

An approach for simultaneous determination for geographical origins of Korean *Panax ginseng* by UPLC-QTOF/MS coupled with OPLS-DA models

Hyuk-Hwan Song, Doo-Young Kim, Soyeun Woo, Hyeong-Kyu Lee, and Sei-Ryang Oh*

Natural Medicine Research Center, Korea Research Institute of Bioscience and Biotechnology, Cheongwon 363-883, Korea

Identification of the origins of *Panax ginseng* has been issued in Korea scientifically and economically. We describe a metabolomics approach used for discrimination and prediction of ginseng roots from different origins in Korea. The fresh ginseng roots from six ginseng cooperative associations (Gangwon, Gaeseong, Punggi, Chungbuk, Jeonbuk, and Anseong) were analyzed by UPLC-MS-based approach combined with orthogonal projections to latent structure-discriminant analysis multivariate analysis. The ginsengs from Gangwon and Gaeseong were easily differentiated. We further analyzed the metabolomics results in subgroups. Punggi, Chungbuk, Jeonbuk, and Anseong ginseng could be easily differentiated by the first two orthogonal components. As a validation of the discrimination model, we performed blind prediction tests of sample origins using an external test set. Our model predicted their geographical origins as 99.7% probability. The robust discriminatory power and statistical validity of our method suggest its general applicability for determining the origins of *P. ginseng* samples.

Keywords: *Panax ginseng*, Metabolomics, Orthogonal projections to latent structure-discriminant analysis, Discrimination of origins

INTRODUCTION

Ginseng is a plant in the genus *Panax* found in the Northern Hemisphere and in East Asian countries. It is estimated that about 65% to 80% of the world's population is using traditional medicine as the primary form of healthcare [1]. Ginseng has been reported to contain polyacetylenes, sesquiterpenes, polysaccharides, peptidoglycans, and vitamins, besides more than 30 ginsenosides have primarily been associated of the major therapeutic effects of ginseng roots [2,3]. It has recently been reported that differences in the bioactivities among species, geographical origin, and extraction methods have been linked to different ratio of ginsenosides [4,5]. As compositions of ginseng roots could vary significantly according to the origin and age, it is important to discriminate

among ginseng roots from different sources [6-9].

Metabolomics has been applied to the classification of plant materials, with principal component analysis (PCA) as the main statistical approach [10-12]. The orthogonal projection to latent structure-discriminant analysis (OPLS-DA) was recently developed, a type of supervised classification [13]. The predictive and orthogonal components supported by OPLS-DA facilitate the interpretation of class discrimination and the prediction of the class membership of unknown samples. Therefore, OPLS-DA is more appropriated than PCA to differentiate origins in cases where many factors can affect metabolite profiles [14,15]. Metabolite profiling has been applied using a number of techniques including NMR or differ-

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Edited by Jae Joon Wee, Korea Ginseng Corp., Korea

Received 18 Jan. 2013, Revised 26 Mar. 2013, Accepted 16 Apr. 2013

* Corresponding author

E-mail: seiryang@kribb.re.kr

Tel: +82-43-240-6110, Fax: +82-43-240-6009

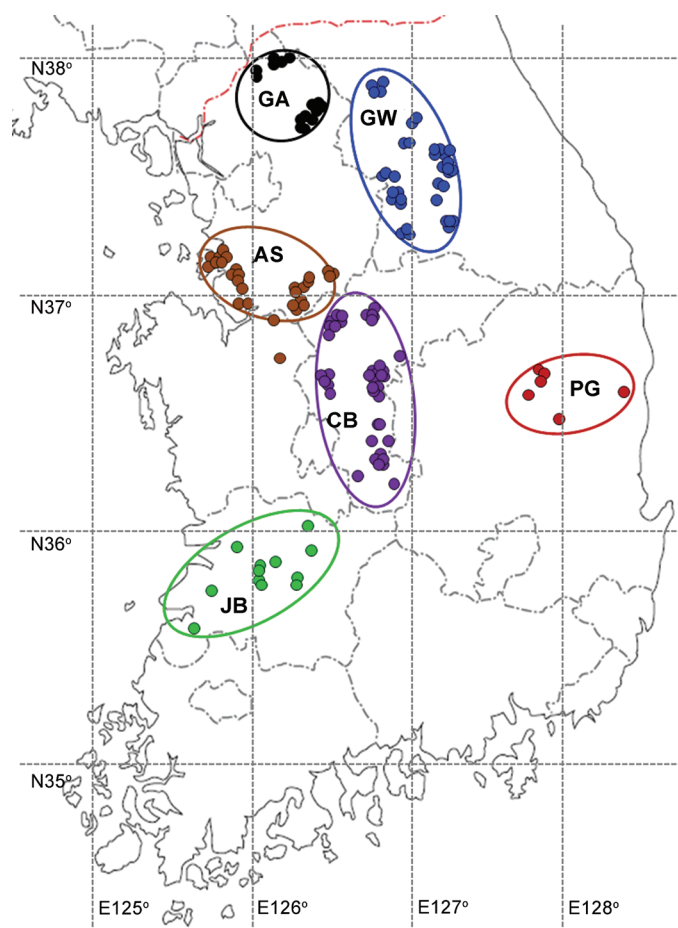


Fig. 1. The origins of ginseng samples. GA, Gaeseong; GW, Gangwon; AS, Anseong; CB, Chungbuk; PG, Punggi; JB, Jeonbuk.

ent combination of LC, GC, and MS [16-21]. Moreover, the quality and profiling of natural products have been reported to vary widely depending on the geographical origin, growing environment, storage condition, and post-harvest processing of the herbal ingredients [12,22,23].

In the present study, a metabolomics approach combining UPLC-MS-based analysis with OPLS-DA was developed for discriminating the six geographical origins (Gangwon, Gaeseong, Punggi, Chungbuk, Jeonbuk, and Anseong) of Korean *P. ginseng* roots. The OPLS-DA model was statistically allowed for differences among their origins. Importantly, validation by predicted model on blind test samples gave a statistical measurement of the reliability of the approach, which is required for practical application.

MATERIALS AND METHODS

Chemicals and materials

Leucine-enkephalin and formic acid were purchased

from Sigma-Aldrich (St. Louis, MO, USA). Acetonitrile and methanol of HPLC grade were obtained from SK Chemical Reagent (Seoul, Korea). All aqueous solutions were prepared with ultrapure water produced by Milli-Q system (18.2 MΩ; Millipore, Bedford, MA, USA). Four hundred twenty-nine Korean *P. ginseng* roots (4 to 6 years old) were collected locally from Gangwon (GW, 90 samples), Gaeseong (GA, 78 samples), Punggi (PG, 18 samples), Chungbuk (CB, 108 samples), Jeonbuk (JB, 36 samples) and Anseong (AS, 99 samples) Insam Cooperative Associations of National Agricultural Cooperative Federation (Nonghyup) during October to November in 2009 (Fig. 1). The roots of each were washed with distilled water to remove adhered soil particles. Each ginseng root body including hair was freeze-dried and ground. The obtained dried ginseng (0.5 g) was weighed, and 5 mL of 70% methanol were added. The methanol extract of ginseng was obtained in an ultrasonic water bath for 60 min at room temperature [24]. The fluid was filtered through a syringe filter (0.22 μm) and injected

directly into the UPLC system.

UPLC-QTOF/MS analysis

P. ginseng metabolite profiling was performed us-

ing an ACQUITY UPLC system (Waters Corporation, Milford, MA, USA) which was equipped with a binary solvent delivery manager, and a sample manger coupled to Micromass Q-TOF Premier mass spectrometer (Wa-

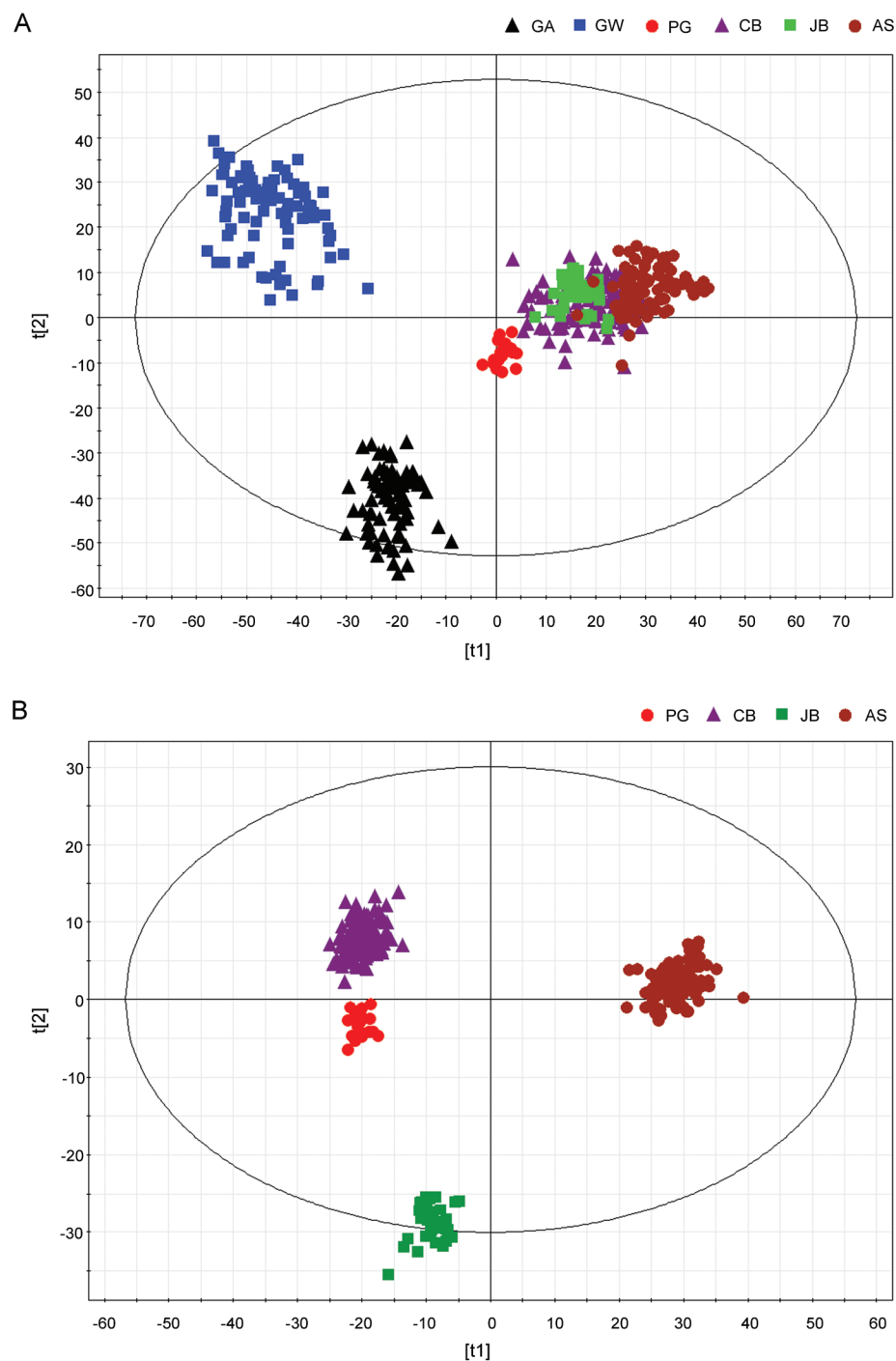


Fig. 2. Multivariate statistical analysis of ginseng samples. (A) Orthogonal projection to latent structure-discriminant analysis (OPLS-DA) score plot of 6 geographical origins (Gaeseong [GA], Gangwon [GW], Anseong [AS], Chungbuk [CB], Punggi [PG], and Jeonbuk [JB]). (B) OPLS-DA score plots of 4 geographical origins (AS, PG, JB, and CB).

ters Corporation) equipped with an electrospray interface. Chromatographic separations were performed on a 2.1×100 mm, 1.7 μm ACQUITY BEH C18 chromatography column. The column temperature was maintained at 35°C, and the mobile phases A and B were water with 0.1% formic acid and acetonitrile with 0.1% formic acid, respectively. The gradient duration program was: 0 min, 10% B; 0 to 7 min, 10% to 33% B; 7 to 14 min, 33% to 56% B; 14 to 21 min, 56% to 100% B; 21 to 23 min, 100% B; 23 to 25 min 10% B. The flow rate was 0.4 mL/min and the injection volume 5 μL with partial loop mode.

The mass spectrometer was operated in a positive ion mode. N₂ was used as the desolvation gas. The desolvation temperature was set to 350°C at a flow rate of 500 L/h and source temperature of 100°C. The capillary and cone voltages were set to 2,700 and 27 V, respectively. The data were collected for each sample from 200 to 1,500 Da with a 0.25-second scan time and a 0.01-second interscan delay over a 25-minute analysis time. Leucine-enkephalin (*m/z* 556.2771) was used as a reference compound.

Chemometric data analysis

The mass raw data were analyzed by the MarkerLynx applications manager ver. 4.1 (Waters, Manchester, UK). The parameters were as follows: retention time range 0.5 to 18.5 min, mass range 200 to 1,500 Da, mass tolerance 0.04 Da, isotopic data were excluded for analysis, noise elimination level was set at 10, mass window and retention time window were set at 0.04 and 0.1 min, respectively. After creating a process for mean-centered and par-scaled data set the Create Dataset window, then, PCA and OPLS-DA were performed to discriminate and predict the geographical origins of ginseng roots. The

resulting two-dimensional matrix data of mass values and their peak intensities were further investigated with SIMCA-P⁺ software 12.0 (Umetrics, Umea, Sweden) for multivariate statistical analysis. In the OPLS method, the systematic variations in *X* were separated into two parts; one that is linear to *Y* and the other is orthogonal to *Y* [15,25], hence, the OPLS model comprises two blocks of model variations: 1) the *Y*-predictive block, which represents the between class variation, and 2) the *Y*-orthogonal block also referred to as the uncorrelated variation, which constitutes the within class variation [26].

RESULTS AND DISCUSSION

In a total ion chromatography (TIC), over thousands ion peaks from each extract of *P. ginseng* were detected in the range of *m/z* 200 to 1,500 at 0.5 to 18.5 min. Because of complexity and similarity of TICs, multivariate analysis such as PCA or OPLS-DA was performed to discriminate the origin of each sample. In the PCA analysis, ginseng samples of 6 geographical origins were not clearly discriminated because the major peaks were severely overlapped (data were not shown). However, in the score plot of OPLS-DA model, three clusters were clearly separated as GW, GA, and the other 4 regions (Fig. 2A) the cross-validated predictive ability (*Q*²) and the variance related to the differences among the classes *R*²(*y*) were found to be significant (*Q*²=0.794, *R*²[*y*]=0.875) with five predictive and five orthogonal (5+5) components. The separation of GW and GA clusters from a cluster of 4 other regions could be easily achieved by combining predictive component 1 with predictive component 2. After the data sets of GW and GA were removed, the rest of samples was also clearly separated into four clusters with *R*²(*y*)=0.965 and *Q*²=0.833 in the OPLS-DA model.

Table 1. The predictive ability (*Q*²) and total variance (*R*²[*y*]) of each OPLS-DA model

Figure	Sample no.	Predictive component	Orthogonal component	<i>R</i> ² (<i>y</i>) (cum)	<i>Q</i> ² (cum)	Models
2A	429	5	5	0.875	0.794	For discrimination of GW, GA, JB, PG, CB, and AS
2B	261	3	8	0.965	0.833	For discrimination of JB, PG, CB, and AS
3A	286 (143 ¹⁾)	5	7	0.914	0.797	For prediction of GW, GA, JB, PG, CB, and AS
3B	72 (32)	1	1	0.967	0.913	For prediction of GW and PG
3C	132 (66)	1	3	0.984	0.949	For prediction of GW and CB
3D	84 (42)	1	2	0.981	0.969	For prediction of GW and JB
3E	126 (63)	1	1	0.979	0.973	For prediction of GW and AS
3F	174 (87)	3	8	0.975	0.801	For prediction of JB, PG, CB, and AS
3G	84 (42)	1	4	0.977	0.596	For prediction of PG and CB

OPLS-DA, orthogonal projection to latent structure-discriminant analysis; GW, Gangwon; GA, Gaeseong; JB, Jeonbuk; PG, Punggi; CB, Chungbuk; AS, Anseong.

¹⁾ No. of blind test set.

(Table 1 and Fig. 2B) This result implied that the tested ginseng roots can be discriminated as six regions by two successive steps of OPLS-DA.

Another critical step in a statistical multivariate analysis is to validate a model on samples not used in building the model itself. For the validation of the model, we randomly took out 143 test samples (1/3 of total; 26 samples of GA, 30 samples of GW, 33 samples of AS, 36 samples of CB, 12 samples of JB, and 6 samples of PG) as blind samples and processed the OPLS-DA prediction model. All of the blind samples of GW and GA were correctly belonged to their origins on the predicted score plot (Fig. 3A) except only one sample (GW-88), which was located between GW cluster and PG of mixed cluster. However, GW 88 was also positioned in GW correctly in the OPLS-DA predicted models with GW and PG samples (Fig. 3B). Also, the other prediction models between GW and each of the rest (CB, JB, or AS) conducted the same results that all of the blind samples were restored to the

corresponding origins (Fig. 3C-E). The predictive components, orthogonal components, $R^2(y)$ values, and Q^2 values of described score plots and predictive plots were shown on Table 1.

The predicted OPLS-DA was processed for the prediction of origins in four overlapped regions. The blind samples of JB and AS were perfectly positioned in their own origins; however, samples of PG and CB were not clearly belonged to their own cluster (Fig. 3F). When the samples of PG and CB were compared directly by OPLS-DA model, the prediction of origins was not successful between PG and CB (Fig. 3G). The classification score (Y calculated) of sample data sets and the prediction score (Y predicted) of blind sample data sets of PG and CB were represented in Fig. 4. PG-4 sample, close to CB cluster in Fig. 3G, was positioned in the borderline of 0.5 as the threshold level. The ginsengs of CB and PG origins had been discriminated statistically, even though some CB and PG blind test samples were located to

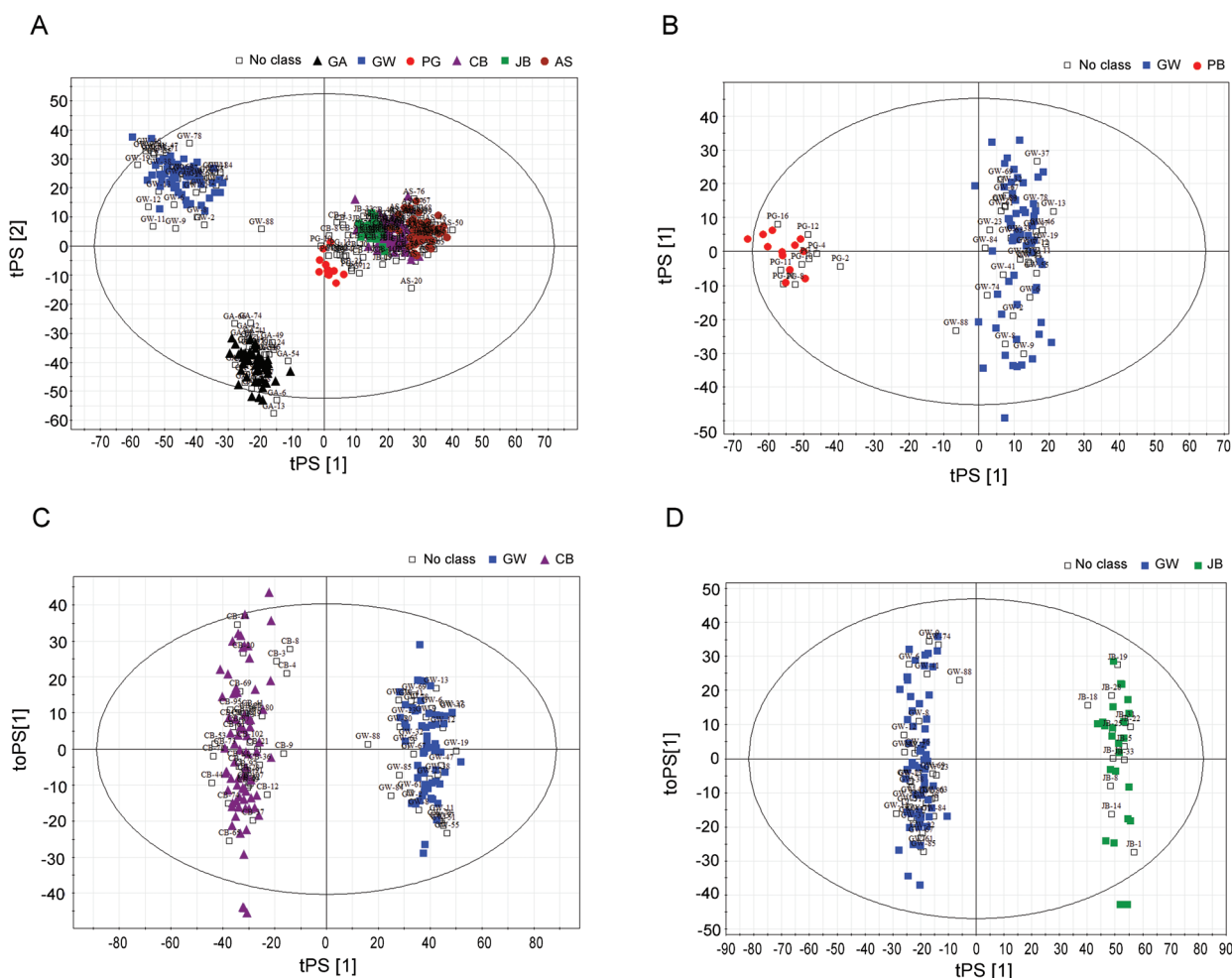


Fig. 3. (Continued)

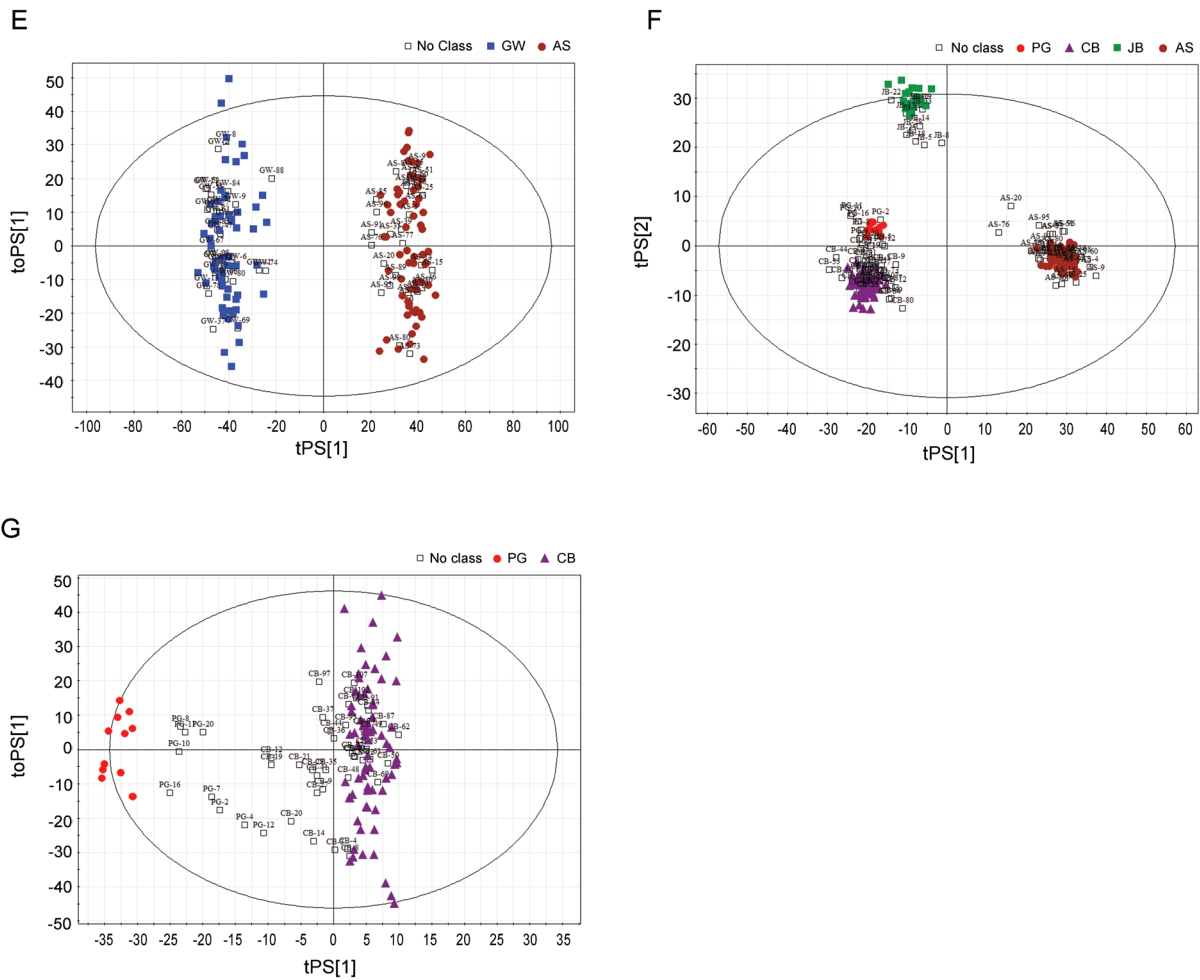


Fig. 3. Predicted score plot of the ginseng for discrimination of geographical origins. (A) Predicted with Gaeseong (GA), Gangwon (GW), Anseong (AS), Chungbuk (CB), Punggi (PG), and Jeonbuk (JB) origins. (B) Predicted with GW and PG origins. (C) Predicted with GW and CB origins. (D) Predicted with GW and JB origins. (E) Predicted with GW and AS origins. (F) Predicted with AS, JB, PG and CB origins. (G) Predicted with PG and CB origins.

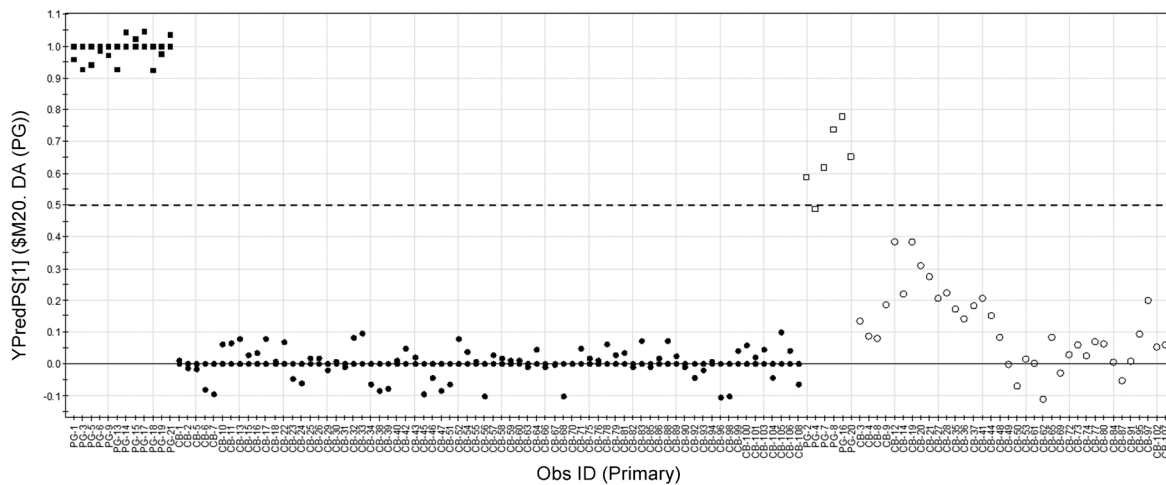


Fig. 4. Prediction of origins of the Punggi and Chungbuk ginseng samples (■ Punggi [PG] ginseng, ● Chungbuk [CB] ginseng, □ no class of Punggi ginseng, and ○ no class of Chungbuk ginseng).

close on borderline. Our results suggested that Korean ginseng could be identified the geographical origin as 99.7% probability. This method was used as a stringent judgment tool in recent discrimination origins and age differentiation of ginseng studies [27-29].

Multivariate models find relations among correlated variables to separate systematic variation from noise. OPLS-DA has the more advantage than an unsupervised PCA method; it separates the predictive variation from the orthogonal variation and can be studied and interpreted separately. In this study, OPLS-DA multivariate analysis showed that the geographical origin of *P. ginseng* cultivated in Korea could be determined by metabolites based on LC-MS data. The discrimination of ginseng from six origins in Korea was successfully performed by OPLS-DA models, which were achieved stage by stage as excluding identified origins from raw data set. Also, each validation of models was applied for predicting their origins with blind test samples that predicted as 99.7% probability. Our results suggested that the growth origin of Korean ginsengs could be discriminated by multivariate models; it could be contributed to development ginseng industry for region characterization in Korea.

ACKNOWLEDGEMENTS

This work was supported from KRIBB Research Initiative Program (KGM1221211). We are grateful to support ginseng samples from Anseong, Baekje, Chungbuk, East Gyeonggi, Geumsan, Jeonbuk, Gaeseong, Gangwon, Ganghwa, and Punggi ginseng cooperative associations.

REFERENCES

1. Akerele O. WHO guidelines for the assessment of herbal medicines. *Fitoterapia* 1992;63:99-104.
2. Huang KC, Williams WM. The pharmacology of Chinese herbs. 2nd ed. Boca Raton: CRC Press, 1998.
3. Sticher O. Getting to the root of ginseng. *Chemtech* 1998;28:26-32.
4. Sengupta S, Toh SA, Sellers LA, Skepper JN, Koolwijk P, Leung HW, Yeung HW, Wong RN, Sasisekharan R, Fan TP. Modulating angiogenesis: the yin and the yang in ginseng. *Circulation* 2004;110:1219-1225.
5. Sievenpiper JL, Arnason JT, Leiter LA, Vuksan V. Decreasing, null and increasing effects of eight popular types of ginseng on acute postprandial glycemic indices in healthy humans: the role of ginsenosides. *J Am Coll Nutr* 2004;23:248-258.
6. Kim N, Kim K, Choi BY, Lee D, Shin YS, Bang KH, Cha SW, Lee JW, Choi HK, Jang DS et al. Metabolomic approach for age discrimination of *Panax ginseng* using UPLC-Q-ToF MS. *J Agric Food Chem* 2011;59:10435-10441.
7. Morita T, Tanaka O, Kohda H. Saponin composition of rhizomes of *Panax japonicus* collected in South Kyushu, Japan, and its significance in oriental traditional medicine. *Chem Pharm Bull (Tokyo)* 1985;33:3852-3858.
8. Sun BS, Xu MY, Li Z, Wang YB, Sung CK. UPLC-Q-TOF-MS/MS analysis for steaming times-dependent profiling of steamed *Panax quinquefolius* and its ginsenosides transformations induced by repetitious steaming. *J Ginseng Res* 2012;36:277-290.
9. Tanaka O, Han EC, Yamaguchi H, Matsuura H, Murakami T, Taniyama T, Yoshikawa M. Saponins of plants of *Panax* species collected in Central Nepal, and their chemotaxonomical significance. III. *Chem Pharm Bull (Tokyo)* 2000;48:889-892.
10. Chassy AW, Bui L, Renaud EN, Van Horn M, Mitchell AE. Three-year comparison of the content of antioxidant microconstituents and several quality characteristics in organic and conventionally managed tomatoes and bell peppers. *J Agric Food Chem* 2006;54:8244-8252.
11. Iniesta MD, Perez-Conesa D, Garcia-Alonso J, Ros G, Periago MJ. Folate content in tomato (*Lycopersicon esculentum*). influence of cultivar, ripeness, year of harvest, and pasteurization and storage temperatures. *J Agric Food Chem* 2009;57:4739-4745.
12. Consonni R, Cagliani LR. Nuclear magnetic resonance and chemometrics to assess geographical origin and quality of traditional food products. *Adv Food Nutr Res* 2010;59:87-165.
13. Deborde C, Maucourt M, Baldet P, Bernillon S, Biais B, Talon G, Ferrand C, Jacob D, Ferry-Dumazet H, de Daruvar A et al. Proton NMR quantitative profiling for quality assessment of greenhouse-grown tomato fruit. *Metabolomics* 2009;5:183-198.
14. Le Gall G, Colquhoun IJ, Davis AL, Collins GJ, Verhoeyen ME. Metabolite profiling of tomato (*Lycopersicon esculentum*) using ¹H NMR spectroscopy as a tool to detect potential unintended effects following a genetic modification. *J Agric Food Chem* 2003;51:2447-2456.
15. Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS). *J Chemom* 2002;16:119-128.
16. Kang J, Choi MY, Kang S, Kwon HN, Wen H, Lee CH, Park M, Wiklund S, Kim HJ, Kwon SW et al. Application of a ¹H nuclear magnetic resonance (NMR) metabolomics approach combined with orthogonal projections to

- latent structure-discriminant analysis as an efficient tool for discriminating between Korean and Chinese herbal medicines. *J Agric Food Chem* 2008;56:11589-11595.
17. Lee EJ, Shaykhtudinov R, Weljie AM, Vogel HJ, Facchini PJ, Park SU, Kim YK, Yang TJ. Quality assessment of ginseng by (1)H NMR metabolite fingerprinting and profiling analysis. *J Agric Food Chem* 2009;57:7513-7522.
 18. 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.
 19. Chan EC, Yap SL, Lau AJ, Leow PC, Toh DF, Koh HL. 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.
 20. Xie GX, Ni Y, Su MM, Zhang YY, Zhao AH, Gao FX, Liu Z, Xiao PG, Jia W. Application of ultra-performance LC-TOF MS metabolite profiling techniques to the analysis of medicinal *Panax* herbs. *Metabolomics* 2008;4:248-260.
 21. 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.
 22. Lee DY, Cho JG, Bang MH, Han MW, Lee MH, Yang DC, Baek NI. Discrimination of Korean ginseng (*Panax ginseng*) roots using rapid resolution LC-QTOF/MS combined by multivariate statistical analysis. *Food Sci Biotechnol* 2011;20:1119-1124.
 23. Mahady GB, Fong HH, Farnsworth NR. Botanical dietary supplements: quality, safety, and efficacy. Lisse: Swets & Zeitlinger Publishers, 2001.
 24. Gao X, Dan M, Zhao A, Xie G, Jia W. Simultaneous determination of saponins in flower buds of *Panax notoginseng* using high performance liquid chromatography. *Biomed Chromatogr* 2008;22:244-249.
 25. Wold S, Trygg J, Berglund A, Antti H. Some recent developments in PLS modeling. *Chemom Intell Lab Syst* 2001;58:131-150.
 26. Bylesjo M, Rantalainen M, Cloarec O, Nicholson JK, Holmes E, Trygg J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J Chemom* 2006;20:341-351.
 27. Kang J, Lee S, Kang S, Kwon HN, Park JH, Kwon SW, Park S. NMR-based metabolomics approach for the differentiation of ginseng (*Panax ginseng*) roots from different origins. *Arch Pharm Res* 2008;31:330-336.
 28. Kim N, Kim K, Lee D, Shin YS, Bang KH, Cha SW, Lee JW, Choi HK, Hwang BY, Lee D. Nontargeted metabolomics approach for age differentiation and structure interpretation of age-dependent key constituents in hairy roots of *Panax ginseng*. *J Nat Prod* 2012;75:1777-1784.
 29. Lee AR, Gautam M, Kim J, Shin WJ, Choi MS, Bong YS, Hwang GS, Lee KS. A multianalytical approach for determining the geographical origin of ginseng using strontium isotopes, multielements, and 1H NMR analysis. *J Agric Food Chem* 2011;59:8560-8567.