JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

ARTICLE

pubs.acs.org/JAFC

Metabolomic Approach for Age Discrimination of *Panax ginseng* Using UPLC-Q-Tof MS

Nahyun Kim,[†] Kemok Kim,[†] Byeong Yeob Choi,[‡] DongHyuk Lee,[‡] Yoo-Soo Shin,[§] Kyong-Hwan Bang,[§] Seon-Woo Cha,[§] Jae Won Lee,[‡] Hyung-Kyoon Choi,[#] Dae Sik Jang,[⊥] and Dongho Lee[*,†]

[†]School of Life Sciences and Biotechnology and [‡]Department of Statistics, Korea University, Seoul 136-713, Korea

[§]Department of Herbal Crop Research, National Institute of Horticultural & Herbal Science, Rural Development Administration, Eumseong 369-873, Korea

[#]College of Pharmacy, Chung-Ang University, Seoul 156-756, Korea

[⊥]College of Pharmacy, Kyung Hee University, Seoul 130-701, Korea

S *Supporting Information*

ABSTRACT: An ultraperformance liquid chromatography−quadrupole time-of-flight mass spectrometry (UPLC-Q-Tof MS)-based metabolomic technique was applied for metabolite profiling of 60 *Panax ginseng* samples aged from 1 to 6 years. Multivariate statistical methods such as principal component analysis and hierarchical clustering analysis were used to compare the derived patterns among the samples. The data set was subsequently applied to various metabolite selection methods for sophisticated classification with the optimal number of metabolites. The results showed variations in accuracy among the classification methods for the samples of different ages, especially for those aged 4, 5, and 6 years. This proposed analytical method coupled with multivariate analysis is fast, accurate, and reliable for discriminating the cultivation ages of *P. ginseng* samples and is a potential tool to standardize quality control in the *P. ginseng* industry.

KEYWORDS: *Panax ginseng*, age discrimination, UPLC-Q-Tof MS, metabolomics, multivariate analysis

INTRODUCTION

Panax ginseng C.A. Meyer is an important and widely used medicinal herb. It is traditionally used as a panacea because of its replenishing and tonic functions. Furthermore, its various biological and pharmacological activities to increase resistance to physical, chemical, and biological stress and boost general vitality have been reported on the basis of studies of its major active constituents, ginsenosides. More than 40 ginsenosides have been isolated and identified, and they are mainly derived from the *P. ginseng* root, which is considered to be the main part for medicinal purposes.[1−3]

Many studies have demonstrated quality variations in *Panax* according to species, geographical origins, cultivation ages, and various environmental conditions.[3−7] Research on different cultivation ages has shown that the components of this herb vary according to its age, and so do its value and quality. However, ginseng age can hardly be determined by the herb's physical appearance alone. Such age determination led to the consumption of incorrectly identified forms of *P. ginseng*, improper use, and undesirable effects, especially for ginseng of cultivation ages 4, 5, and 6 years, which are the most in demand in the market. Therefore, a reliable method to discriminate the cultivation ages of *P. ginseng* is required for quality control of this medicinal herb and prevention of its adulteration in the market.

Metabolomics, a fine combination of analytical and statistical techniques, provides major insights into the similarities and differences of samples in various environmental conditions by quantitatively and qualitatively measuring the dynamic range of metabolites in organisms. The diverse metabolome data obtained by using this approach enables comparison among samples based on multivariate statistical methods such as principal component analysis (PCA) and hierarchical clustering analysis (HCA). Metabolomics has been applied to research on natural products in various ways, especially for quality control of medicinal plants.[8−10] Metabolomic approaches based on various analytical techniques, including gas chromatography−mass spectrometry (GC-MS), liquid chromatography−mass spectrometry (LC-MS), and nuclear magnetic resonance (NMR), have been applied for metabolite profiling of ginseng extracts.[11−17] In particular, there are several reports of differences among cultivation ages of ginseng revealed by GC-MS and ^{1}H NMR analyses, but the results were limited to certain ages and not applicable to all ages from 1 to 6 years.[18−20]

Here, we show the combined application of ultraperformance liquid chromatography (UPLC) for analysis of nonvolatile compounds of ginseng and quadrupole time-of-flight mass spectrometry (Q-Tof MS) for obtaining multiple levels of structural information to discriminate the cultivation ages of *P. ginseng* samples aged from 1 to 6 years. This combination offers high selectivity, sensitivity, and accuracy for nontargeted analysis of nonvolatile compounds in ginseng. The analyzed metabolites pass through several computational stages such as raw data generation, data treatment, and metabolite selection to derive

Received: May 1, 2011
Revised: September 2, 2011
Accepted: September 2, 2011
Published: September 15, 2011

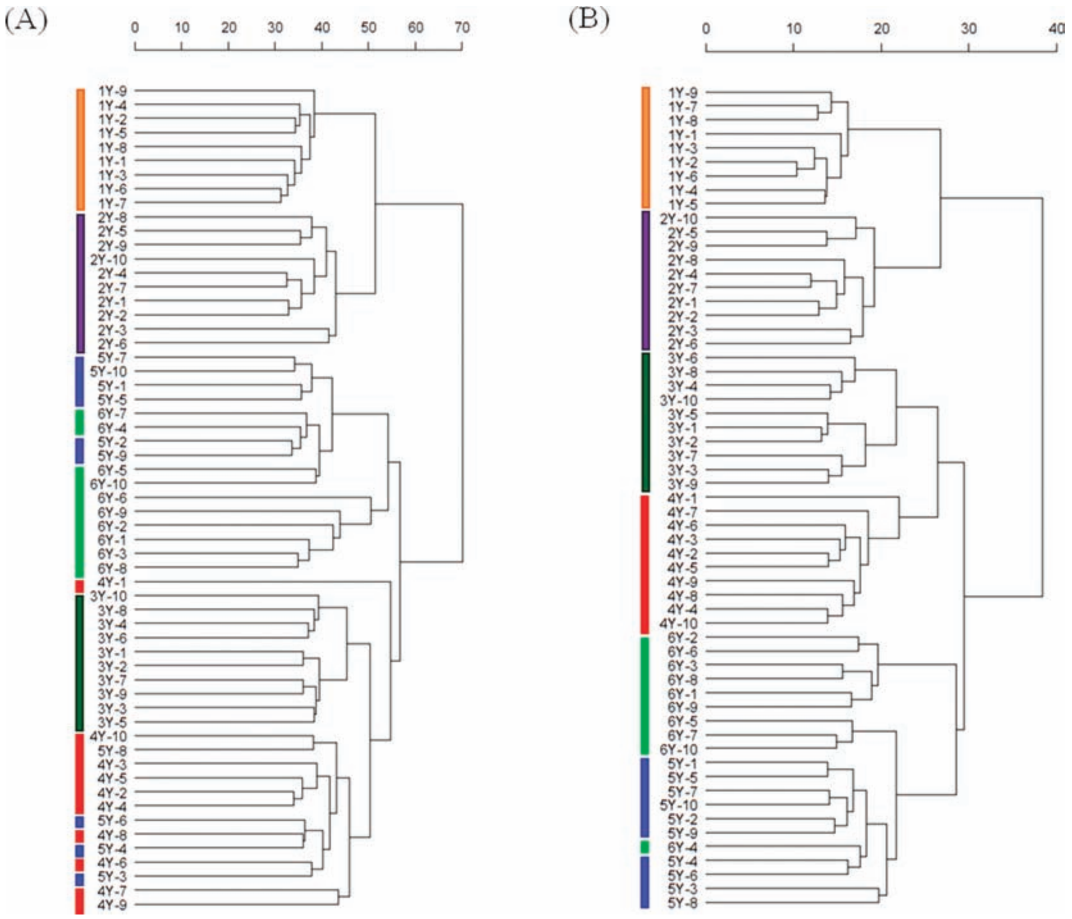

Figure 1. HCA dendrogram of *Panax ginseng* extracts aged from 1 to 6 years with detected metabolites (A) and selected metabolites (B).

the best result for discriminating sample groups. Various metabolite selection methods, including random forest (RF), prediction analysis of microarray (PAM), and partial least squares-discriminant analysis (PLS-DA), have been applied to improve the interpretability of age discrimination data. These statistical processes play an important role in metabolomic investigation for the effective discrimination of samples and identification of marker metabolites representing sample groups.

MATERIALS AND METHODS

Reagents. Acetonitrile and methanol for the preparation and analysis of samples were purchased from Honeywell Birdick & Jackson (Muskegon, MI). Purified water was obtained from an aqua MAX-ultra system (Young Lin, Anyang, Korea). Leucine-enkephalin and formic acid were purchased from Sigma-Aldrich (St. Louis, MO) and Duksan (Seoul, Korea), respectively. The reference solution was 50 pg/μL leucine-enkephalin in 50:50 acetonitrile/water with 0.1% formic acid. All solvents and samples were filtered through 0.2 μm membrane filters before analysis.

Sample Preparation. *P. ginseng* samples for this study were cultivated for 6 years under a regulated system by the Rural Development Administration, Suwon, Republic of Korea. Voucher specimens were deposited at the School of Life Sciences and Biotechnology, Korea University, Seoul, Korea. The main root of *P. ginseng* was cut and freeze-dried (Eyela, Tokyo, Japan). For the UPLC-Q-Tof MS analyses, ginseng samples were extracted with an optimized method for detecting diverse ginseng metabolites.[11] Then, 5 mg of each powdered *P. ginseng* sample was subjected to ultrasonic extraction in 500 μL of 70% MeOH for 20 min. After extraction, the samples were centrifuged at 12000 rpm for 20 min, and the supernatant was filtered through a 0.2 μm membrane filter. The extracts were then dissolved with 50% MeOH to obtain a final concentration of 2 mg/mL. For reliable results, 10 replicates of each sample group aged from 1 to 6 years were prepared.

UPLC-Q-Tof MS Analysis. The samples were analyzed by using an Acquity UPLC system (Waters, Milford, MA) with a Micromass Q-Tof Micro mass spectrometer (Waters, Manchester, U.K.). An Acquity UPLC BEH C_{18} column (2.1 × 100 mm, 1.7 μm) was used to perform the chromatographic separation of 5 μL of each sample injected into a gradient system at a flow rate of 500 μL/min. The mobile phase consisted of 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B). The starting eluent was 10% B. Its proportion was held constant for 0.5 min, increased linearly to 30% from 0.5 to 2.5 min, to 60% from 2.5 to 6 min, and to 90% from 6 to 9 min, held constant at 100% until 10.5 min, returned to the initial composition (10% B) at 10.5 min, and then held constant for 4.5 min to re-equilibrate the column. This UPLC elution condition was optimized to detect the maximal number of metabolites in *P. ginseng*, especially to separate ginsenosides for identifying markers. The column and sample managers were maintained at 35 and 15 °C, respectively. The mass spectrometer was operated in negative ion mode and set to the total ion chromatogram (TIC) mode. The optimized MS conditions were as follows: capillary voltage of 2800 V, cone voltage of 35 V, source temperature of 100 °C, desolvation temperature of 250 °C, and desolvation gas flow rate of 600 L/h. To ensure that mass was measured accurately, leucine-enkephalin was used as the reference lock-mass compound at a concentration of

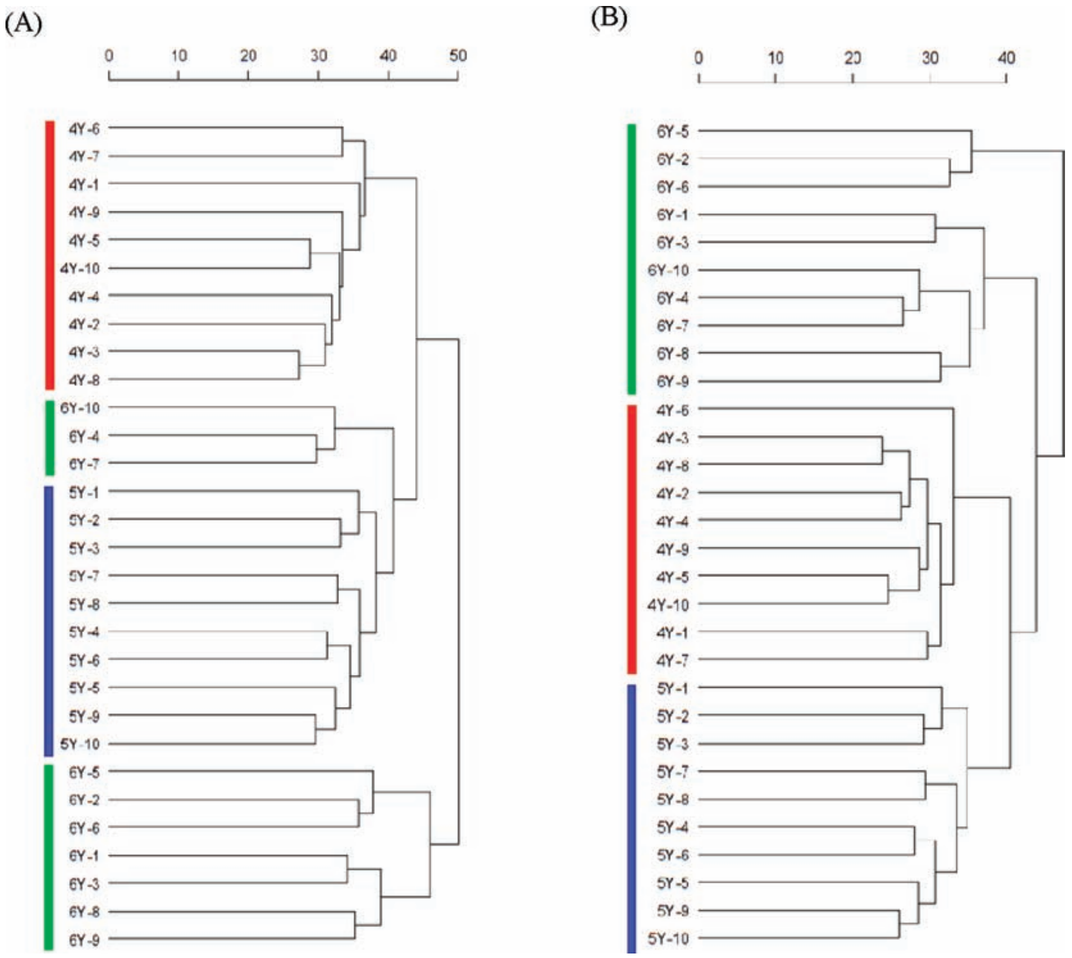

Figure 2. HCA dendrogram of *Panax ginseng* extracts aged from 4 to 6 years with detected metabolites (A) and selected metabolites (B).

500 pg/μL at a flow rate of 2 μL/min, and the $[M - H]^-$ ion at 554.2615 Da was detected over 15 min of analysis.

Metabolite Profiling. The developed UPLC-Q-Tof MS method was used to analyze 60 samples of *P. ginseng* main root extracts. For metabolite profiling, 5 μL of each sample was introduced in the UPLC system, and 20% acetonitrile was used as a blank sample for every five samples to prevent interference among the samples.

Raw Data Generation. The metabolites acquired by UPLC-Q-Tof MS-based metabolomic profiling were analyzed with MassLynx version 4.1 (Waters, Manchester, U.K.). Chromatographic data were preprocessed and normalized so that all samples could be compared in the same condition. Retention times (RT) of 1−10.5 min, mass range from 200 to 1500 Da, mass tolerance of 0.05 ppm, noise elimination level of 6.00, minimum intensity of 15%, and RT tolerance of 0.01 min were set to align the peak RT and calculate the peak intensity of each sample. The intensities of all detected peaks in a single sample were calculated on the basis of both RT and m/z data of each peak, and the ion intensities of each peak were normalized against the sum of the peak intensities in each sample by using the MarkerLynx XS Application Manager (Waters, Milford, MA). Every sample was applied to this raw data-generation process to enable data treatment, metabolite selection, and multivariate analysis.

Data Treatment. The generated raw data were applied to a data treatment process for dealing with missing values and conversion to a proper data set for classification. First, metabolites having >75% of the number of zero values across all samples were eliminated by assuming that they did not possess specific patterns and influence to characterize the samples. Second, to reduce heteroscedasticity, zero values in the data set except the eliminated metabolites were replaced with the smallest value larger than zero to fit the further data treatment. Log transformation with two bases was used to even the importance of each metabolite regardless of the amount in the samples so that all metabolites were equally evaluated over the data set for the data analysis. Third, the values were replaced with *k*-nearest neighbor imputation providing accurate estimation with a substantial and delicate approach to interpolate missing data.[21] Finally, the data set was normalized by l_2 norm for unifying the influence of each sample and scaled by unit variance for unifying the influence of each metabolite so that correlated samples could be connected to each other.[22] A more detailed description of each data treatment process is in the Supporting Information.

Metabolite Selection. A growing concern to obtain a desirable result is the identification of metabolites relevant to discrimination. Total metabolites processed according to specific data treatment procedures are used as potential targets for discrimination. However, the classification of samples with a massive volume of data is quite challenging because numerous metabolites are detected by metabolome analysis, especially UPLC-Q-Tof MS-based analysis. To decrease the sample size and improve data interpretability by selecting influential metabolites within the processed metabolite list for discrimination of samples, the following classification methods were used: RF, PAM, and PLS-DA. Each method has its unique technique to select optimal numbers of metabolites having high significance to discriminate. RF is a classification method based on decision tree learning as an algorithm developed by Breiman.[23] It uses random selection of metabolites in a

Table 1. Cross-Validation Accuracy of Each Age of *Panax ginseng* Ranging from 1 to 6 Years by Different Classification Methods

		CV accuracy (n = 59)						
classification method	no. of selected metabolites	1 year	2 years	3 years	4 years	5 years	6 years	mean
RF	119	1.000	1.000	1.000	0.998	0.900	0.948	0.974
PAM	1146	1.000	0.900	1.000	0.900	0.796	0.966	0.926
PLS-DA	198	1.000	1.000	1.000	1.000	0.996	1.000	0.999

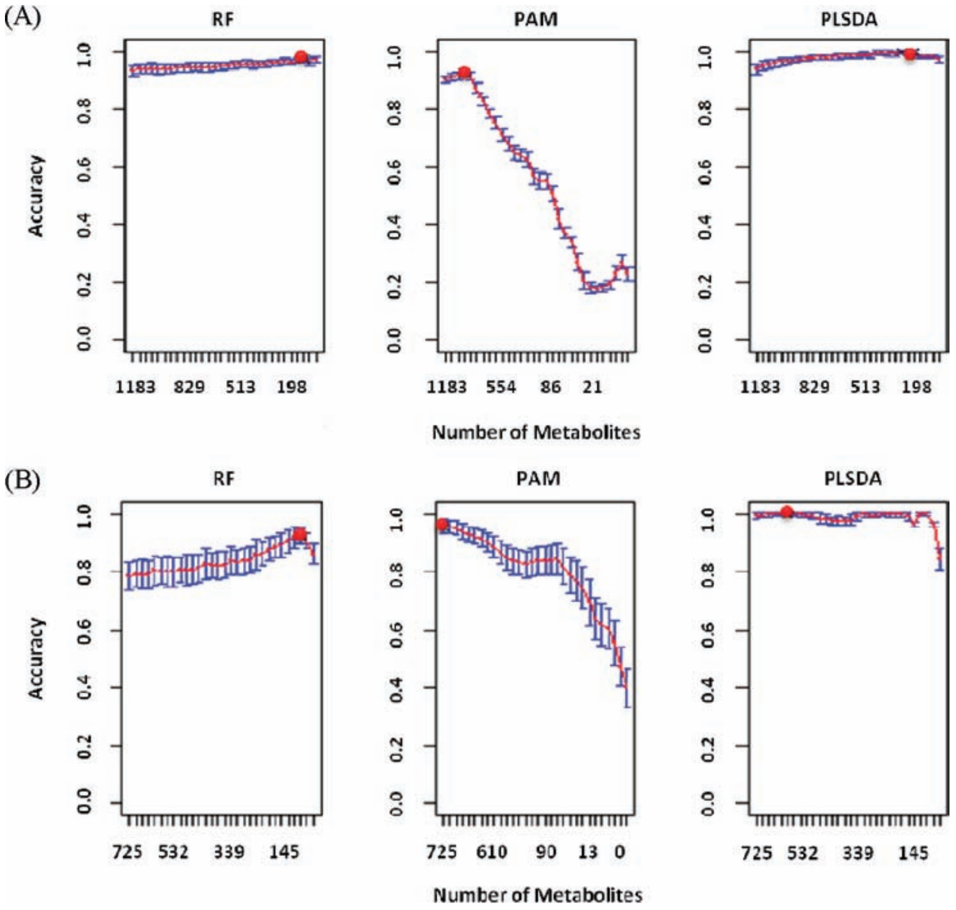

Figure 3. Mean accuracy of three different classification methods (RF, PAM, and PLS-DA) to select the optimal number of metabolites to discriminate ages of *Panax ginseng* ranging from 1 to 6 years (A) and from 4 to 6 years (B).

data set and forms classification trees, a forest, and the forest predicts the class of a new sample by deciding how closely related the sample and the tree are across all trees. PAM is a centroid classification method using the nearest shrunken centroid methodology proposed by Narashiman.[24] It calculates a standardized centroid for each class and chooses the class of a new sample by measuring its distance to the class centroid. PLS-DA is a dimension reduction method based on both group and metabolite information.[25] It finds new components in an independent metabolite space attempting to maximize the separation between groups of samples. In addition, to enhance the classification accuracy and prevent data overfitting, data validation and selection of the optimal number of metabolites for each classification method are required.[26] In this experiment, we used backward metabolite elimination to choose the optimal number of metabolites and 10-fold cross-validation (CV) to maximize the classification accuracy. In the CV procedure, the samples were randomly divided into 10 equal parts, and a model was built by using 9 of the 10 CV groups as training data. The remaining part was used as test data by fitting to the resulting model to calculate the error rate and predict the accuracy of each class.[27] The details of each classification method and backward selection are further described in the Supporting Information.

Multivariate Analysis. All processed data of the *P. ginseng* samples were analyzed by multivariate statistical techniques such as PCA and HCA to evaluate the similarities and differences of the tested samples with the MarkerLynx XS Application Manager and R+ package (R Foundation for Statistical Computing, Vienna, Austria), respectively. PCA is an unsupervised method for pattern analysis and the predominantly used multivariate statistical method to visualize all acquired data in two- or three-dimensional score plots by reducing all detected metabolites to several new principal components.[10] HCA is a clustering method to compare patterns of similarities and dissimilarities by measuring distances among samples, and a dendrogram represents the relationships among samples.[10]

RESULTS AND DISCUSSION

Metabolite Profiling and Data Processing. The chromatographic results of UPLC-Q-Tof MS for analyzing the 60 *P. ginseng* samples have barely shown the visible difference of ages (Supporting Information, Figure S1). Consequently, all detected metabolites were processed for comprehensive evaluation of the samples. The chromatographic data were transferred to a 3-dimensional data set composed of RT, *m*/*z*, and ion intensity to align the data, which is one of the most important tasks for multivariate analysis to obtain precise and stable results from all analyzed samples. The aligned data were then processed through a series of data treatment procedures to refine metabolites used for classification of *P. ginseng* age.

Multivariate Analysis of Detected Metabolites. The data treatment process identified 1361 metabolites from the 60 *P. ginseng* samples. By PCA, one of the 1-year-old samples was excluded as an outlier, which is identifiable by its location outside the 2-dimensional tolerance ellipse plot.[21] Therefore, 59 samples, which are located inside the ellipse, were further analyzed. As HCA explains the hierarchy of clusters illustrating the relationships among samples, the samples aged 1 and 2 years, which were the closest, merged into clusters (Figure 1A). Similarly, the samples aged 3 and 4 years as well as those aged 5 and 6 years merged into clusters. Moreover, the samples aged 3 and 4 years progressively merged with those aged 5 and 6 years to form a bigger cluster that finally merged with the cluster of samples aged 1 and 2 years to create a whole tree structure. However, the HCA dendrogram showed that the 1−3-year-old ginseng samples clustered clearly, but several samples aged 4, 5, and 6 years were mixed with one another. This confirms that ginseng samples aged from 1 to 3 years can be clearly discriminated according to their detected metabolites, whereas samples aged from 4 to 6 years cannot be discriminated in this manner.

To clearly discriminate the cultivation ages that are most in demand, data from 30 samples, aged from 4 to 6 years, were separated from the data of all samples and the relationships were compared. By additional analysis, 1155 metabolites were detected from these ginseng samples, and multivariate analysis was performed by using these metabolites. The HCA dendrogram showed that the 4- and 5-year-old samples merged as clusters and the 6-year-old samples subsequently merged together (Figure 2A). It is also observed that three samples aged 6 years were mixed with those aged 5 years, which is similar to the HCA result in Figure 1A. The 2-dimensional PCA score plots showing clustering and scattering among the samples are shown in detail in the Supporting Information (Figures S2 and S3). On the basis of the 2-dimensional PCA score plots, we further processed the 3-dimensional PCA score plots that allow us to more easily see the distinction of clustering patterns among the samples [Supporting Information, Figures S4 (A) and S5 (A)].

Metabolite Selection for Classification. To identify the relevant metabolites for age discrimination of *P. ginseng*, we applied classification methods reducing the large amount of data to the minimal size, thereby increasing the accuracy for age discrimination. For example, in the case of the 59 ginseng samples, although 1361 metabolites were detected, not all were influential in discriminating sample ages. Only some affect the classification of ginseng ages; others can even interrupt the classification and should be eliminated to increase the accuracy.

Table 2. Confusion Matrices between True Class and Predicted Class of Each Age of *Panax ginseng* Ranging from 1 to 6 Years by Different Classification Methods [RF (A), PAM (B), and PLS-DA (C)]

	predicted class						
true class	1 year	2 years	3 years	4 years	5 years	6 years	prediction accuracy
			(A) RF				
1 year	450	0	0	0	0	0	1.000
2 years	0	500	0	0	0	0	1.000
3 years	0	0	500	0	0	0	1.000
4 years	0	0	0	499	1	0	0.998
5 years	0	0	0	9	450	41	0.900
6 years	0	0	0	0	26	474	0.948
			(B) PAM				
1 year	450	0	0	0	0	0	1.000
2 years	50	450	0	0	0	0	0.900
3 years	0	0	500	0	0	0	1.000
4 years	0	0	0	450	50	0	0.900
5 years	0	0	0	0	398	102	0.796
6 years	0	0	0	0	17	483	0.966
			(C) PLS-DA				
1 year	450	0	0	0	0	0	1.000
2 years	0	500	0	0	0	0	1.000
3 years	0	0	500	0	0	0	1.000
4 years	0	0	0	500	0	0	1.000
5 years	0	0	2	0	498	0	0.996
6 years	0	0	0	0	0	500	1.000

Table 3. Cross-Validation Accuracy of Each Age of *Panax ginseng* Ranging from 4 to 6 Years by Different Classification Methods

		CV accuracy (n = 30)			
classification method	no. of selected metabolites	4 years	5 years	6 years	mean
RF	73	0.995	0.804	0.974	0.924
PAM	725	1.000	0.999	0.879	0.959
PLS-DA	605	1.000	1.000	1.000	1.000

The results of RF, PAM, and PLS-DA were compared; for the 59 *P. ginseng* samples, these methods chose 119, 1146, and 198 metabolites from the 1361 metabolites, respectively. The CV accuracy of each classification method for each age ranged from 0.926 to 0.999 (Table 1). Especially, PLS-DA had 0.999 accuracy, indicating that this classification method can discriminate the ages of *P. ginseng* with 99.9% accuracy. Although RF and PAM were less accurate, their accuracies are sufficiently high for age discrimination. The mean accuracy of each classification method shown in Figure 3A explains how we selected the optimal number of metabolites, marked with a red dot for the highest point of accuracy. Moreover, the confusion matrices shown in Table 2 indicate the prediction accuracy of each classification method based on 10-fold cross-validation. The numbers in rows indicate multiples of the number of samples and 50 times iteration. As an example, the numbers in the rows and columns of Table 2C

Table 4. Confusion Matrices between True Class and Predicted Class of Each Age of *Panax ginseng* Ranging from 4 to 6 Years by Different Classification Methods [RF (A), PAM (B), and PLS-DA (C)]

	predicted class			
true class	4 years	5 years	6 years	prediction accuracy
		(A) RF		
4 years	995	5	0	0.995
5 years	125	804	71	0.804
6 years	3	23	974	0.974
		(B) PAM		
4 years	1000	0	0	1.000
5 years	0	999	1	0.999
6 years	0	121	879	0.879
		(C) PLS-DA		
4 years	1000	0	0	1.000
5 years	0	1000	0	1.000
6 years	0	0	1000	1.000

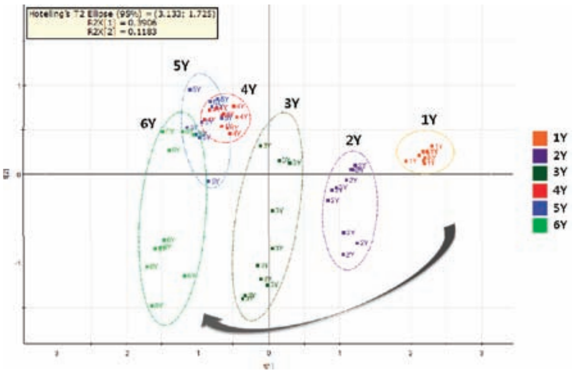

Figure 4. PCA 2D score plot of *Panax ginseng* extracts aged from 1 to 6 years with selected metabolites: 1Y, 2Y, 3Y, 4Y, 5Y, and 6Y represent 1-, 2-, 3-, 4-, 5-, and 6-year-old ginseng, respectively.

explain how accurately the test samples were predicted by PLS-DA and the resultant prediction accuracy. The samples of all the ages except 5 years were exactly matched with the true class having 100% accuracy. Finally, 301 metabolites satisfying at least two or all the classification methods were used for further statistical analysis by improving the data interpretability.

For the 4−6-year-old ginseng samples, RF, PAM, and PLS-DA selected 73, 725, and 605 metabolites from 1155 metabolites, respectively. The CV accuracy ranged from 0.924 to 1.000, and PLS-DA had 1.000 accuracy for all ages (Table 3). This means that 100% classification of 4-, 5-, and 6-year-old ginseng samples was possible by PLS-DA in this data set. The mean accuracy of each classification method is shown in Figure 3B, and the selected numbers of metabolites with the highest accuracy are marked with a red dot. Confusion matrices were indicated with the same method as already described (Table 4), and, finally, 606 selected metabolites were used for further multivariate analysis. The margin accuracy showing variations in accuracy among the ages of each method is indicated in the Supporting Information [Figures S6 (A) and (B)].

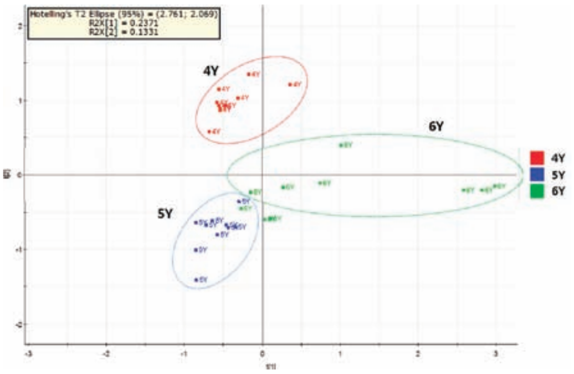

Figure 5. PCA 2D score plot of *Panax ginseng* extracts aged from 4 to 6 years with selected metabolites: 4Y, 5Y, and 6Y represent 4-, 5-, and 6-year-old ginseng, respectively.

Multivariate Analysis of Selected Metabolites. To confirm the importance of metabolite selection for discrimination, the 301 and 606 metabolites selected from 59 and 30 samples, respectively, by RF, PAM, and PLS-DA were used for PCA and HCA. For the 1−6-year-old ginseng samples, the 2-dimensional PCA plots with the 301 selected metabolites showed better clustering than those with the total metabolites (Figure 4). PC 1 and PC 2 demonstrated 50.89% of the total variance, indicating that the data interpretability with the selected metabolites improved through the metabolite selection process. The HCA dendrogram indicated more organized hierarchy by merging clusters according to the ages and shortened distance among the samples than the one with the total metabolites (Figure 1B). For the 4−6-year-old ginseng samples, the 2-dimensional PCA score plot with the 606 selected metabolites indicated higher variance of 37.02% than the PCA result with the total metabolites (Figure 5), and the HCA dendrogram in Figure 2B described successful discrimination among the ages of 4, 5, and 6 years. These HCA results with the selected metabolites in comparison with the total metabolites confirmed the importance of the metabolite selection process for more precise classification (Figures 1 and 2). More obvious results can be observed by the 3-dimensional PCA score plots in Figures S4 (B) and S5 (B) of the Supporting Information.

In conclusion, this study presents a novel finding that different cultivation ages of *P. ginseng* can be successfully discriminated by using the proposed UPLC-Q-Tof MS-based metabolomic approach together with precise statistical analysis. By this approach, ginseng samples aged 4, 5, and 6 years, which are the most in demand in the ginseng market, can be precisely classified on the basis of selected metabolites. Further investigations on various ginseng samples that show environmental variations should be performed to build a more extensive and robust model. Moreover, this method, along with the identification of potential biomarkers for each age, could be used as an effective tool for quality control in the *P. ginseng* industry.

■ ASSOCIATED CONTENT

S Supporting Information. The statistical methods for data treatment and metabolite selection used in this study are described in detail. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Phone: +82 2 3290 3017. Fax: +82 2 953 0737. E-mail: dongholee@korea.ac.kr.

Funding Sources

This work was supported by BioGreen21 (No. 20070501034007) and Agenda (No. 20100501030008), Rural Development Administration, Republic of Korea.

REFERENCES

(1) Ruan, C.-C.; Liu, Z.; Li, X.; Liu, X.; Wang, L.-J.; Pan, H.-Y.; Zheng, Y.-N.; Sun, G.-Z.; Zhang, Y.-S.; Zhang, L.-X. Isolation and characterization of a new ginsenoside from the fresh root of *Panax ginseng*. *Molecules* **2010**, *15*, 2319–2325.

(2) Lu, J. M.; Yao, Q.; Chen, C. Ginseng compounds: an update on their molecular mechanisms and medical applications. *Curr. Vasc. Pharmacol.* **2009**, *7*, 293–302.

(3) Shi, W.; Wang, Y.; Li, J.; Zhang, H.; Ding, L. Investigation of ginsenosides in different parts and ages of *Panax ginseng*. *Food Chem.* **2007**, *102*, 664–668.

(4) Angelova, N.; Kong, H. W.; van der Heijden, R.; Yang, S. Y.; Choi, Y. H.; Kim, H. K.; Wang, M.; Hankemeier, T.; van der Greef, J.; Xu, G.; Verpoorte, R. Recent methodology in the phytochemical analysis of ginseng. *Phytochem. Anal.* **2008**, *19*, 2–16.

(5) Assinewe, V. A.; Baum, B. R.; Gagnon, D.; Arnason, J. T. Phytochemistry of wild populations of *Panax quinquefolius* L. (North American ginseng). *J. Agric. Food Chem.* **2003**, *51*, 4549–4553.

(6) Lim, W.; Mudge, K. W.; Vermeylen, F. Effects of population, age, and cultivation methods on ginsenoside content of wild American ginseng (*Panax quinquefolium*). *J. Agric. Food Chem.* **2005**, *53*, 8498–8505.

(7) Wang, C. Z.; McEntee, E.; Wicks, S.; Wu, J. A.; Yuan, C. S. Phytochemical and analytical studies of *Panax notoginseng* (Burk.) FH Chen. *J. Nat. Med.* **2006**, *60*, 97–106.

(8) Fukusaki, E.; Kobayashi, A. Plant metabolomics: potential for practical operation. *J. Biosci. Bioeng.* **2005**, *100*, 347–354.

(9) Okada, T.; Afendi, F. M.; Altaf-Ul-Amin, M.; Takahashi, H.; Nakamura, K.; Kanaya, S. Metabolomics of medicinal plants: the importance of multivariate analysis of analytical chemistry data. *Curr. Comput.-Aided Drug Des.* **2010**, *6*, 179–196.

(10) Sumner, L. W.; Mendes, P.; Dixon, R. A. Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry* **2003**, *62*, 817–836.

(11) Chan, E. C.; Yap, S. L.; Lau, A. J.; Leow, P. C.; Toh, D. F.; Koh, H. L. Ultra-performance liquid chromatography/time-of-flight mass spectrometry based metabolomics of raw and steamed *Panax notoginseng*. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 519–528.

(12) Dan, M.; Su, M.; Gao, X.; Zhao, T.; Zhao, A.; Xie, G.; Qiu, Y.; Zhou, M.; Liu, Z.; Jia, W. Metabolite profiling of *Panax notoginseng* using UPLC-ESI-MS. *Phytochemistry* **2008**, *69*, 2237–2244.

(13) Kang, J.; Lee, S.; Kang, S.; Kwon, H. N.; Park, J. H.; Kwon, S. W.; Park, S. NMR-based metabolomics approach for the differentiation of ginseng (*Panax ginseng*) roots from different origins. *Arch. Pharm. Res.* **2008**, *31*, 330–336.

(14) Lee, E. J.; Shaykhutdinov, R.; Weljie, A. M.; Vogel, H. J.; Facchini, P. J.; Park, S. U.; Kim, Y. K.; Yang, T. J. Quality assessment of ginseng by ^{1}H NMR metabolite fingerprinting and profiling analysis. *J. Agric. Food Chem.* **2009**, *57*, 7513–7522.

(15) Toh, D. F.; New, L. S.; Koh, H. L.; Chan, E. C. Ultra-high performance liquid chromatography/time-of-flight mass spectrometry (UHPLC/TOFMS) for time-dependent profiling of raw and steamed *Panax notoginseng*. *J. Pharm. Biomed. Anal.* **2010**, *52*, 43–50.

(16) Wang, Y.; Pan, J. Y.; Xiao, X. Y.; Lin, R. C.; Cheng, Y. Y. Simultaneous determination of ginsenosides in *Panax ginseng* with different growth ages using high-performance liquid chromatography-mass spectrometry. *Phytochem. Anal.* **2006**, *17*, 424–430.

(17) Xie, G.; Plumb, R.; Su, M.; Xu, Z.; Zhao, A.; Qiu, M.; Long, X.; Liu, Z.; Jia, W. Ultra performance LC/TOF MS analysis of medicinal *Panax* herbs for metabolomic research. *J. Sep. Sci.* **2008**, *31*, 1015–1026.

(18) Lin, W. N.; Lu, H. Y.; Lee, M. S.; Yang, S. Y.; Chen, H. J.; Chang, Y. S.; Chang, W. T. Evaluation of the cultivation age of dried ginseng radix and its commercial products by using ^{1}H-NMR fingerprint analysis. *Am. J. Chin. Med.* **2010**, *38*, 205–218.

(19) Qiu, Y.; Lu, X.; Pang, T.; Ma, C.; Li, X.; Xu, G. Determination of radix ginseng volatile oils at different ages by comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry. *J. Sep. Sci.* **2008**, *31*, 3451–3457.

(20) Shin, Y. S.; Bang, K. H.; In, D. S.; Kim, O. T.; Hyun, D. Y.; Ahn, I. O.; Ku, B. C.; Kim, S. W.; Seong, N. S.; Cha, S. W.; Lee, D.; Choi, H. K. Fingerprinting analysis of fresh ginseng roots of different ages using ^{1}H-NMR spectroscopy and principal components analysis. *Arch. Pharm. Res.* **2007**, *30*, 1625–1628.

(21) Enot, D. P.; Lin, W.; Beckmann, M.; Parker, D.; Overy, D. P.; Draper, J. Preprocessing, classification modeling and feature selection using flow injection electrospray mass spectrometry metabolite fingerprint data. *Nat. Protoc.* **2008**, *3*, 446–470.

(22) Scholz, M.; Gatzek, S.; Sterling, A.; Fiehn, O.; Selbig, J. Metabolite fingerprinting: detecting biological features by independent component analysis. *Bioinformatics* **2004**, *20*, 2447–2454.

(23) Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.

(24) Tibshirani, R.; Hastie, T.; Narasimhan, B.; Chu, G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6567–6572.

(25) Barker, M.; Rayens, W. Partial least squares for discrimination. *J. Chemom.* **2003**, *17*, 166–173.

(26) Radmacher, M. D.; McShane, L. M.; Simon, R. A paradigm for class prediction using gene expression profiles. *J. Comput. Biol.* **2002**, *9*, 505–511.

(27) Qiu, Y.; Rajagopalan, D.; Connor, S. C.; Damian, D.; Zhu, L.; Handzel, A.; Hu, G.; Amanullah, A.; Bao, S.; Woody, N. Multivariate classification analysis of metabolomic data for candidate biomarker discovery in type 2 diabetes mellitus. *Metabolomics* **2008**, *4*, 337–346.