS column (0.25 mm diameter and 0.25 μm thickness, 30 m length; Agilent, Palo Alto, CA, USA). The temperature of injection, interface, and ion source was 290, 280, and 230 °C, respectively. The flow rate of carrier gas (helium) was 1.0 mL min^{-1}. The GC program was set for 2 min at 150 °C, followed by an increase up to 320 °C, with a ramping rate of 15 °C min^{-1} and hold time of 10 min. The volume of the injected samples was 1.0 μL, and the split ratio was 10:1. The mass spectra were analyzed using the Lab solutions GCMS solution software version 4.11 (Shimadzu, Kyoto, Japan).
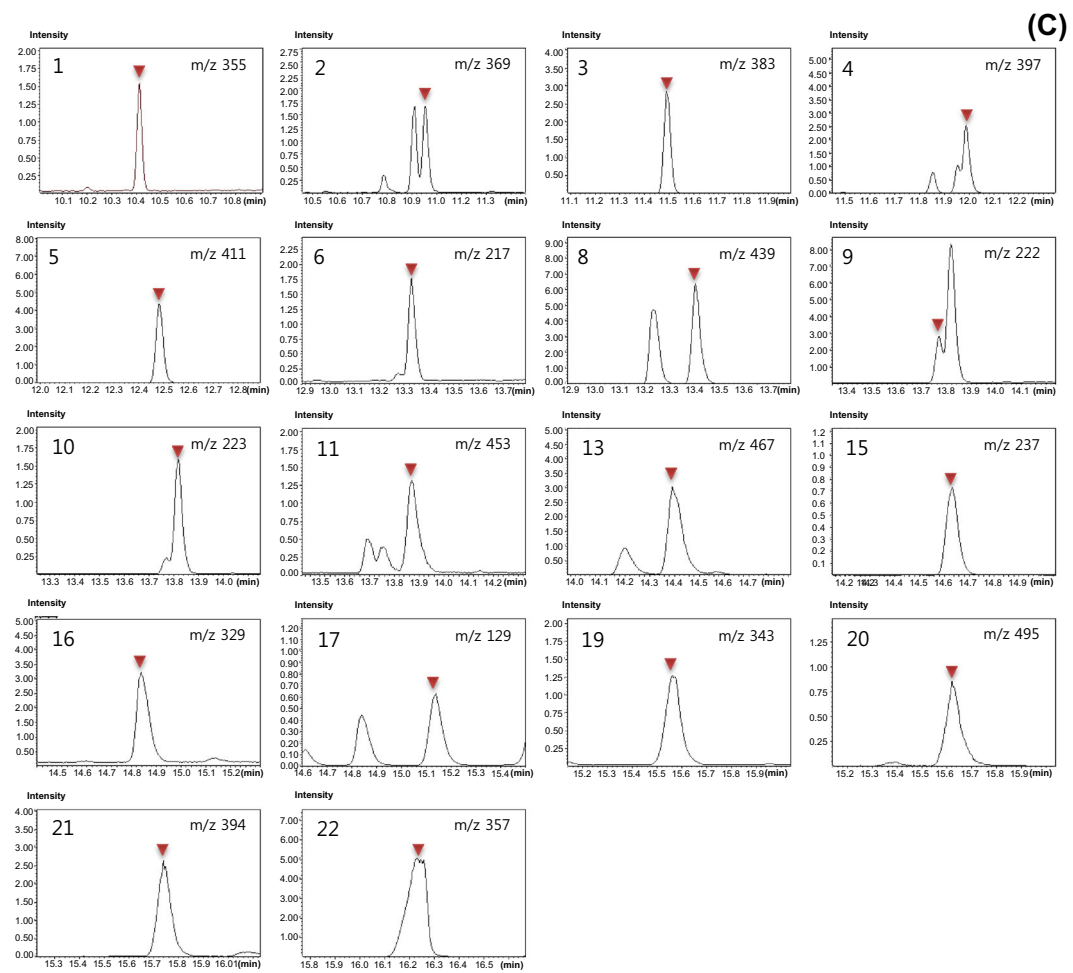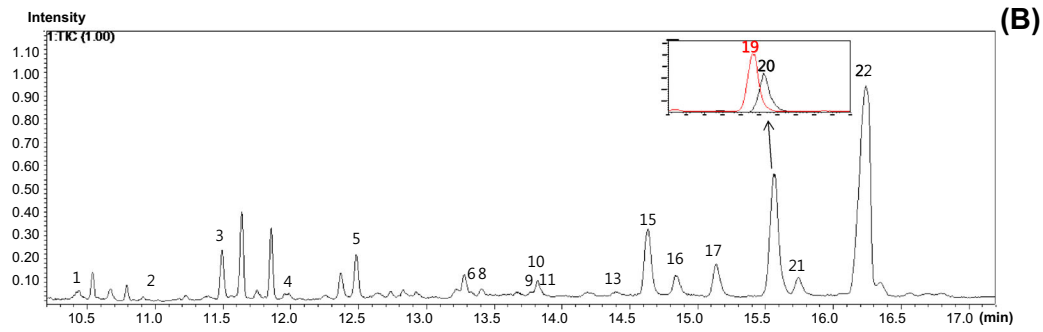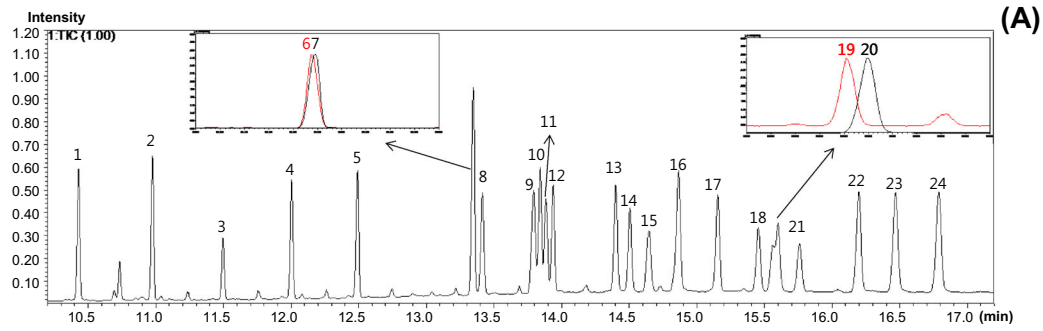
Carotenoids were extracted as previously described [16]. In brief, 3 mL 0.1% ascorbic acid in ethanol (w/v) were added to each sample (0.1 g), vortexed for 20 s, and kept for 5 min in a water bath at 85 °C. After saponification, samples were left on ice for 5 min, and 0.1 mL of internal standard (β-apo-8′-carotenal in ethanol; 25 µg mL^{-1}), and 1.5 mL distilled water were added to each tube. Hexane (1.5 mL) was then added, and samples were centrifuged at $1200 \times g$ for 5 min. The carotenoid layer was then dried under a nitrogen stream and dissolved in 0.25 mL methanol/dichloromethane (50:50, v/v). Resulting samples (20 µL each) were analyzed by HPLC in an Agilent 1100 instrument (Agilent Technologies, Massy, France), equipped with a photodiode array detector set at 450 nm. The column, mobile phase, and elution program were previously described [16].

Statistical analyses

All analyses were performed in triplicate, at least. All data files were scaled for multivariate analysis using unit variance. Brassicaceae data were analyzed by PCA (SIMCA-P version 13, Umetrics, Umeå, Sweden) to visualize relationships among vegetables and compounds. The patterns observed in the PCA score plots and Brassicaceae samples' loading plots were used to explain the dispersion of vegetables in the diagrams. The OPLS-DA model was calculated in SIMCA-P version 13 using phytochemicals data and species data as the Y matrix. Pearson's correlation analysis was performed in SAS version 9.4 software package (SAS Institute, Cary, NC, USA), and Multi-Experiment Viewer version 4.9.0 was used for HCA and visualization as heat map. A simple self-organizing map [SOM; (http://kanaya.naist.jp/SOM/]) was used in BL-SOM analysis, in which correlation and clustering were performed among the levels of 28 metabolites with standardization procedures.

Results and discussion

Lipophilic metabolite profiling of nine Brassicaceae vegetables

Metabolomics allows visualizing differences in metabolite patterns among diverse foods. In the present study, GC–MS and HPLC were used to identify and quantify 28 lipophilic compounds in the samples of the nine Brassicaceae species analyzed. The carotenoids violaxanthin, antheraxanthin, lutein, zeaxanthin, β-cryptoxanthin, 13Z-β-carotene, E-β-carotene, α-carotene, and 9Z-β-carotene were detected by HPLC through the co-elution and retention time with standards as seen in our previous study [16]. Additionally, 19 types of lipophilic compounds in most vegetables were detected by GC–MS analysis. Quantification was performed using selected ions (Fig. 1).

Evaluation and classification of nine Brassicaceae vegetables using chemometrics

Multivariate analysis using metabolites' data is particularly useful approach to find underlying structures in complicated biological systems [17–19]. Differences in metabolite levels were evaluated through a PCA, which explored the structure of the data obtained from GC–MS and HPLC, as evidenced in the PCA scores plot (Fig. 2). The scores of principal component 1 (PC 1) and 2 (PC 2), plotted in the abscissa and ordinate, respectively, accounted for 69.8% of the total variance within all species data (Fig. 2A, B). Radish sprouts belong to genus *Raphanus* L., whereas the other species belong to genus *Brassica* L., a difference that was successfully captured in PC 1. The metabolites loaded in PC 1 and PC 2 were compared to investigate which contributed the most to the observed pattern. The predominant metabolite in PC 1 was α-tocopherol, although other 26 metabolites (excluding amyrins) had positive loading scores in PC 1. As a result, radish sprouts, which were strongly affected by α-tocopherol and had the highest concentrations of the other 26 metabolites, appeared separated in the plot.

To verify if PCA could be used as a tool for distinguishing *Brassica* spp., this technique was applied to the datasets originated from the leaves of Brussels sprouts, cabbage, Chinese cabbage, kale, pak choi, and red cabbage (Fig. 3). The first two PCs accounted for 75.5% of the total variance, and *Brassica* spp. were separated into two groups along PC 2 (Fig. 3A, circled within the dotted line). The predominant contributors in PC 2 were α-amyrin and cholesterol, which separated *B. oleracea* varieties from *B. rapa* varieties. Loading plots (Fig. 3B) indicated that *B. rapa* varieties had higher contents of carotenoids and phytosterols, except stigmasterol and campesterol, than *B. oleracea* varieties. Similarly, Brussels sprouts, cabbage, and red cabbage were also clearly grouped by PCA.

OPLS-DA can be used to maximize differences between samples as well as to identify markers for their classification. In this study, all samples were clearly separated in the OPLS-DA score plots (Fig. 4A). The quality of the OPLS-DA model can be explained by goodness of fit (R^2) and predictive ability (Q^2) values, which were 0.747 and 0.937 in our model; in addition, our model seemed to have an excellent ability as its $Q^2 > 0.9$ [20]. External validation aims to address the accuracy of a model in samples from different species. To confirm the performance of our OPLS-DA model, the four samples were randomly left as a test data set and the OPLS-DA was established with training samples (Fig. 4B). The R_X^2 and Q^2 values of this
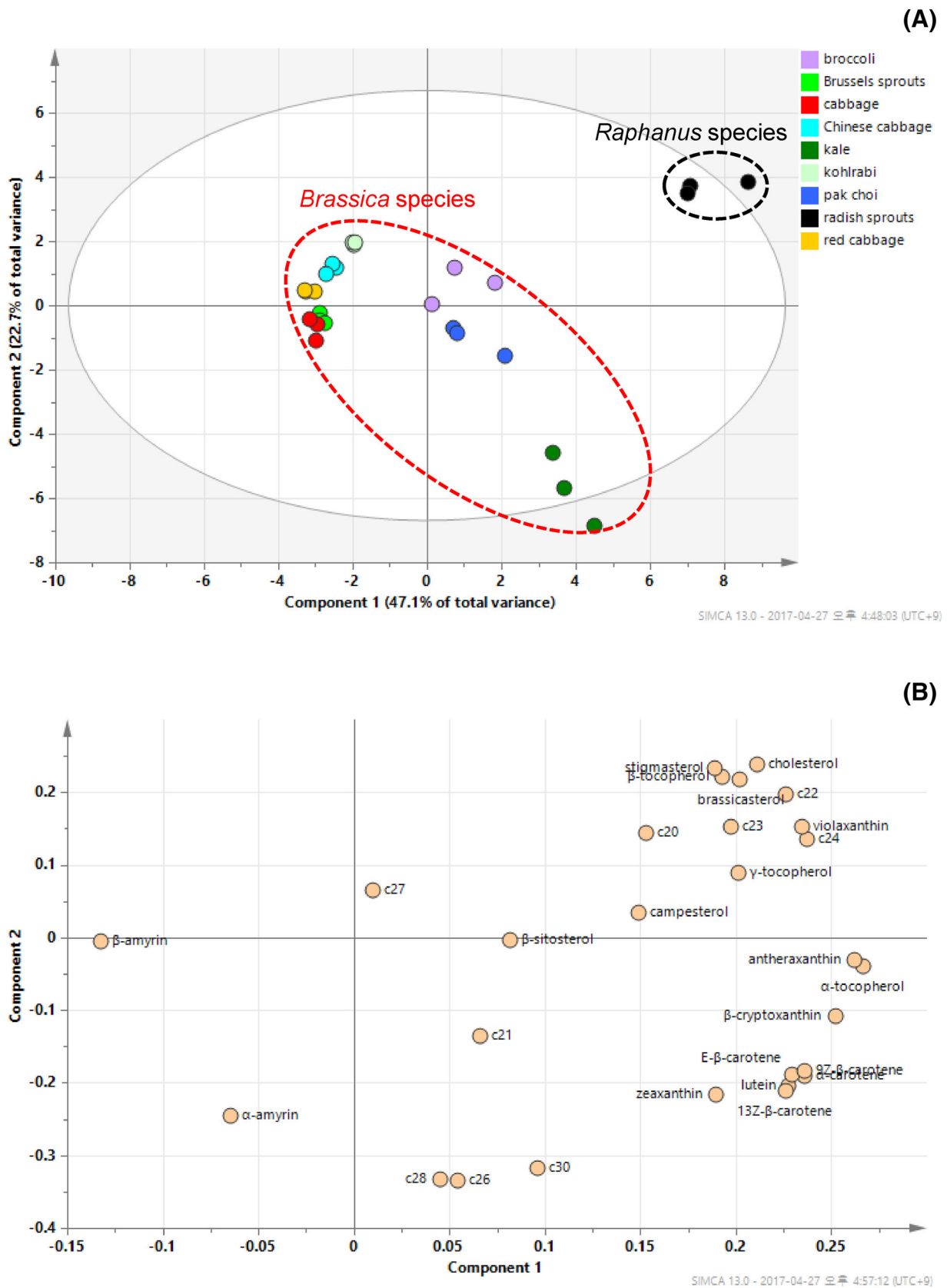
Fig. 2 (**A**) Scores and (**B**) loading plots of principal components 1 and 2 of the principal components analysis (PCA) results obtained for the metabolites of nine Brassicaceae vegetables
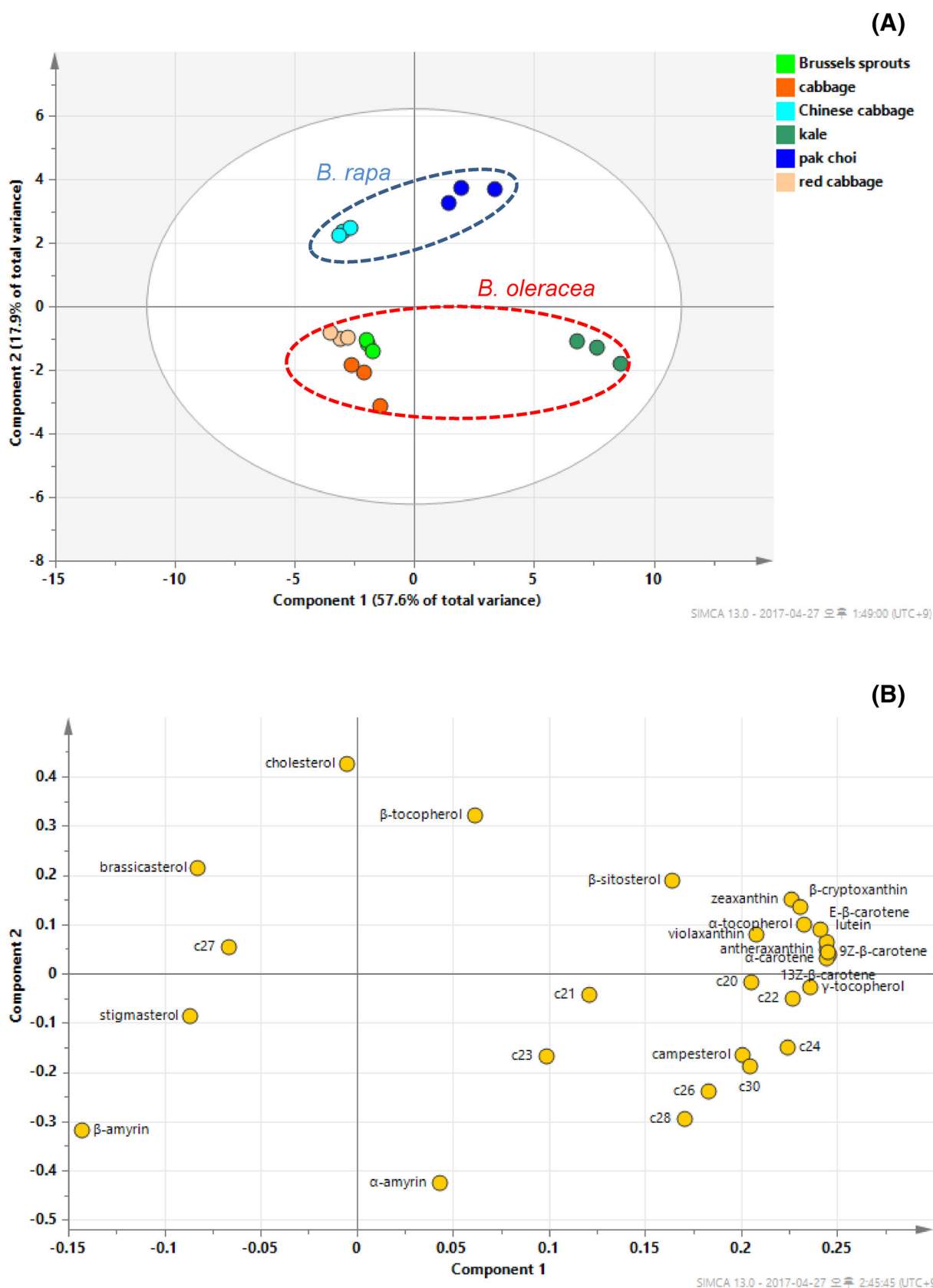
(A)

(B)

Fig. 3 (**A**) Scores and (**B**) loading plots of principal components 1 and 2 of the principal components analysis (PCA) results obtained for the metabolites of six *Brassica* vegetables
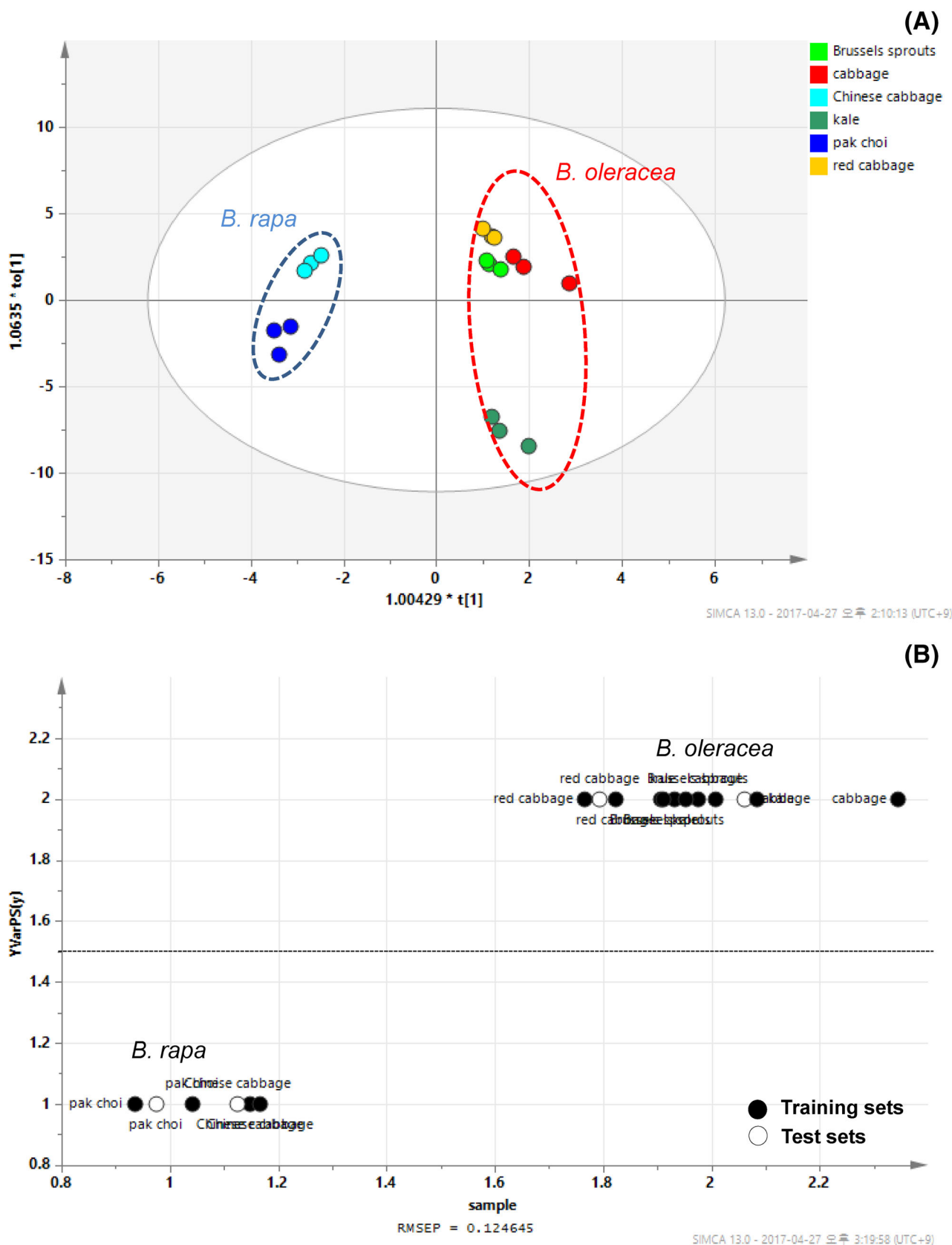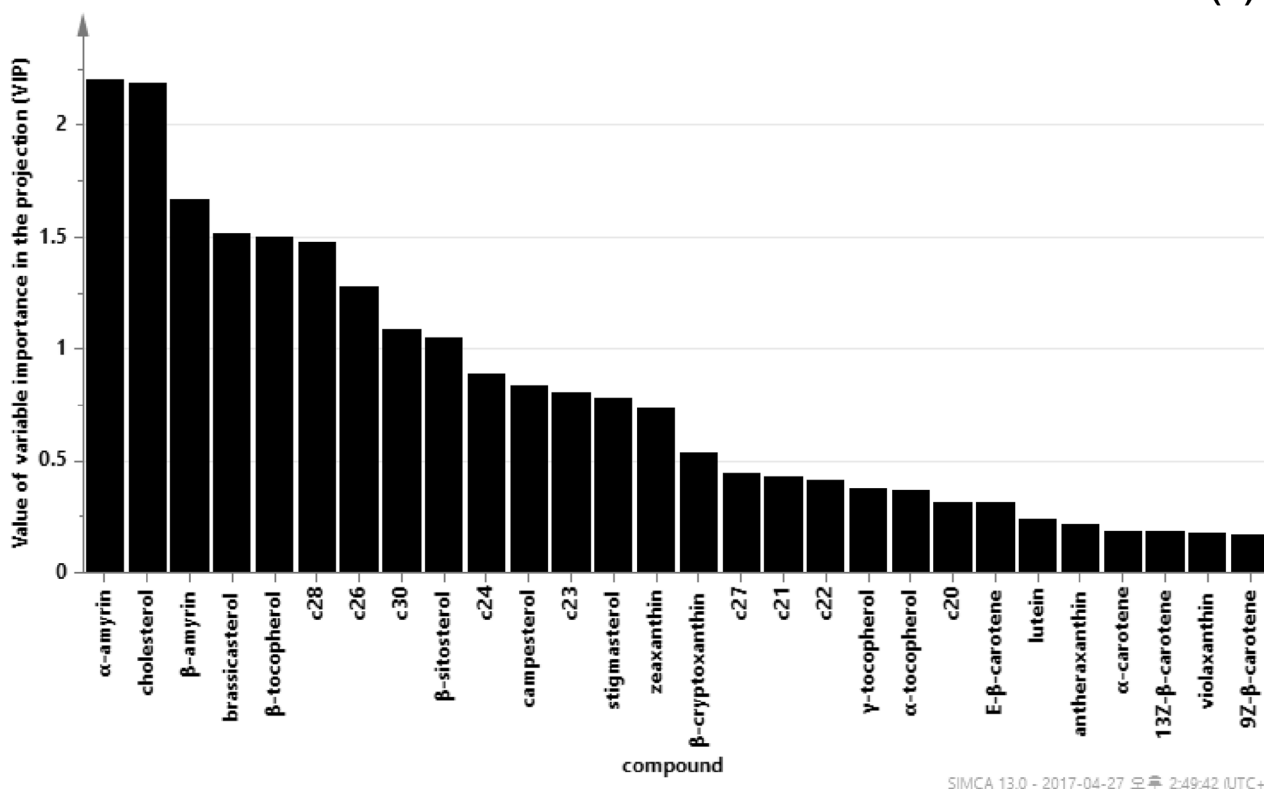
Fig. 4 (**A**) Score plots and (**B**) external validation test of the orthogonal projection to latent structure-discriminant analysis (OPLS-DA) model derived from the metabolite data of six *Brassica* vegetables, and **C** variable importance in the projection (VIP)

(C)

Fig. 4 continued

model were 0.829 and 0.853, respectively, and $Q^2 > 0.5$ indicates a good predictive ability. The variables important in the projection (VIP) value explains the contribution of variables to the projection, and VIP > 1 is used as a criterion to identify the most important variables to the model [19]. In the present study, nine metabolites, namely α- and β-amyrins, cholesterol, brassicasterol, β-tocopherol, octacosanol, hexacosanol, triacontanol, and β-sitosterol, presented VIP > 1, indicating their important contribution to discriminate between *B. oleracea* and *B. rapa* varieties (Fig. 4C).

Pearson's correlation analysis was performed to identify relationships between metabolites, which were classified according to a color scale, with red indicating a positive correlation and green a negative correlation, and color intensity indicating the strength of the correlation (Fig. 5). The metabolite-to-metabolite correlation matrix resulting from Pearson's correlations was used as input for the HCA of the 28 metabolites, in which those with the highest correlations were clustered. These analyses identified two groups (boxed within dotted lines in Fig. 5): one comprising c21, c26, c27, c28, c30, and amyrins, and another composed mainly of phytosterols, carotenoids, tocopherols, and policosanols with positive correlations. Phytosterols are biosynthesized by the mevalonate pathway [21, 22],

while the non-mevalonate pathway, also called the mevalonic acid (MVA)-independent pathway, promotes the synthesis of carotenoids and tocopherols [22, 23]. These two pathways have common precursors, such as isopentenyl diphosphate (IPP), geranylgeranyl diphosphate (GPP), and farnesyl diphosphate (FPP), which might explain the positive correlations found within the second group of metabolites. Among the three β-carotenes, E-β-carotene was the most abundant, and it was highly correlated with 9Z-β-carotene ($r = 0.9883$, $p < 0.0001$) and 13Z-β-carotene ($r = 0.9884$, $p < 0.0001$).

Batch learning SOM analysis was developed by Kanaya et al. [24] to replace self-organizing map (SOM) analysis, which had a low reproducibility. BL-SOM is a multivariate statistical analysis method that uses existing SOM matrices from PCA datasets [9], allowing visualizing relative amounts of metabolites and differences between samples in a large dataset with high reproducibility [24–27]. In the present study, the SOMs of nine Brassicaceae vegetables were obtained (Fig. 6A), each comprising neurons colored according to the amount of metabolites. A 6 × 5 matrix shows the patterns of phytochemical levels in the different species (Fig. 6B). For example, in the SOM of kale (matrix number 5), the upper-right neurons, corresponding to carotenoids except violaxanthin, are indicated
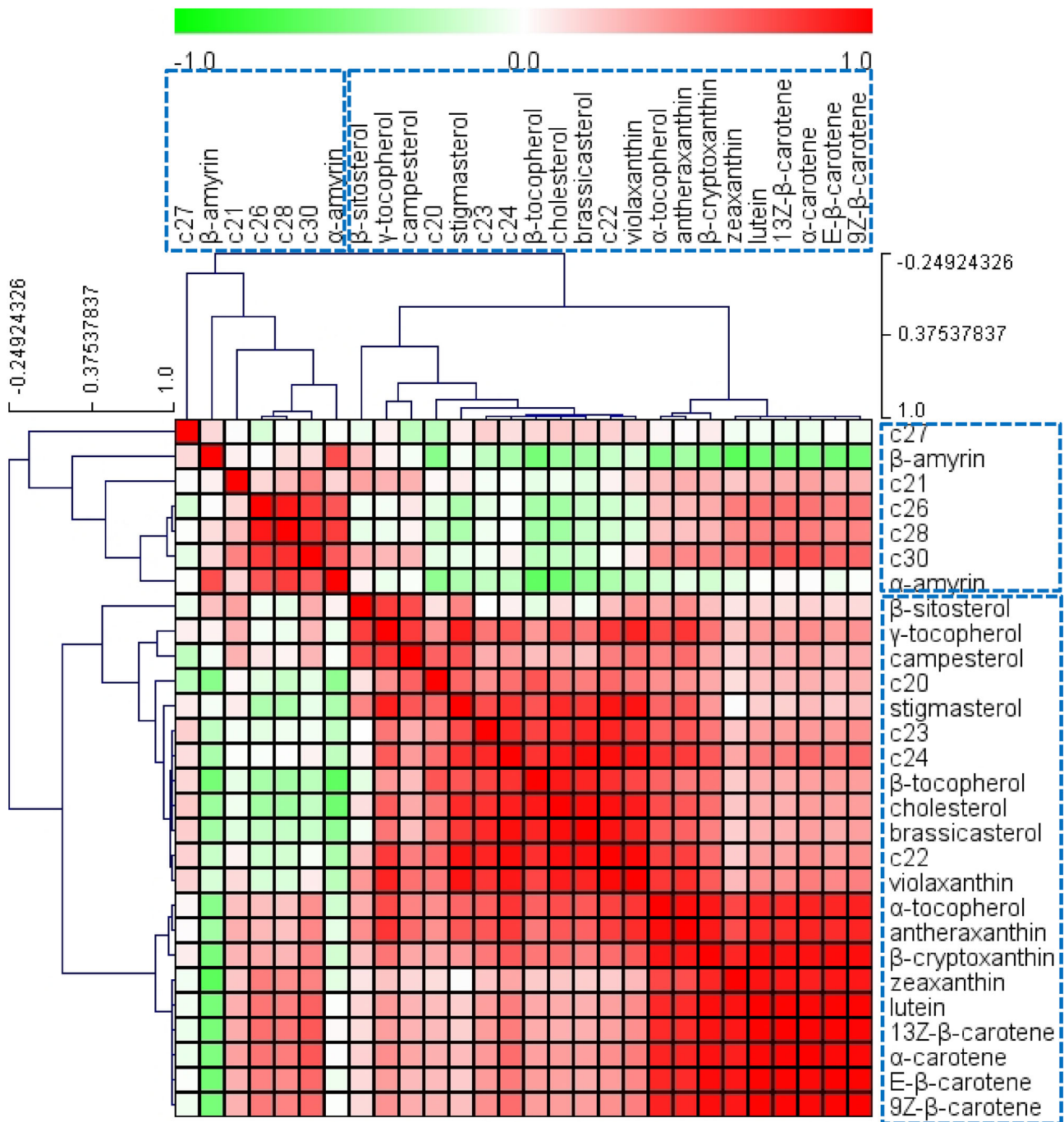
Fig. 5 Correlation matrix of the 28 metabolites identified from the nine Brassicaceae vegetables analyzed. Each square indicates the Pearson's correlation coefficient obtained between a pair of compounds, and the intensity of green or red (negative or positive correlation, respectively) corresponds to the value of the correlation coefficient. Hierarchical clusters are indicated at the top and left of the figure

in red, meaning kale has higher contents of these metabolites than the other eight Brassicaceae. The distance among neurons within SOMs also indicates the correlation between metabolites: Closely located neurons have high positive correlations (Fig. 6C). BL-SOM of these metabolites clustered metabolites that are metabolically related. Apart from violaxanthin, all carotenoids were clustered in the upper-middle to the upper-right, amyrins were clustered on the bottom-right, and phytosterols were clustered on the left and at the bottom-center. Thus, BL-SOM results allowed visualization of correlations among metabolites and characterizing samples.
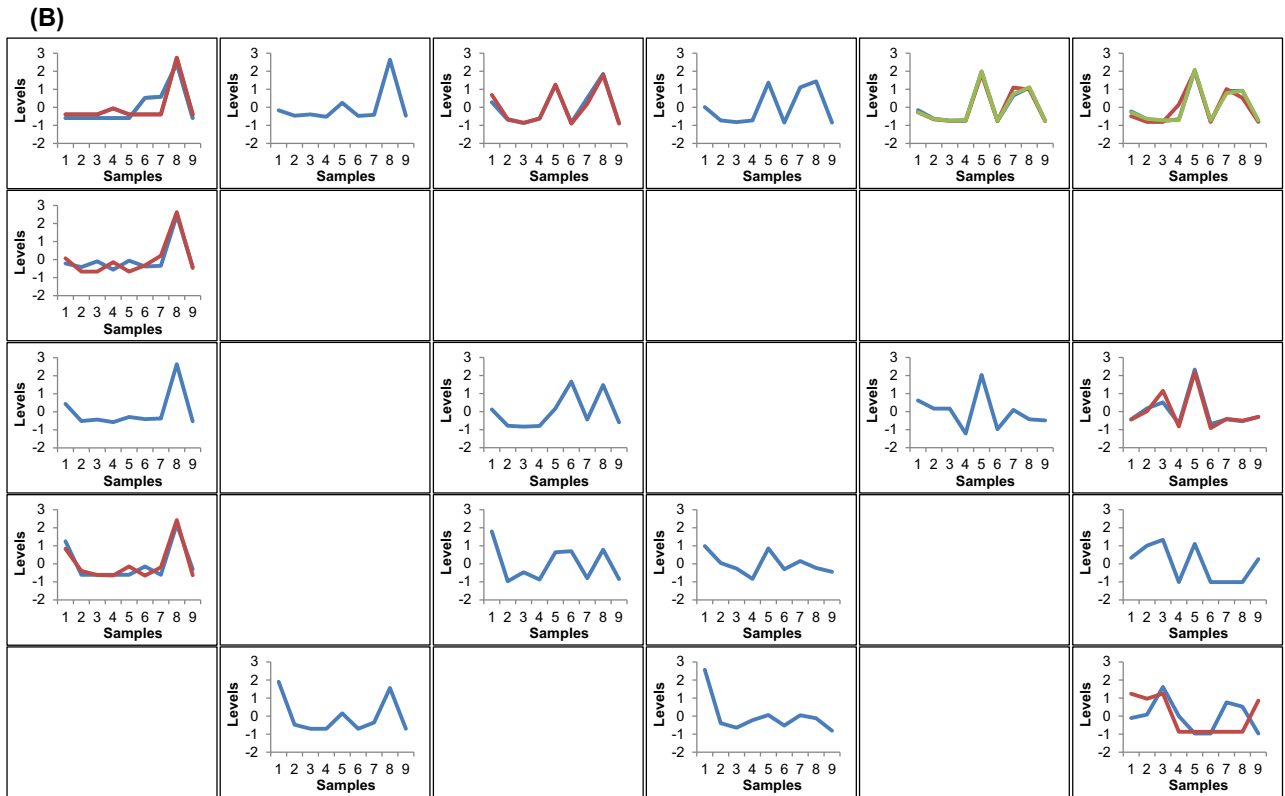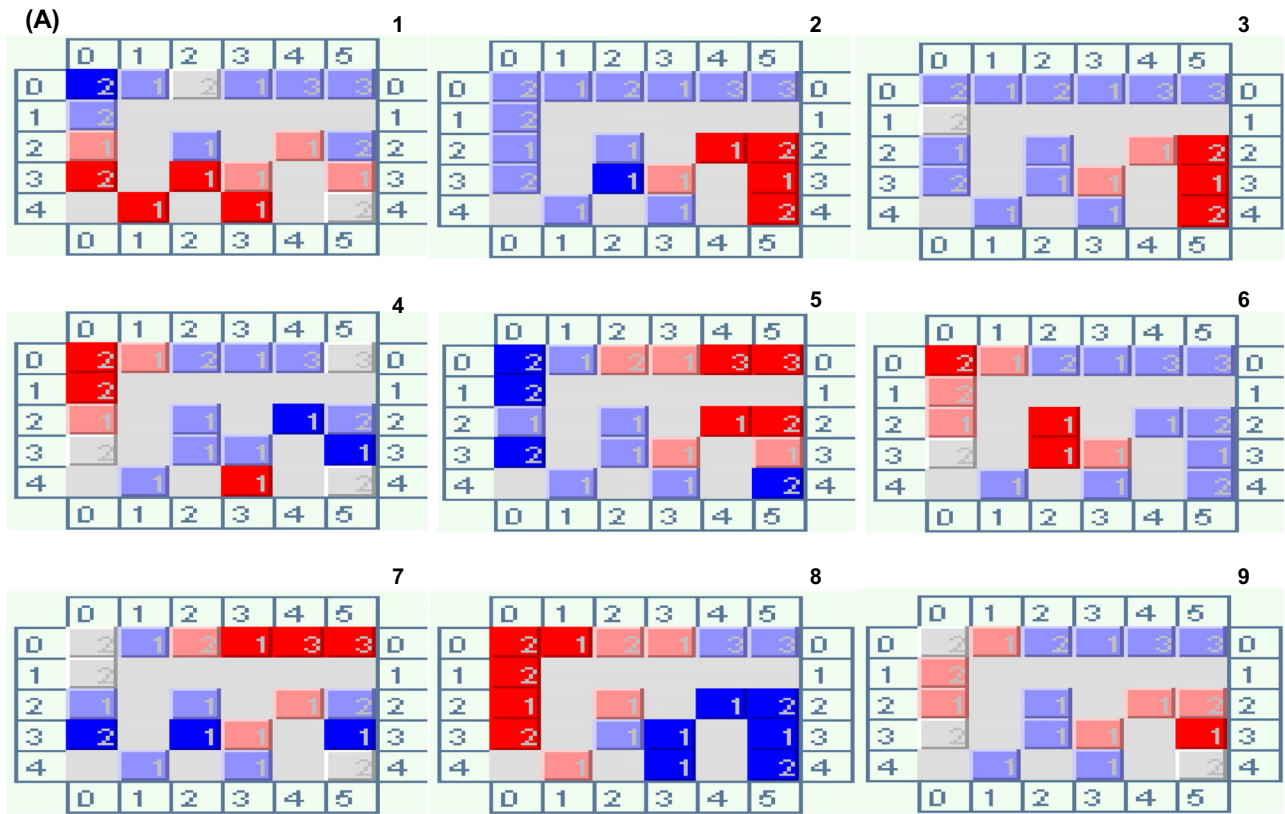
(A)

(B)

◄**Fig. 6** Batch learning self-organizing map (BL-SOM) analysis (**A**) based on the 6 × 5 SOMs derived from the principal components analysis (PCA) of the 28 metabolites. The SOMs show metabolites' clustering (numbers), and neurons within SOMs are arranged in two-dimensional lattice matrices. Color of SOMs corresponds to the relative amounts of compounds: red (most increased), pink (increased), pale blue (decreased), and blue (most decreased). 1 broccoli; 2 Brussels sprouts; 3 cabbage; 4 Chinese cabbage; 5 kale; 6 kohlrabi; 7 pak choi; 8 radish sprouts; 9 red cabbage. The amounts of compounds in Brassicaceae (**B**) and the metabolite clusters (**C**) displayed in the 6 × 5 matrices obtained in the BL-SOM analysis. In (**B**), Brassicaceae vegetables are displayed in the X-axis and standardized compounds levels in the Y-axis

Composition and content of lipophilic metabolites in nine Brassicaceae vegetables

Tocopherols, phytosterols, policosanols, and amyrins were quantified in a previous study [8]. Previously, the c26 (hexacosanol) was reported as the main policosanol in leaf samples [8, 11]. In our study, this compound was the highest in leaf samples [12.72–388.06 µg g^{-1} of dry weight (DW)] (Table 1). Furthermore c22 (docosanol) was the highest in flower, root, and sprout samples (17.65–223.00 µg g^{-1} of DW). In general, kale and radish sprouts had high policosanols contents. Campesterol and β-sitosterol were the main phytosterols, and α-tocopherol was the main tocopherol in Brassicaceae [8, 11]. The results obtained in the present study confirmed that campesterol (542.89–1182.18 µg g^{-1} of DW), β-sitosterol (638.76–1900.15 µg g^{-1} of DW), and α-tocopherol (4.25–254.32 µg g^{-1} of DW) contents were high in the nine Brassicaceae, especially in broccoli, which contained high levels of phytosterols, and radish sprouts, which contained high levels of tocopherols (Table 1). Lutein (0.46–251.81 µg g^{-1} of DW) and E-β-carotene (0.66–243.61 µg g^{-1} of DW) were the predominant carotenoids (Table 2). In the Chinese cabbage, for example, lutein comprised about 59.58% of the total carotenoids, which is in agreement with the previous findings [11] where lutein was found in plant leaves at a higher ratio than other components. Carotenoids play an essential role as accessory light-harvesting pigments [28] and, therefore, the low levels of carotenoids found in kohlrabi, in which the edible part was the root, were expected. Total carotenoids levels were much higher in kale, pak choi, and radish sprouts than in other vegetables (Table 2).

In conclusion, 28 metabolites were identified in nine Brassicaceae spp. samples by GC–MS and HPLC, including amyrins, carotenoids, tocopherols, phytosterols, and policosanols. To visualize the several components, PCA, OPLS-DA, HCA, and BL-SOM were used. The PCA separated the nine species into two groups, *Brassica* sp. and *Raphanus* sp. (Fig. 2). A separation between *B. rapa* and *B. oleracea* was observed in PCA and OPLS-DA plots, and several variables were identified as candidate biomarkers that could be used in *Brassica* sp. authentication (Figs. 3, 4). Pearson's correlations and HCA indicated a positive correlation between carotenoids, phytosterols, and tocopherols, and their clustering (Fig. 5). BL-SOM demonstrated patterns in the relative quantity changes of each metabolite within the different species and allowed discriminating Brassicaceae vegetables by using relative differences in the amounts of components (Fig. 6). These

(C)

	0	1	2	3	4	5	
0	β-tocopherol brassicasterol	c24	α-tocopherol antheraxanthin	β-cryptoxanthin	α-carotene E-β-carotene 9Z-β-carotene	lutein zeaxanthin 13Z-β-carotene	0
1	c23 cholesterol						1
2	c22		c20		c30	c26 c28	2
3	stigmasterol violaxanthin		campesterol	c21		α-amyrin	3
4		γ-tocopherol		β-sitosterol		c27 β-amyrin	4
	0	1	2	3	4	5	

Fig. 6 continued

Table 1 Composition and content ($\mu g\ g^{-1}$) of lipophilic compounds in nine vegetables

Compound	Broccoli	Brussels sprouts	Cabbage	Chinese cabbage	Kale	Kohlrabi	Pak choi	Radish sprouts	Red cabbage
Policosanols									
c20	3.70 ± 0.58	2.75 ± 0.25	2.70 ± 0.18	2.74 ± 0.27	3.76 ± 0.43	5.31 ± 0.97	3.11 ± 0.23	5.12 ± 0.41	2.95 ± 0.27
c21	2.91 ± 0.25	2.43 ± 0.20	2.27 ± 0.77	1.98 ± 0.38	2.84 ± 0.81	2.26 ± 0.35	2.48 ± 0.43	2.29 ± 0.53	2.18 ± 0.44
c22	74.28 ± 18.12	10.54 ± 0.59	15.65 ± 2.48	6.27 ± 1.16	25.85 ± 2.13	17.65 ± 5.55	19.78 ± 1.99	223.00 ± 8.82	9.12 ± 3.83
c23	1.67 ± 1.72	1.55 ± 1.03	2.51 ± 0.67	1.12 ± 0.31	2.65 ± 1.29	1.64 ± 0.28	1.78 ± 1.12	10.26 ± 4.25	1.60 ± 0.21
c24	14.07 ± 3.04	5.67 ± 0.51	7.52 ± 1.76	3.70 ± 0.40	26.51 ± 2.10	5.21 ± 0.11	6.75 ± 0.59	96.13 ± 19.40	5.43 ± 0.77
c26	43.06 ± 14.38	119.11 ± 27.72	161.45 ± 28.67	12.72 ± 3.28	388.06 ± 166.03	9.86 ± 3.17	43.82 ± 8.62	29.57 ± 7.87	60.56 ± 12.26
c27	4.09 ± 1.87	5.00 ± 1.28	12.40 ± 8.32	4.61 ± 1.34	ND	ND	8.25 ± 2.26	7.15 ± 0.57	ND
c28	26.17 ± 2.54	43.13 ± 9.57	84.33 ± 14.90	12.18 ± 3.14	121.51 ± 19.32	8.87 ± 2.37	27.37 ± 5.81	23.70 ± 3.18	31.34 ± 3.51
c30	48.48 ± 6.69	36.59 ± 3.74	36.75 ± 12.88	ND	86.33 ± 30.86	6.30 ± 1.23	34.57 ± 11.52	20.73 ± 1.80	19.28 ± 1.46
Vitamin E									
α-Tocopherol	109.75 ± 19.22	21.31 ± 1.93	4.43 ± 0.78	26.47 ± 13.31	196.53 ± 5.42	6.52 ± 0.52	131.53 ± 46.72	254.32 ± 26.46	4.25 ± 2.62
β-Tocopherol	NDb	ND	ND	ND	ND	2.27 ± 0.15	2.38 ± 0.40	6.11 ± 0.23	ND
γ-Tocopherol	30.88 ± 7.05	2.65 ± 0.39	ND	ND	10.15 ± 1.22	ND	4.06 ± 1.15	26.81 ± 2.29	ND
Sterol									
Cholesterol	12.81 ± 2.25	ND	ND	9.17 ± 2.11	ND	6.01 ± 0.55	15.37 ± 2.28	57.51 ± 1.58	3.52 ± 0.25
Brassicasterol	ND	ND	ND	11.75 ± 1.40	ND	ND	ND	111.85 ± 7.73	ND
Campesterol	1182.18 ± 129.41	542.89 ± 26.92	659.12 ± 113.17	565.56 ± 74.24	915.20 ± 41.98	929.22 ± 124.68	579.71 ± 27.35	947.83 ± 23.50	569.49 ± 38.54
Stigmasterol	102.97 ± 29.89	ND	ND	ND	ND	25.94 ± 2.80	ND	156.23 ± 21.52	16.94 ± 2.72
β-Sitosterol	1900.15 ± 181.69	795.48 ± 64.15	705.02 ± 132.79	858.19 ± 175.87	965.74 ± 34.84	748.28 ± 68.08	959.57 ± 48.03	899.45 ± 41.92	638.76 ± 31.17
Terphen									
α-Amyrin	15.08 ± 2.24	22.65 ± 1.19	26.30 ± 7.32	ND	23.82 ± 1.14	ND	ND	ND	14.27 ± 0.77
β-Amyrin	7.16 ± 0.52	6.23 ± 0.73	7.18 ± 1.38	ND	ND	ND	ND	ND	5.85 ± 0.35

Each value is the mean of three replications ± standard deviation

[a] c20, eicosanol; c21, heneicosanol; c22, docosanol; c23, tricosanol; c24, tetracosanol; c26, hexacosanol; c27, heptacosanol; c28, octacosanol; c30, triacontanol

[b] ND not detected

Table 2 Composition and content ($\mu g\ g^{-1}$) of carotenoids in nine vegetables

Compound	Broccoli	Brussels sprouts	Cabbage	Chinese cabbage	Kale	Kohlrabi	Pak choi	Radish sprouts	Red cabbage
Violaxanthin	3.01 ± 0.45	1.09 ± 0.04	0.73 ± 0.07	0.69 ± 0.06	1.47 ± 0.13	0.66 ± 0.05	1.41 ± 0.42	5.52 ± 0.71	0.71 ± 0.03
Antheraxanthin	0.76 ± 0.08	0.12 ± 0.02	ND	0.14 ± 0.04	1.02 ± 0.06	ND	0.52 ± 0.19	1.27 ± 0.13	ND
Lutein	49.41 ± 8.82	11.23 ± 0.92	1.98 ± 0.80	9.83 ± 3.38	251.81 ± 13.66	0.46 ± 0.15	151.11 ± 40.60	151.98 ± 32.06	0.71 ± 0.56
Zeaxanthin	ND[a]	ND	ND	0.69 ± 0.02	2.04 ± 0.09	ND	1.32 ± 0.40	0.96 ± 0.07	ND
β-Cryptoxanthin	1.91 ± 0.23	0.93 ± 0.10	0.80 ± 0.01	0.93 ± 0.08	3.70 ± 0.25	0.78 ± 0.01	3.35 ± 0.85	3.79 ± 0.28	0.78 ± 0.01
13Z-β-Carotene	15.56 ± 1.62	3.10 ± 0.72	0.57 ± 0.23	0.66 ± 0.65	96.73 ± 13.74	0.22 ± 0.05	52.15 ± 11.76	56.70 ± 12.13	0.34 ± 0.24
α-Carotene	3.16 ± 0.54	0.73 ± 0.19	0.14 ± 0.04	0.31 ± 0.22	14.21 ± 3.63	0.06 ± 0.01	7.36 ± 1.90	9.67 ± 1.66	0.10 ± 0.06
E-β-carotene	46.03 ± 8.28	10.19 ± 0.70	1.79 ± 0.84	2.42 ± 1.71	243.61 ± 18.45	0.66 ± 0.19	172.78 ± 41.56	162.34 ± 28.10	0.80 ± 0.65
9Z-β-carotene	15.34 ± 2.31	3.61 ± 0.23	0.61 ± 0.22	0.82 ± 0.49	86.48 ± 1.57	0.26 ± 0.07	47.17 ± 15.09	58.74 ± 9.45	0.28 ± 0.19

Each value is the mean of three replications ± standard deviation

[a] *ND* not detected

multivariate statistical analyses techniques can be used to establish plant metabolite profiles, which can later be used to select cultivars with a specific compound. Plant identification and differentiation at the species and population levels are important to plant scientists and breeders [29]. Qualitative variations in the phytochemical profiles of Brassicaceae vegetables could contribute to differences in health-promoting properties. Thus, multivariate characterization using metabolic profiling should be used in phenotype visualization and discrimination.

References

1. Salunkhe DK, Kadam SS (1998) Handbook of vegetable science and technology: production, composition. Storage and Processing. Marcel Dekker Inc, New York, pp 533–538
2. Howard LA, Jeffery EH, Wallig MA, Klein BP (1997) Retention of phytochemicals in fresh and processed broccoli. J Food Sci 62:1098–1100
3. Femina A, Robertson JA, Waldron KW (1998) Cauliflower (Brassica oleracea L.), globe artichoke (Cynara scolymus L.) and chicory witloof (Cichorium intybus L.) processing by-products as source of dietary fibre. J Sci Food Agric 77:511–518
4. Kumar S, Andy A (2012) Health promoting bioactive phytochemicals from Brassica. Int Food Res J 19:141–152
5. Park WT, Kim JK, Park S, Lee SW, Li X, Kim YB, Uddin MR, Park NI, Kim SJ, Park SU (2012) Metabolic profiling of glucosinolates, anthocyanins, carotenoids, and other secondary metabolites in kohlrabi (Brassica oleracea var. gongylodes). J Agric Food Chem 60:8111–8116
6. Park SY, Lim SH, Ha SH, Yeo YS, Park WT, Kwon DY, Park SU, Kim JK (2013) Metabolite profiling approach reveals the interface of primary and secondary metabolism in colored cauliflowers (Brassica oleracea L. ssp. botrytis). J Agric Food Chem 61:6999–7007
7. Iniguez-Luy FL, Voort AV, Osborn TC (2008) Development of a set of public SSR markers derived from genomic sequence of a rapid cycling Brassica oleracea L. genotype. Theor Appl Genet 117:977–985
8. Kim TJ, Lee KB, Baek SA, Choi JH, Ha SH, Lim SH, Park SY, Yeo YS, Park SU, Kim JK (2015) Determination of lipophilic metabolites for species discrimination and quality assessment of nine leafy vegetables. J Korean Soc Appl Biol Chem 58:909–918
9. Kim JK, Cho MR, Baek HJ, Ryu TH, Yu CY, Kim MJ, Fukusaki E, Kobayashi A (2007) Analysis of metabolite profile data using batch-learning self-organizing maps. J Plant Biol 50:517–521
10. Kim JK, Lee SY, Chu SM, Lim SH, Suh SC, Lee YT, Cho HS, Ha SH (2010) Variation and correlation analysis of flavonoids and carotenoids in Korean pigmented rice (Oryza sativa L.) cultivars. J Agric Food Chem 58:12804–12809
11. Baek SA, Jung YH, Lim SH, Park SU, Kim JK (2016) Metabolic profiling in Chinese cabbage (Brassica rapa L. subsp. pekinensis) cultivars reveals that glucosinolate content is correlated with carotenoid content. J Agric Food Chem 64:4426–4434
12. Fonville JM, Richards SE, Barton RH, Boulange CL, Ebbels TMD, Nicholson JK, Holmes E, Durnas ME (2010) The

13. Kim YB, Park SY, Park CH, Park WT, Kim SJ, Ha SH, Valan Arasu M, Al-Dhabi NA, Kim JK, Park SU (2016) Metabolomics of differently colored Gladiolus cultivars. Appl Biol Chem 59:597–607
14. Kim JK, Park SY, Lim SH, Yeo Y, Cho HS, Ha SH (2013) Comparative metabolic profiling of pigmented rice (Oryza sativa L.) cultivars reveals primary metabolites are correlated with secondary metabolites. J Cereal Sci 57:14–20
15. Hirai MY, Yano M, Goodenowe DB, Kanaya S, Kimura T, Awazuhara M, Arita M, Fujiwara T, Saito K (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in Arabidopsis thaliana. Proc Natl Acad Sci USA 101:10205–10210
16. Park SY, Choi SR, Lim SH, Yeo YS, 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
17. Messerli G, Partovi Nia V, Trevisan M, Kolbe A, Schauer N, Geigenberger P, Chen J, Davison AC, Fernie AR, Zeeman SC (2007) Rapid classification of phenotypic mutants of Arabidopsis via metabolite fingerprinting. Plant Physiol 143:1484–1492
18. 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
19. Jumtee K, Bamba T, Fukusaki E (2009) Fast GC-FID based metabolic fingerprinting of Japanese green tea leaf for its quality ranking prediction. J Separation Sci 32:2296–2304
20. Eriksson L, Johansson E, Kettaneh-Wold N, Wold S (2001) Multi- and megavariate data analysis principles and applications. Umetrics AB, Umeå
21. Disch A, Hemmerlin A, Bach TJ, Rohmer M (1998) Mevalonate-derived isopentenyl diphosphate is the biosynthetic precursor of ubiquinone prenyl side chain in tobacco BY-2 cells. Biochem J 331:615–621
22. Laule O, Furholz A, Chang HS, Zhu T, Wang X, Heifetz PB, Gruissem W, Lange BM (2003) Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in Arabidopsis thaliana. Proc Natl Acad Sci USA 100:6866–6871
23. Arigoni D, Sagner S, Latzel C, Eisenreich W, Bacher A, Zenk MH (1997) Terpenoid biosynthesis from 1-deoxy-D-xylulose in higher plants by intramolecular skeletal rearrangement. Proc Natl Acad Sci USA 94:10600–10605
24. Kanaya S, Kinouchi M, Abe T, Kudo Y, Yamada Y, Nishi T, Mori H, Ikemura T (2001) Analysis of codon usage diversity of bacterial genes with a self-organizing map (SOM): characterization of horizontally transferred genes with emphasis on the E. coli O157 genome. Gene 276:89–99
25. Kohonen T (1982) Self-organized formation of topologically correct feature maps. Biol Cybern 43:59–69
26. Kohonen T (1990) The self-organizing map. Proc IEEE 78:1464–1480
27. Kohonen T, Oja E, Simula O, Visa A, Kangas J (1996) Engineering applications of the self-organizing map. Proc IEEE 84:1358–1384
28. Young AJ (1991) The photoprotective role of carotenoids in higher plants. Physiol Plant 83:702–708
29. 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