LWT - Food Science and Technology 129 (2020) 109454

Contents lists available at ScienceDirect

LWT - Food Science and Technology

journal homepage: www.elsevier.com/locate/lwt

# Discovery study of integrative metabolic profiles of sesame seeds cultivated in different countries

Bo Mi Lee[a], Eun Mi Lee[b], Dong Jin Kang[c], Jeong-Ah Seo[d], Hyung-Kyoon Choi[e], Young-Suk Kim[f,**], Do Yup Lee[b,*]

[a] Department of Bio and Fermentation Convergence Technology, BK21 Plus Program, Kookmin University, Seoul, Republic of Korea
[b] Department of Agricultural Biotechnology, Center for Food and Bioconvergence, And Research Institute for Agriculture and Life Sciences, CALS, Seoul National University, Republic of Korea
[c] National Agricultural Products Quality Management Service, Gimcheon, Republic of Korea
[d] School of Systems Biomedical Science, Soongsil University, Seoul, Republic of Korea
[e] College of Pharmacy, Chung-Ang University, Seoul, Republic of Korea
[f] Department of Food Science and Engineering, Ewha Womans University, Seoul, Republic of Korea

ARTICLE INFO

*Keywords:*
Sesame seed
Primary & secondary metabolites
Region-specific metabolomics features
GC-MS
LC-MS

ABSTRACT

Nutrition-compositional evaluation is critical for quality control and value determination of agricultural products. Sesame seed is an important crop worldwide, enriched with primary nutrients and bioactive compounds. Here, we metabo-typed sesame seeds cultivated in Korea (N = 20) and imported from China (N = 11), South Asia (India and Pakistan), and Africa (Nigeria and Ethiopia) based on 243 metabolites determined by the combination of GC-MS and LC-MS. First, PLS-DA coupled to VIP analysis proposed ten-metabolite recomposite for biomarker. Receiver operating characteristic (ROC) analysis demonstrated the best predictability of the ten-metabolite recomposite according to the four areas (AUCs ranged 0.93–1.00). We further interrogated the unique metabolic features of 20 Korean domestic samples. Unbiased metabo-typing of the profiles showed the strong association according to the longitudinal coordination of the cultivated regions. The resultant 3 groups presented the regional specificity in metabolite contents based on chemical enrichment analysis. Particularly, the northwest region was featured by enriched amino acids, sugar alcohols, and pyrrolidinones. The central region presented relatively higher contents in various types of sugar alcohols, indoles, and fatty acids whereas the east province and south coast showed higher contents in dicarboxylic acids and dipeptides.

## 1. Introduction

Sesame (*Sesamum indicum* L.) is an important oilseed crop worldwide, and considered as an good resource for bioactive compounds, including unsaturated fatty acid, protein, vitamin B1, mineral, dietary fiber, antioxidants, and other beneficial plant compounds (Namiki, 2007; Zhang et al., 2019). Previous studies have reported that sesame seeds and the oil products may have hypocholesterolemic effects, help anticancer activity, help lower blood pressure, and prevent oxidative stress attenuation (Elleuch, Bedigian, & Zitoun, 2011).

The bioactivities may vary according to nutritional characters, which are influenced by multiple factors such as genotype (variety), environment, and mostly their interaction. Like other agricultural products, sesames have been bred and domesticated for a long time (more than 3000 years). Thus, physiological diversity including nutritional composition cannot be classified or predicted by morphological characters or DNA marker (Laurentin, Ratzinger, & Karlovsky, 2008). Alternatively, omics technology (e.g. genomics, transcriptomics, and metabolomics) is more adequate approach to identify biological trait and to accordingly categorize them.

Among the technologies, metabolomics provides a direct information on a range of metabolic composition inaccessible by other molecular profiling. Indeed, metabolic profiling allows precise evaluation of compound characters, the final readout through biochemical activity interacted with genetic and environmental factors (E. M. Lee et al., 2019). The comprehensive information on metabolite profiles can be translated for nutritional identity in food matrix and authenticity of cultivation origin of agricultural products.

In our current study, we performed untargeted metabolite profiling of sesame seeds cultivated in 20 provinces of Korea and 5 major

* Corresponding author. Department of Agricultural Biotechnology, Seoul National University, Republic of Korea.
** Corresponding author. Department of Food Science and Engineering, Ewha Womans University, Seoul, Republic of Korea.
*E-mail addresses:* yskim10@ewha.ac.kr (Y.-S. Kim), rome73@snu.ac.kr (D.Y. Lee).

https://doi.org/10.1016/j.lwt.2020.109454
Received 27 September 2019; Received in revised form 16 April 2020; Accepted 17 April 2020
Available online 23 April 2020
0023-6438/ 

exporters to Korea (China, India, Pakistan, Nigeria, and Ethiopia). We aimed first to construct robust biomarker model for discriminating Korean domestic ones from the imported ones. Subsequently, we deepened the understanding of region-specific features in metabolite profiles among 20 Korean domestic sesames. The result suggested the comprehensive metabolite profiling may be an applicable and promising approach to origin declaration of agricultural product and ingredient quality for producers, manufacturers, and consumers.

## 2. Materials and methods

### 2.1. Sample collection

All Korean samples were harvested in 2017 year and collected by the National Agricultural Products Quality Management Service. The samples were cultivated in Icheon (Gyeonggi-do), Inje (Gangwon-do), Goesan, Cheongju, Chungju (Chungcheongbuk-do), Dangjin, Gongju, Seocheon (Chungcheongnam-do), Namwon, Jeonju, Wanju, Iksan (Jeollabuk-do), Yeosu, Muan, Gwangju (Jeollanam-do), Yeongju, Andong (Gyeongsangbuk-do), Hapcheon, Jinju (Gyeongsangnam-do), and Jeju (Jejudo). The variety information was retrospectively obtained and summarized in Table 1.

Chinese sesame seeds harvested in 2017 were purchased from online suppliers (500–600 g) and air-transported. The regions were Tonghua (Jilin province), Zhumadian (Henan province), Huainan (Anhui province), and Liuzhou (Guangxi Zhuangzu autonomous region) (Table S1). Sesame seeds from India, Pakistan, Nigeria, and Ethiopia (500–600 g) harvested in 2017 were obtained from Ottogi Co., Ltd, CJ CheilJedang Co., Ltd, and Daesang Co., Ltd, which are the major suppliers for imported sesame in Korea designated by Korea Agro-Fisheries Trade Corporation (http://www.at.or.kr/home/apen000000/index.action). All samples were stored at −80 °C in a deep freezer until use. Overall workflow was provided in Fig. 1, which included the sample information, mass-spectrometric analysis, and statistical analysis conducted in this study.

### 2.2. Metabolite extraction

Sesame seeds (20 mg per sample) were lyophilized and ground using Mixer Mill MM400 (Retsch GmbH & Co., Germany). The fine-powders were mixed with 1 mL of extraction solvent (methanol:isopropanol:water, 3:3:2, v/v/v). The mixtures were sonicated for 5 min and centrifuged for 5 min (16,100 × *g* at 4 °C). The supernatants were aliquoted and then transferred to new 1.5 mL tubes (200 μL for GC-TOF MS and 200 μL for LC-Orbitrap MS). The aliquots were concentrated to complete dryness in a speed vacuum concentrator (SCANVAC, Korea).

### 2.3. GC-TOF MS analysis

Dried extracts were derivatized as follows: 1) 5 μL of 40 mg/mL methoxyamine hydrochloride (Sigma-Aldrich, St. Louis, MO, USA) in pyridine (Thermo, USA) (90 min at 800 rpm at 30 °C) and 2) 45 μL of N-methyl-N-trimethylsilyltrifluoroacetamide (MSTFA + 1% TMCS; Thermo, USA) for 1 h (Fiehn et al., 2008; E. M.; Lee et al., 2019). A mixture of 13 fatty acid methyl esters (FAMEs, C8 – C30) was added to the derivatives as retention index marker that corrected retention time shift during GC-TOF MS analysis. The derivatives (0.5 μL) were injected by an Agilent 7693 ALS (Agilent Technologies, Wilmington, DE, USA) in splitless mode. The metabolites were separated by an RTX-5Sil MS column (Restek, Gellefonte, PA, USA) and Agilent 7890B gas chromatography (Agilent Technologies). Helium (99.9999% pure) flow was set to 1 mL/min. The oven temperature was set to 50 °C for 1 min, ramped at 20 °C/min to 330 °C, and held for 5 min.

Mass spectrometry was conducted on a Leco Pegasus HT time of flight mass spectrometer controlled by Chroma TOF software 4.50 version (LECO, St. Joseph, MI, USA). Mixtures of 30 compounds were analyzed every 10 samples for quality control purpose (Ji et al., 2018). The transfer line temperature and ion source temperature were set to 280 °C and 250 °C, respectively. Mass spectra were collected ranging from m/z 85–500 at 17 spectra s-1 and 1850 V detector voltage after 300 s solvent delay.

Pre-processing was done using ChromaTOF software, in which apex mass, spectrum, peak purity, S/N, and retention time were collected and exported to server computer. *.peg file was converted to *.txt file and to *.netCDF file for following data evaluation, *Binbase* algorithm. The algorithm processed raw data file (.txt file) using the criteria as follows: less than 10 peaks with intensity higher than $10^7$ counts $s^{-1}$ for chromatogram validation, spectra similarity score above 800 for RI detection, 5th order polynomial regression-based retention index computation, retention time window of ± 2 s, and unique mass validation against apex mass (Fiehn, Wohlgemuth, & Scholz, 2005; E. M.; Lee et al., 2019; J.-E.; Lee, Cho, Kim, & Lee, 2016).

Quantitative value for each metabolite was computed using on peak height based on a quant ion. The resultant data was tabulated for all metabolite entries that were identified against Fiehn library (Kind et al., 2009). The detailed information on processed data is provided in supplementary files, which include compound name, retention time, quant mass, InChiKeys, Pubchem ID, and SMILES. Compound mixture with 30 metabolites was analyzed for quality control purpose. The data for the quality control is provided in supplementary information.

### 2.4. LC-orbitrap MS analysis

The dried metabolites were reconstituted with 100 μL of 80% MeOH. The reconstituents were separated by a Waters Acquity UPLC BEH C18 column (2.1 mm × 150 mm, 1.7 μm) and Ultimate-3000 UPLC system (Thermo Fisher Scientific, MA, USA). The mobile phase A was water (0.1% formic acid, v/v) and B was acetonitrile with 0.1% formic acid (v/v). The gradient was as follows: 0–2.0 min, 0% B; 2.0–30.0 min, 0%–100% B; 30.0–32.0 min, 100% B; 32.0–32.1 min, 100%–0% B; 32.1–35.0 min, 0% B (E. M. Lee et al., 2019; Y. Li et al., 2017). Mass spectrometry was performed on a Q-Exactive Plus instrument (Thermo Fisher Scientific, MA, USA) in positive mode controlled by Xcalibur 4.0 and Q-Exactive Tune software. Full scan (MS1) was

**Table 1**
The list of origin of sesame seeds cultivated in Korea.

| Country | No. | Province | City | Variety |
|---|---|---|---|---|
| Korea | 1 | Gyeonggi-do | Icheon | Danbaek 24%, Yangbaek 22%, Ansan 18% |
| | 2 | Gangwon-do | Inje | Danbaek 21%, Yangbaek 20%, Ansan 24% |
| | 3 | Chungcheongbuk-do | Goesan, Cheongju, Chungju | Danbaek 26%, Yangbaek 21%, Ansan 18% |
| | 4 | Chungcheongnam-do | Dangjin, Gongju, Seocheon | Danbaek 28%, Yangbaek 23%, Ansan 14% |
| | 5 | Jeollabuk-do | Namwon, Jeonju, Wanju, Iksan | Danbaek 21%, Yangbaek 23%, Ansan 21% |
| | 6 | Jeollanam-do | Yeosu, Muan, Gwangju | Danbaek 25%, Yangbaek 23%, Ansan 16% |
| | 7 | Gyeongsangbuk-do | Yeongju, Andong | Danbaek 25%, Yangbaek 27%, Ansan 13% |
| | 8 | Gyeongsangnam-do | Hapcheon, Jinju | Danbaek 21%, Yangbaek 23%, Ansan 21% |
| | 9 | Jejudo | Jeju | Local variety |

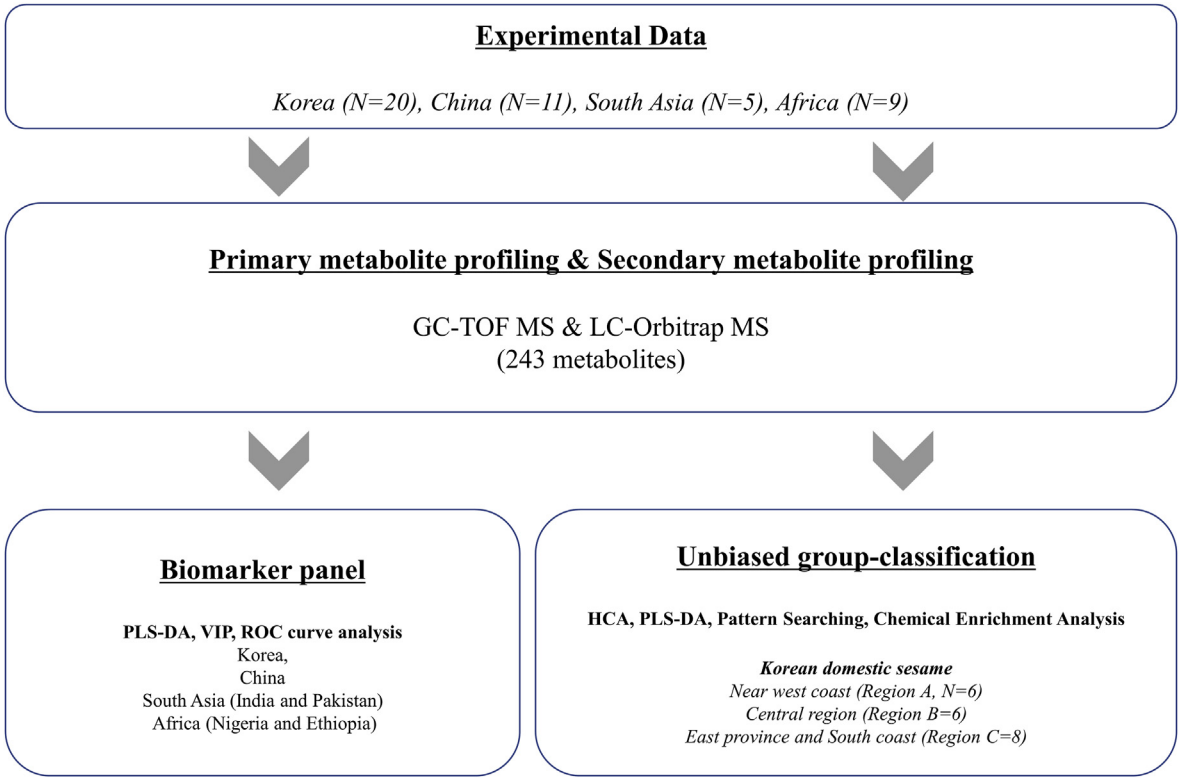



**Fig. 1.** Experimental workflow sample information, mass-spectrometric analysis, and statistical analysis conducted in this study.

ranged from 100 to 1500 Da with resolution of 70,000 FWHM and MS/MS was done in data-dependent manner (HCD: 30 eV, resolution: 17,500 FWHM). Data process was done using Compound Discover (ver 2.0, Thermo Fisher Scientific, USA). Criteria for compound identification were as follows: 1) mass tolerance of 5 ppm and maximum retention time shift of 1 min for peak alignment 2) MS1 tolerance of 5 ppm, MS2 tolerance of 10 ppm, assignment threshold of 70% similarity for identification parameter. Relative intensity for each metabolite was calculated using peak area based on base peak chromatogram (BPC) corresponding to mass-to-charge ratio on specific retention time that was identified. The detailed information on processed data is provided in supplementary files, which include compound name, MS/MS fragments ion spectrum, extracted mass, predicted formula, molecular weight, RT [min], area (Max.), *mzCloud* best match score, InChiKeys, Pubchem ID and SMILES. The pooled samples were analyzed for the quality control and the data is provided in supplementary information.

### 2.5. Statistical analysis

Statistical analysis was performed on all continuous variables (243 identified compounds). The profiles from GC-TOF MS and LC-Orbitrap MS were normalized based on MSTUS method (Karpievitch, Nikolic, Wilson, Sharman, & Edwards, 2014) implemented in NOREVA (http://idrb.zju.edu.cn/noreva/(B. Li et al., 2017). Student's *t*-test was performed for statistical significance using EXCEL (Microsoft Office 2010). Following autoscaling (mean-centered and divided by the standard deviation of each variable), principal component analysis (PCA), partial least squares-discriminant analysis (PLS-DA), and variable importance projection (VIP) analysis were conducted SIMCA 15 (Umetrics AB, Umea, Sweden). Hierarchical clustering analysis (HCA) with dendrogram plot was performed using MetaboAnalyst (distance measure, Pearson correlation; clustering algorithm, complete linkage) (Chong et al., 2018). Validation of PLS-DA model was conducted based on random permutation plot (100 times iteration). The result is generally determined by the criteria including Y-axis intercepts and slope in which all permuted R2 and Q2 values (the left on x-axis) are lower than the original values on the right. ANOVA testing of cross-validated predictive residuals (CV-ANOVA) was performed for assessing the reliability of PLS-DA model. Chemical enrichment analysis was applied to acquire chemical cluster with statistical significance (p value) and median of XlogP value based on ChemRICH (Barupal & Fiehn, 2017; E. M.; Lee et al., 2019).

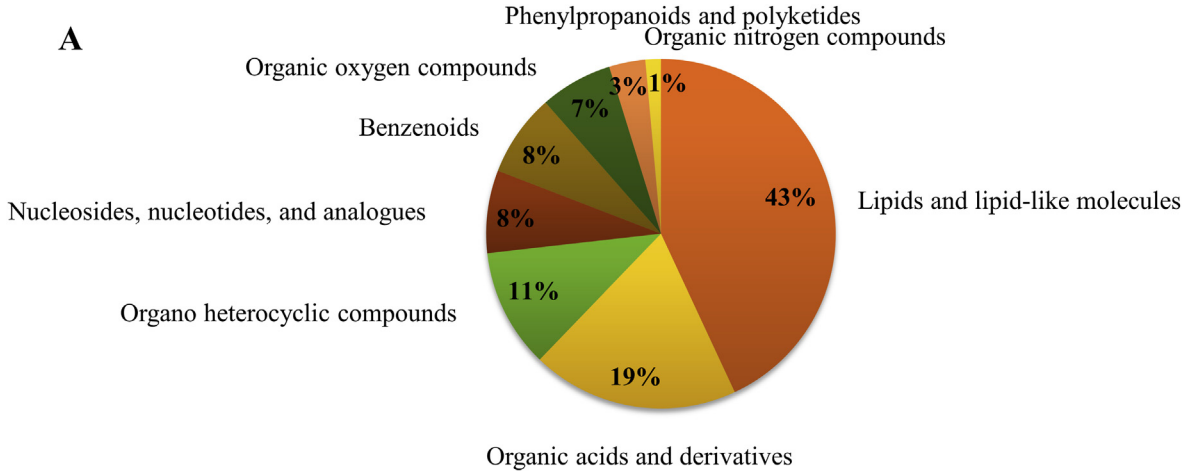
## 3. Results

### 3.1. Unique metabolic traits of sesame seeds cultivated in Korea and 5 other countries

Untargeted profiling of primary and secondary metabolite was conducted using GC-TOF MS and LC-Orbitrap MS. The integrative MS analysis determined 243 compounds. Raw and processed data is available at [http://calslab.snu.ac.kr/lms2/board.list?mcode=1312&cate=1312]. The chemical ontology analysis systematically categorized the identified metabolites as depicted in Fig. 2A. The uppermost portion was lipids and lipid-like molecules (43%, 90 compounds) and the next ones were organic acids (and derivatives) and organo-heterocyclic compounds (Fig. 2B).

First, unsupervised multivariate statistic, PCA suggested the distinctive profiles particularly between the Korean sesames and the imported ones mainly by pc1 (Fig. S1). The subsequent PLS-DA showed the high levels of explained variance ($R^2Y$ = 0.816) and prediction validity ($Q^2$ = 0.591) (Fig. 3A). Loading scatter plot presented the Korean sesame was best characterized by the high correlation with a range of amino acids (e.g. threonine, tyrosine, valine, glutathione, isoleucine, glutamine, and serine) (Fig. S2). The CV-ANOVA and random permutation plot (100-time iteration) validated the performance of the discriminant model for the four international regions, Korea, China, South Asia (India and Pakistan), and Africa (Nigeria and Ethiopia) (Fig. S2).

A




B

| Superclass | Class |
|---|---|
| Lipids and lipid-like molecules (90) | Fatty Acyls (82) |
| | Prenol lipids (5) |
| Organic acids and derivative (40) | Carboxylic acids and derivatives (36) |
| | Hydroxy acids and derivatives (3) |
| Organo heterocyclic compounds (23) | Indoles and derivatives (7) |
| | Pyridines and derivatives (5) |
| | Imidazopyrimidines (4) |
| Nucleosides, nucleotides, and analogues (16) | Purine nucleotides (10) |
| Benzenoids (16) | Benzene and substituted derivatives (15) |
| Organic oxygen compounds (14) | Organooxygen compounds (13) |
| Phenylpropanoids and polyketides (7) | Cinnamic acids and derivatives (3) |
| Organic nitrogen compounds (3) | Organonitrogen compounds (3) |

**Fig. 2.** Chemical classification of identified metabolites in sesame seeds. The classification was done based on chemical taxonomy provided by HMDB (http://www.hmdb.ca). A total of 243 compounds (86%) were categorized into 8 super classes (A) and 12 classes (B).

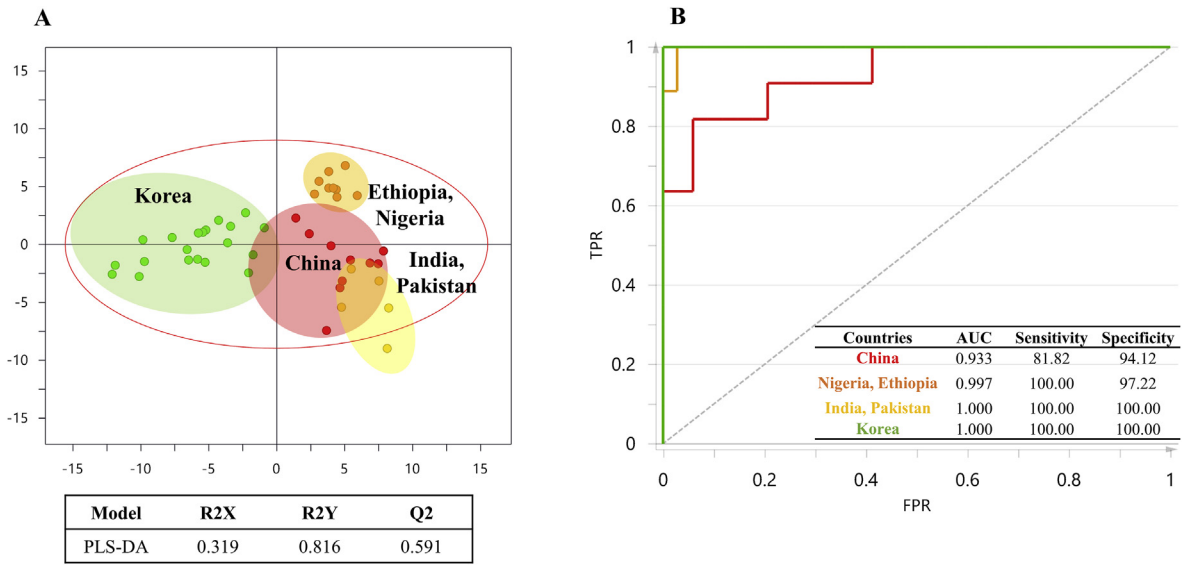
For biomarker construction, ten metabolites were determined based on VIP analysis including glycerol, thiamine, scopoletin, carnitine, raffinose, glycerol α-phosphate, palmitic acid, 1-monopalmitin, 9-Oxo-10(E),12(E)-octadecadienoic acid, and histamine (VIP > 1.5). (Fig. S3). The predictive model of the ten-metabolite recomposite showed the highest level of discrimination power in which AUCs ranged from 0.933 to 1.000 for all pairwise comparison (Fig. 3B). Particularly, the distinctiveness of the Korean and South Asian ones showed perfect discrimination with 100% of sensitivity and specificity. Although potential confounding effects on the metabolomic profiles may exist (e.g. post-harvest processing and representativeness), the result showed the significant difference in the chemical composition of the Korean sesame, which may be applied to monitoring the domestically cultivated and commercially distributed sesames from the imported ones.

### *3.2. The metabolic features of sesames cultivated in 20 different domestic regions of Korea*

Since the sesames cultivated in Korea showed the similar level of variety mixture as described in materials and methods, we assumed that the metabolomic physiology would reflect regional specificity (e.g. environmental traits). Thus, we first performed unbiased group-classification based on hierarchical clustering analysis to interrogate latent association between metabolite composition and geographical characteristics (Fig. 4A). Indeed, we observed close linkage of the metabolomic profiles along with longitudinal coordination (Fig. 4B). Subsequent PLS-DA model validated goodness of fit ($R^2Y$ = 0.940 and $Q^2$ = 0.732) (Fig. 4C). Next, we determined the metabolic features specific to pre-defined region (Region A, B, and C) based on *Pattern-Hunter* module in *MetaboAnalyst* (E. M. Lee et al., 2019; Xia & Wishart, 2016). The statistical analysis identified the metabolites that presented either significantly higher (positive correlation) or lower abundances (negative correlation) in specific group compared to the others. The correlated metabolites were then systematically classified based on chemical structural similarity and ontology mapping (Fig. 5).

Six regions were classified as Region A, which were geographically located near northwest area. The area was best featured by relatively higher contents in the compounds with higher hydrophilicity (x-axis, lower XlogP values). The list included various types of amino acids, particularly, sulfur amino acids (cysteine, methionine, and methionine sulfoxide). Others were sugar alcohols, pyrrolidinones, which was contrast to relatively lower contents in a range of indoles (indole-3-pyruvic acid, *N*-methyltryptamine, skatole, indirubin, tryptoline, indole-3-acrylic acid), coumarins (4-hydroxycoumarin, 7-hydroxycoumarine, coumarin, esculetin, scopoletin) (Fig. S4).

On contrary, Region B, the central region of Korea presented relatively lower contents in amino acids (beta alanine, threonine, serine, and homoserine), dipeptides, purine nucleosides, and cinnamates




| Model | R2X | R2Y | Q2 |
|---|---|---|---|
| PLS-DA | 0.319 | 0.816 | 0.591 |

**Fig. 3.** Multivariate statistical model that discriminated the international origin of sesame seed profiles. (A) The score plot of the sesame profiles based on partial least squares-discriminant analysis (PLS-DA) The level of explained variance and prediction validity were presented ($R^2Y$ = 0.816) and prediction validity ($Q^2$ = 0.591). (B) Receiver operating characteristic (ROC) curve analysis for the biomarker panel composed of 10 metabolites (glycerol, thiamine, scopoletin, carnitine, raffinose, glycerol α-phosphate, palmitic acid, 1-monopalmitin, 9-Oxo-10(E),12(E)-octadecadienoic acid, and histamine).

whereas higher levels were determined in sugar alcohols, indoles, and saturated FA. Sugar alcohols were glycerol, myo-inositol, xylitol, lyxitol, lactitol, erythritol, arabitol, and iditol (Fig. S5).

The metabolite profiles of sesame seed cultivated in Region C, located in east province and south coast, showed characteristic enrichment in various types of dicarboxylic acids. The list included argininosuccinic acid, fumaric acid, malic acid, oxalic acid and succinic acid. Likewise, higher contents in dipeptides and lower contents in sugar alcohols were unique metabolic feature in this area, which was contrast to the profiles of Region B (Fig. S6). The dipeptides included glycylproline, L-saccharopine, leucylproline, valylproline, α-Aspartylphenylalanine.

## 4. Discussion

The current study laid emphasis on two perspectives. The first aim was to discriminate the domestically-grown and –distributed ones (Korean sesame seeds) from the ones imported from China, South Asia (India and Pakistan), and Africa (Ethiopia and Nigeria). Indeed, we proposed the biomarker panel recomposed with 10 metabolites that showed the highest levels of prediction power for the Korean ones from the imported ones. There were undissolved factors (e.g. storage, transportation, and variety) among the cultivation countries that may have influenced the nutritional compositions. Nonetheless, we used imported sesame seeds designated by Korea Agro-Fisheries Trade Corporation, which authenticated the country of origin and year, mandatory information by import origin labeling regulation. Thus, the predictive model may be applied to the determination method for deceptive marking of cultivation country that have been reported for various types of agricultural products.

The second goal was to characterize and compare the metabolic profiles of the sesame seeds domestically cultivated in Korea. They were not cultivated as single variety, but the co-cultivation ratio of multiple variety was comparably similar among the different regions. Besides, the samples were collected and stored under the control of the Korean government agency for research purpose. The fact provided the plausibility that the metabolic uniqueness we identified in the study was mainly affected by environmental factors. Indeed, we dissolved and characterized the sesame profiles among 20 different samples of Korea, in which Danbaek, Yangbaek (breeding from Danbaek), and Ansan were co-cultivated with marginally different ratios (approximately 24%, 23%, and 18%, respectively) with other varieties including wild types (the rest of 35%). The distinctive metabolic profiles of the varieties with “heterogenetic homogeneity” cultivated in localized area suggests the substantial effects of environmental factors on nutritional characters, which in turn can be applied to metabolic markers of regional distinctiveness in nutritional quality accompanied by more detailed geography-association parameter (e.g. soil trait).

Genotyping technique (e.g. NGS) provides the authentic information on variety, but it hardly reflects environmental effects; thus, the application is limited by the cases in which 1) identical variety is cultivated in different regions and 2) more than two varieties or wild types are cultivated. Stable isotope ratio analysis has been regarded as the best tool for determining geographic origins of various agricultural products including sesame seeds since stable isotope signals ($^{13}C$) are affected by the environmental factors (Horacek et al., 2015). The isotope analysis acquires more stable signal and requires less sample preparation; however, relatively small number of variables is applicable for the discriminant model construction due to the limitation of labeling atoms.

On contrary, several hundreds of metabolic signatures determined by un-targeted metabolomic profiling can potentially be candidacy for discriminant variable, which facilitates the construction of biomarker re-composite. In addition to determining geographic origin, the technology is ideal to globally characterize metabolic traits and to evaluate the nutritional value like the second aim of our study. Accordingly, we identified the health-beneficial compounds that showed region-specific abundances. For instance, hydroxycinnamic acids (coumaric acid and ferulic acid) highly-enriched in Region A are free radical scavengers (Graf, 1992). Methylated flavone, nobiletin the most abundant in Region B has been reported for multifunctional effects in vivo and in vitro including neural, cardiovascular, respiratory, and digestive systems protection (Huang et al., 2016). Sinapinic acid, a featured compound in Region C has been associated with natural antioxidant, antiinflammation, anticancer efficacy (Chen, 2016).

For conclusive discovery, our exploratory findings should be validated considering sample size, yearly changes in climate, and investigation on the causality (e.g. geographic trait). Particularly, metabolic phenotyping research is often performed in untargeted and hypothesis-free manner without prior knowledge of the molecular

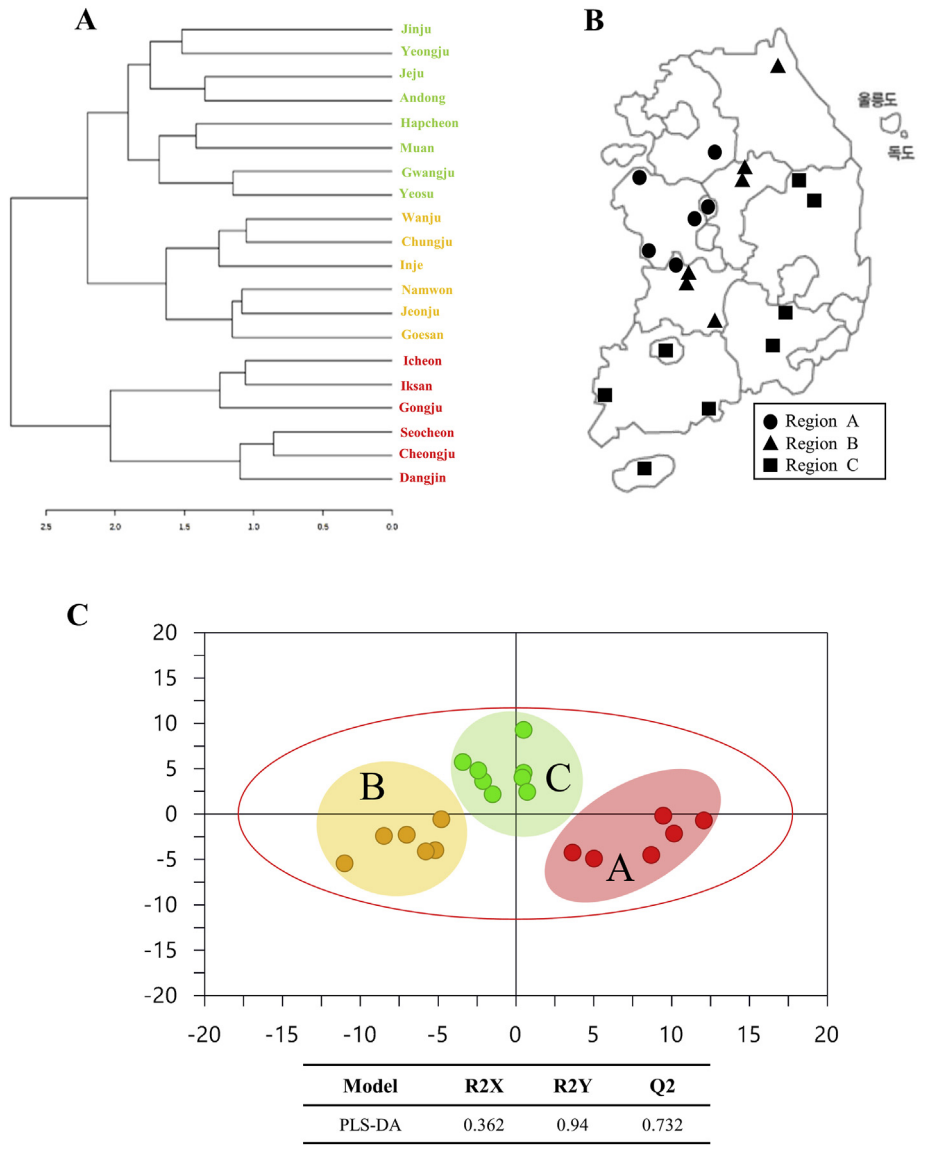



**Fig. 4.** Metabolomic phenotyping of sesame seeds cultivated in 20 different Korean provinces (A) Unbiased clustering analysis of the profiles based on hierarchical clustering analysis (HCA). (B) The resultant geographic mapping of the location of the cultivation regions (C) Score plot of the sesame profiles using PLS-DA model.

targets (Blaise et al., 2016) which limit meaningful calculation of statistical power before experiment. However, the proper sample size would be ninety based on univariate statistics (e.g. ANOVA) to meet false discovery rate of 0.2, which needs to be considered in validation study.

## 5. Conclusion

Metabolomics provide unique information on metabolite contents that vary according to genetic and environmental factors, which is inaccessible by other molecular profiling. Thus, comprehensive metabolite profiling of agricultural products can provide biomarkers for determination of cultivation origin and re-evaluate the nutritional significance. Accordingly, based on integrative metabolomic profiling by GC- and LC-MS, we constructed the discriminant model for the Korean sesame seeds from the other countries' ones (China, India, Pakistan, Nigeria, and Ethiopia). The biomarker signature can be applied for authentic methodology to guarantee the interest of consumers that often prefer domestic agricultural product to imported ones. In addition, we evaluated the sesame profiles of 20 different samples that were domestically cultivated in Korea. The unbiased classification proposed the longitudinal coordination of the metabolite profiles, and chemical enrichment analysis identified the characteristic quantity in the metabolite contents.

## CRediT authorship contribution statement

**Bo Mi Lee:** Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Eun Mi Lee:** Formal analysis, Investigation. **Dong Jin Kang:** Validation, Investigation, Resources. **Jeong-Ah Seo:** Software, Formal analysis, Visualization. **Hyung-Kyoon Choi:** Methodology, Validation, Writing - review & editing. **Young-Suk Kim:** Conceptualization, Methodology, Validation, Writing - original draft, Writing - review & editing, Project administration, Funding acquisition. **Do Yup Lee:** Conceptualization, Methodology, Software, Validation, Data curation, Writing - original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition.

## Acknowledgments

This work was supported by Korea Institute of Planning and

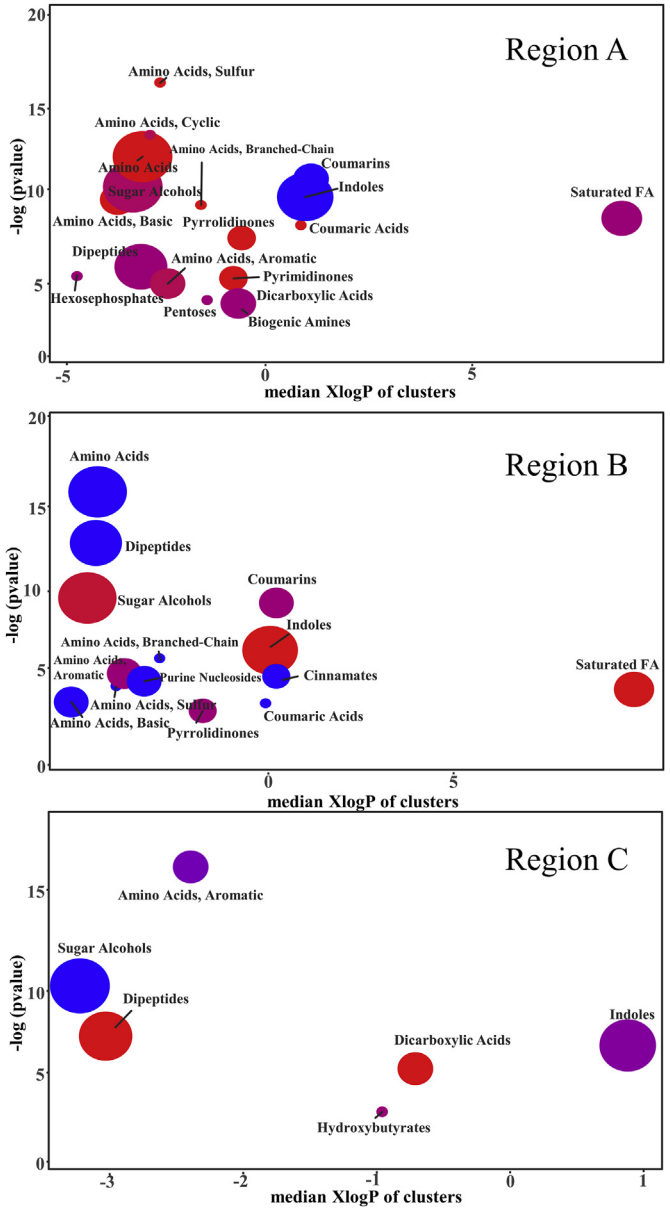



**Fig. 5.** Chemical similarity enrichment analysis of 243 compounds in the four cultivation regions. *X*-axis indicates partition coefficient of clusters, and *y*-axis presents statistical significance (Kolmogorov–Smirnov test). Node color was determined by the ratio of higher or lower abundant metabolites in each cluster (red = higher, blue = lower, purple = mixed). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Evaluation for Technology in Food, Agriculture, Forestry and Fisheries (IPET) through Advanced Production Technology Development Program (316081–04). This work was also funded by the National Research Foundation of Korea (NRF-2019R1A2C1007264) funded by the Korean Ministry of Science, ICT and Future Planning.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.lwt.2020.109454.

## References

Barupal, D. K., & Fiehn, O. (2017). Chemical Similarity Enrichment Analysis (ChemRICH) as alternative to biochemical pathway mapping for metabolomic datasets. *Scientific Reports, 7*(1), 14567.

Blaise, B. J., Correia, G., Tin, A., Young, J. H., Vergnaud, A.-C., Lewis, M., ... Holmes, E. (2016). Power analysis and sample size determination in metabolic phenotyping. *Analytical Chemistry, 88*(10), 5179–5188.

Chen, C. (2016). Sinapic acid and its derivatives as medicine in oxidative stress-induced diseases and aging. *Oxidative Medicine and Cellular Longevity,* 3571614 2016.

Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., ... Xia, J. (2018). MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. *Nucleic Acids Research, 46*(W1), W486–W494.

Elleuch, M., Bedigian, D., & Zitoun, A. (2011). Sesame (*Sesamum indicum L.*) seeds in food, nutrition, and health. *Nuts and seeds in health and disease prevention* (pp. 1029–1036). Elsevier.

Fiehn, O., Wohlgemuth, G., & Scholz, M. (2005). Setup and annotation of metabolomic experiments by integrating biological and mass spectrometric metadata. *International Workshop on Data Integration in the Life Sciences,* 224–239.

Fiehn, O., Wohlgemuth, G., Scholz, M., Kind, T., Lee, D. Y., Lu, Y., ... Nikolau, B. (2008). Quality control for plant metabolomics: Reporting MSI-compliant studies. *The Plant Journal, 53*(4), 691–704.

Graf, E. (1992). Antioxidant potential of ferulic acid. *Free Radical Biology and Medicine, 13*(4), 435–448.

Horacek, M., Hansel-Hohl, K., Burg, K., Soja, G., Okello-Anyanga, W., & Fluch, S. (2015). Control of origin of sesame oil from various countries by stable isotope analysis and DNA based markers—a pilot study. *PloS One, 10*(4).

Huang, H., Li, L., Shi, W., Liu, H., Yang, J., Yuan, X., et al. (2016). The multifunctional effects of nobiletin and its metabolites in vivo and in vitro. *Evidence-based Complementary and Alternative Medicine, 2016,* 14 2918796.

Ji, D. Y., Park, S. H., Park, S. J., Kim, K. H., Ku, C. R., Shin, D. Y., ... Lee, E. J. (2018). Comparative assessment of Graves' disease and main extrathyroidal manifestation, Graves' ophthalmopathy, by non-targeted metabolite profiling of blood and orbital tissue. *Scientific Reports, 8*(1), 9262.

Karpievitch, Y. V., Nikolic, S. B., Wilson, R., Sharman, J. E., & Edwards, L. M. (2014). Metabolomics data normalization with EigenMS. *PloS One, 9*(12), e116221.

Kind, T., Wohlgemuth, G., Lee, D. Y., Lu, Y., Palazoglu, M., Shahbaz, S., et al. (2009). FiehnLib: Mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Analytical Chemistry, 81*(24), 10038–10048.

Laurentin, H., Ratzinger, A., & Karlovsky, P. (2008). Relationship between metabolic and genomic diversity in sesame (Sesamum indicum L.). *BMC Genomics, 9*(1), 250.

Lee, J.-E., Cho, Y. U., Kim, K. H., & Lee, D. Y. (2016). Distinctive metabolomic responses of Chlamydomonas reinhardtii to the chemical elicitation by methyl jasmonate and salicylic acid. *Process Biochemistry, 51*(9), 1147–1154.

Lee, E. M., Park, S. J., Lee, J.-E., Lee, B. M., Shin, B. K., Kang, D. J., ... Lee, D. Y. (2019). Highly geographical specificity of metabolomic traits among Korean domestic soybeans (*Glycine max*). *Food Research International, 120*, 12–18.

Li, Y., Jin, Y., Yang, S., Zhang, W., Zhang, J., Zhao, W., ... Lu, K. (2017). Strategy for comparative untargeted metabolomics reveals honey markers of different floral and geographic origins using ultrahigh-performance liquid chromatography-hybrid quadrupole-orbitrap mass spectrometry. *Journal of Chromatography A, 1499*, 78–89.

Li, B., Tang, J., Yang, Q., Li, S., Cui, X., Li, Y., ... Zhu, F. (2017). Noreva: Normalization and evaluation of MS-based metabolomics data. *Nucleic Acids Research, 45*(W1), W162–W170.

Namiki, M. (2007). Nutraceutical functions of sesame: A review. *Critical Reviews in Food Science and Nutrition, 47*(7), 651–673.

Xia, J., & Wishart, D. S. (2016). Using MetaboAnalyst 3.0 for comprehensive metabolomics data analysis. *Current Protocols in Bioinformatics, 55*(1) 14.10. 11-14.10. 91.

Zhang, Y., Li, D., Zhou, R., Wang, X., Dossa, K., Wang, L., ... Zhang, X. (2019). Transcriptome and metabolome analyses of two contrasting sesame genotypes reveal the crucial biological pathways involved in rapid adaptive response to salt stress. *BMC Plant Biology, 19*(1), 66.