scientific reports

OPEN

Development of a leaf metabolite-based intact sample distinguishing algorithm for the three varieties of *Panax Vietnamensis*

Ranran Cheng[1], Young Cheol Yoon[1], Cheol Woon Jung[1], Tae Ha Kim[1], Qiang Wang[2], Woohyeon Cho[3], Tae-Jin Yang[3], Thi Hong Van Le[4], Chan Jae Cho[5], Jae Hyun Kim[5], Gyu Hwan Hyun[6], Jeong Hill Park[6], Sung Won Kwon[1,6]✉ & Sun Jo Kim[7]✉

Panax vietnamensis, indigenous to Vietnam and southern China, is classified into three subspecies: *Panax vietnamensis* Ha et Grushv. (PVV), *Panax vietnamensis* var. *fuscidiscus* (PVF), and *Panax vietnamensis* var. *langbianensis* (PVL). A method to distinguish these varieties in their intact form is absent, which poses a possible risk of misclassification. Here, we aimed to devise a plant metabolite-based discrimination algorithm for the three varieties, without causing significant damage to individual plants. A multivariate analysis on mass spectral data of PVV, PVF, and PVL revealed that a peak at *m/z* 426, which was subsequently identified as an indole alkaloid glycoside, was exclusive to PVF and therefore clearly distinguished PVF from PVV and PVL. Additionally, global metabolic profiling was conducted to elucidate the discrimination markers between PVV and PVL, and lysophospholipids and hydroxy fatty acids were selected as potential discrimination markers. The performance of these markers was validated by cross-validation using machine learning algorithm.

Ginseng has been used in traditional medicine for over 4000 years due to its multifaceted effects in physiological processes such as cancer treatment[1,2], immune enhancement[3], insulin resistance prevention[4], and hypertension management[5]. *Panax vietnamensis*, naturally occurring in the Ngoc Linh and Hoang Lien Son mountainous regions of Vietnam, is renowned for its abundant medicinal components and high research value, which contribute significantly to its economic importance. Similar to other ginseng varieties, *P. vietnamensis* is abundant in ginsenosides, particularly in protopanaxadiol-, protopanaxatriol-, and ocotillol-type ginsenosides. It can be further classified into three main subspecies, including *P. vietnamensis* Ha et Grushv. (PVV), *P. vietnamensis* var. *fuscidiscus* (PVF), and *P. vietnamensis* var. *langbianensis* (PVL).

Reported in 1985, PVV was the first officially recognized variety of *P. vietnamensis*, and it is the most prevalent ginseng variety in Vietnam[6]. In 2001, it was revealed that nucleotide differences in 18S and the *matK* gene explain the distinctive genetic pattern of *P. vietnamensis*.[7] Soon after, in 2003, PVF was identified subsequent to its discovery at the northernmost border of Vietnam and the southernmost frontier of Yunnan, China, and it was distinguished by a nucleotide difference in the *trnK* gene[8]. Finally, in 2016, a third variety (i.e., PVL) was identified based on its morphological characteristics and distinguishable DNA barcodes: it originated from Lang Bian Mountain on the Lam Vien Plateau in southern Vietnam[9]. Although these three varieties are categorized into a single species, *P. vietnamensis*, it is anticipated that the market values of these three varieties would not be identical. One reason for this is the substantial difference in the composition of ginsenosides and other saponins[10]. As a result, it is critical to develop a discrimination method specifically designed for these varieties.

[1]College of Pharmacy, Seoul National University, Seoul 08826, Republic of Korea. [2]Henan Academy of Sciences, No. 228, Chongshili Road, Zhengdong New District, Zhengzhou 450002, Henan Province, China. [3]Department of Agriculture, Forestry and Bioresources, Plant Genomics and Breeding Institute, College of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Republic of Korea. [4]Faculty of Pharmacy, University of Medicine and Pharmacy at Ho Chi Minh City, Ho Chi Minh City 70000, Vietnam. [5]Department of Global Innovative Drugs, Chung-Ang University, Seoul 06974, Republic of Korea. [6]Research Institute of Pharmaceutical Sciences, Seoul National University, Seoul 08826, Republic of Korea. [7]College of Pharmacy, Chonnam National University, Gwangju 61186, Republic of Korea. ✉email: swkwon@snu.ac.kr; sunjokim@jnu.ac.kr

Currently, PVV, PVF, and PVL, aged 4 years or older, are generally distinguished based on their morphological characteristics[9]. However, the three types of Vietnamese ginseng exhibit similarities in various plant parts, such as roots, stems, leaves, and flowers, making visual differentiation challenging when these varieties are interplanted. Additionally, it is imperative to emphasize that ginseng plants under 3 years of age are virtually inconspicuous by visual inspection, which significantly increases the probability of misclassification as an unintended subspecies. It is anticipated that unsatisfactory quality control caused by misclassification would result in significant financial losses, particularly in the case of ginseng cultivation, which requires an estimated duration of 4 to 6 years to yield the final product. Necessarily, investigations have been made to devise other methods for distinguishing different *P. vietnamensis* varieties. DNA-based classification methods have generally been used; for example, PVF and PVL were discovered based on their genetic characteristics[11–13]. Because of the complexity of DNA-based classification procedures, however, efforts have also been made to establish a rapid and cost-effective method using plant metabolites as marker compounds[10,11]. Nevertheless, there has been no research undertaken to discriminate among three different varieties of *P. vietnamensis* or to discover phytochemicals other than ginsenoside derivatives as markers. Therefore, it is necessary to devise a novel method for distinguishing PVV, PVF, and PVL while minimizing the damage to individual ginseng plants to preserve their value.

In this study, we aimed to establish a plant metabolite-based discrimination algorithm for the three varieties of *P. vietnamensis* while ensuring that individual plants sustained minimal damage. After collecting leaf samples from 25 PVV, 25 PVF, and 12 PVL, their extracts were first analyzed by flow injection analysis coupled to mass spectrometry (FIA-MS) to elucidate discrimination markers. Notably, a peak with a *m/z* of 426 was exclusively detected in PVF. After collecting the corresponding fraction by preparative thin layer chromatography (TLC), this peak was subsequently determined to be a novel indole alkaloid glycoside, as confirmed by high-resolution MS, ^{1}H-NMR, ^{13}C-NMR, ^{1}H-^{1}H COSY, HMBC, and HSQC. To distinguish PVV from PVL, liquid chromatography–mass spectrometry (LC–MS) analysis was performed for the global phytometabolite profiling, followed by a supervised clustering analysis to identify potential discriminative markers. Notably, lysophospholipids and hydroxy fatty acids (HFAs) were significant contributors to the classification of PVV and PVL. Correspondingly, six lysophospholipids and four HFAs were selected as potential markers, and their discriminative performance was validated via cross-validation using a machine learning algorithm.

Results

Morphological characteristics have little power in the discrimination of three varieties of *Panax vietnamensis*

The sequential information of the nucleotides of the 18S ribosomal RNA gene and *matK* gene was used to confirm the collected PVV, PVF, and PVL, according to the findings of previous studies by Komatsu et al., Zhu et al., and Duy et al.[7–9] As shown in Fig. 1, 25 PVF leaf samples were collected in Yunnan Province (China) and Lai Chau Province (Vietnam). Twenty-five PVV leaf samples were collected in Quang Nam Province and Kon Tum Province, Vietnam. Twelve PVL leaf samples were collected in Kon Tum Province and Lam Dong Province, Vietnam.

At first, we sought to confirm the accuracy of morphological assessment in differentiating PVV, PVF, and PVL. According to the reference and an expert engaged in *P. vietnamensis* cultivation in Vietnam, several morphological distinctions within the leaf were generally employed for the discrimination without damaging individual ginsengs[9]. Here, four fresh leaf samples from each group were randomly selected for the assessment (Fig. 2a). First, the leaves are known to taper more sharply toward the leaf base in PVF compared to PVV and PVL. When the petiole angles were calculated as a red line in Fig. 2a, however, there were no significant differences and no observable tendency (Fig. 2b). Secondly, PVV is considered to have a wider leaf shape, with a length-to-width ratio generally less than 2.5. In contrast, PVF features longer leaves with a length-to-width ratio that typically surpasses 3.0. Likewise, it failed to serve as an accurate indicator of the discrimination (Fig. 2c). Finally, it is generally recognized that the leaf apex of all three varieties is acuminate in shape, with PVF exhibiting a somewhat longer apex length than PVL. Our inspection revealed no statistically significant differences among the varieties when we calculated the apex length-to-leaf length ratio (Fig. 2d). Although some other distinct morphological aspects have been reported, relying solely on empirical observations of leaf shape requires extensive experience; otherwise, it may lead to decreased classification accuracy. Therefore, a more precise method than visual observation is required for classification.

Panax vietnamensis, but not the three varieties, is accurately classified by Pvdm12

The specially designed gene primer Pvdm12 was employed and converted into a Kompetitive Allele-Specific PCR (KASP) marker in this study. Three distinct varieties of Vietnamese ginseng were labeled and tested against various global ginseng types. DNA extraction, PCR amplification, and final fluorescence measurement were performed. Fluorescein amidite (FAM) relative fluorescence values were plotted on the *x*-axis, and Hexachloro-fluorescein (HEX) relative fluorescence values were plotted on the *y*-axis, resulting in Fig. 3. PVV, PVF, and PVL, as varieties of *P. vietnamensis*, successfully clustered within the blue circle, whereas the remaining other ginseng species, lacking specific marker binding, formed a distinct group within the green circle. This experiment validates the classification of PVV, PVF, and PVL as *P. vietnamensis*, but it is not applicable for distinguishing between the varieties.

Partial least squares-discriminant analysis (PLS-DA) on MS data and TLC visualization reveal a PVF-specific compound

Spectral features ranging from *m/z* 200 to 1700 were acquired from three varieties via FIA-MS analysis. Using an integer number of *m/z* as a feature, PLS-DA was performed by MetaboAnalyst 6.0. As shown in Fig. 4a, PVV, PVF, and PVL were clearly discriminated by PLS-DA. With respective R^2 and Q^2 values of 1.00 and 0.88, the

Fig. 1. Collection sites of *Panax vietnamensis*: four different provinces in Vietnam and one province in China.

PLS-DA model demonstrated a high degree of predictability and effectively accounts for most of the variability observed in the data. The variable importance projection (VIP) score of each feature was ranked from highest to lowest in order to determine the most contributing features for the development of the model. Notably, it was revealed that m/z 426 had an outstanding contribution to the model prediction (Fig. 4b). This feature, m/z 426, presented exclusively in PVF but was absent in both PVV and PVL (Fig. 4c).

In the meantime, the methanol extracts of each variety were inspected under a UV lamp (254 nm) after development on the TLC plate. A clear band with an R_f of approximately 0.67 was observed to exist exclusively in PVF samples (Fig. 4d, Supplementary Fig. S1). The region was then isolated using preparative TLC and analyzed by high-resolution MS to verify its molecular weight. Remarkably, the compound showed a mass-to-charge ratio of 426.14, which was the identical value to the feature that was the most outstanding contributor to the PLS-DA prediction. Based on MS spectral inspection, it was ascertained that the exact mass of the compound is 381.14, with a formate adduct form at m/z 426.14 (Supplementary Fig. S2). ^{1}H-NMR and ^{13}C-NMR combined with ^{1}H-^{1}H COSY, HMBC, and HSQC identified the compound as an indole alkaloid N-glycoside (Fig. 4e, Supplementary Figs. S3–S7, and Supplementary Table S1). Notably, this indole alkaloid N-glycoside was elucidated as a novel compound of *P. vietnamensis* in a recent study in which it was able to differentiate *P. vietnamensis* from *P. ginseng*, *P. notoginseng*, *P. japonicus*, and *P. quinquefolius* and was named panaxindole by the authors[14]. However, that study insisted that the compound was isolated from PVV, whereas our spectral data and TLC analysis indicated the compound was exclusively present in PVF, leading us to believe that PVV was misclassified. In summary of this section, this indole alkaloid N-glycoside, panaxindole, can be a PVF marker for differentiation.
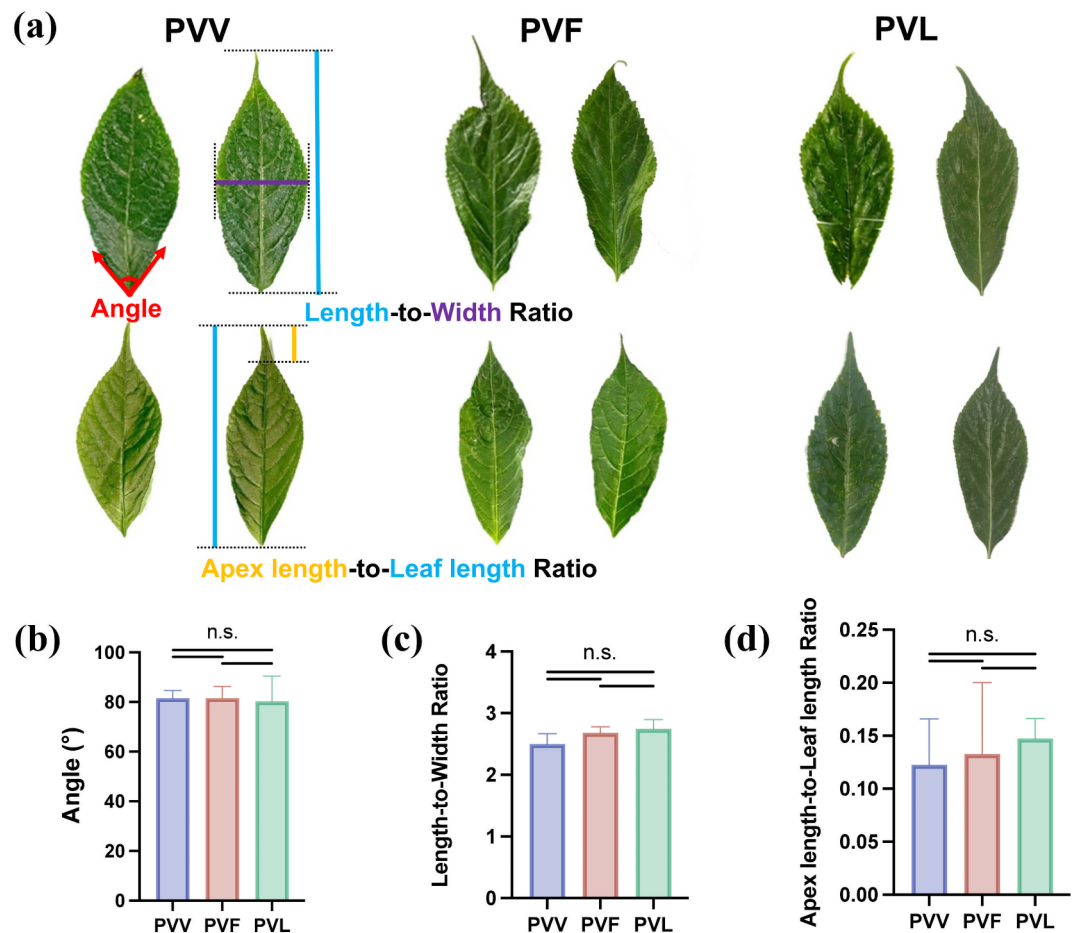
Fig. 2. The external morphology of three groups of *Panax vietnamensis* leaves. (**a**) Four leaves for each group: (**b**) leaf base angle, (**c**) leaf length-to-width ratio, and (**d**) apex length-to-leaf length ratio. The data are expressed as the mean ± standard deviation.

Lysophospholipids and HFAs are potential markers for differentiating between PVV and PVL

The absence of panaxindole in *P. vietnamensis* indicates that the sample is either PVV or PVL. To identify distinguishing markers between PVV and PVL, an LC−Q/ToF analysis was conducted, which was then followed by the identification of chromatographic peaks using two separate in-house libraries containing phytometabolites and primary metabolites (e.g., amino acids), in addition to using the public libraries. After the removal of duplicated compounds, 130 metabolites were annotated from PVV and PVL. The orthogonal PLS-DA (OPLS-DA) analysis using these annotated metabolites clearly differentiated PVV and PVL with only a slight risk of overfitting (Fig. 5a). Notably, multiple lysophospholipids, all of which were lysophosphatidylethanolamines (LPEs) or lysophosphatidylcholines (LPCs), and HFAs (FA + nO) were included in 15 of the most contributing predictive components (Fig. 5b). Most of the annotated lysophospholipids were more enriched in PVL, whereas HFAs were quantified as significantly higher in PVV (Fig. 5c). Although some were statistically non-significant, all LPCs and LPEs showed a consistent trend of PVL-enriched (Fig. 5d). Likewise, HFAs also exhibited a consistent trend of PVV enrichment (Fig. 5e).

Therefore, we aimed to examine the predictability of these lipid classes in PVV and PVL differentiation. A model was generated using the quantitative values of LPCs, LPEs, and HFAs that were included in the 15 most contributing predicted components, namely FA 18:2 + 2O, FA 18:3 + 1O, FA 18:3 + 2O, FA 18:4 + 2O, LPC 16:0, LPC 18:2, LPC 18:3, LPE 16:0, LPE 18:2, and LPE 18:3. Employing the set of these variables, the model differentiated PVV and PVL with 100% accuracy (95% confidence interval: 73.5–100%), as shown in the confusion matrix (Fig. 5f). Receiver operating characteristic (ROC) analysis using the constructed model also showed perfect classification with the area under the curve = 1.000 (95% confidence interval: 1.000–1.000) (Fig. 5g). In light of a recent study's finding that a set of ginsenosides could distinguish between PVV and PVF[10], we investigated the predictability of ginsenosides in the differentiation of PVV and PVL. Based on the current metabolomic analysis, 14 ginsenosides and their derivatives were identified. In line with previous studies, their expression levels were found to differ significantly among the three varieties (Supplementary Fig. S8). A model was trained and tested following the identical procedure using nine significant ginsenosides detected in the current study, each with a -log(false discovery rate) greater than 2.00 and a fold change exceeding 2.00. We were able to observe that the model was inferior to the former model, although it showed decent predictability
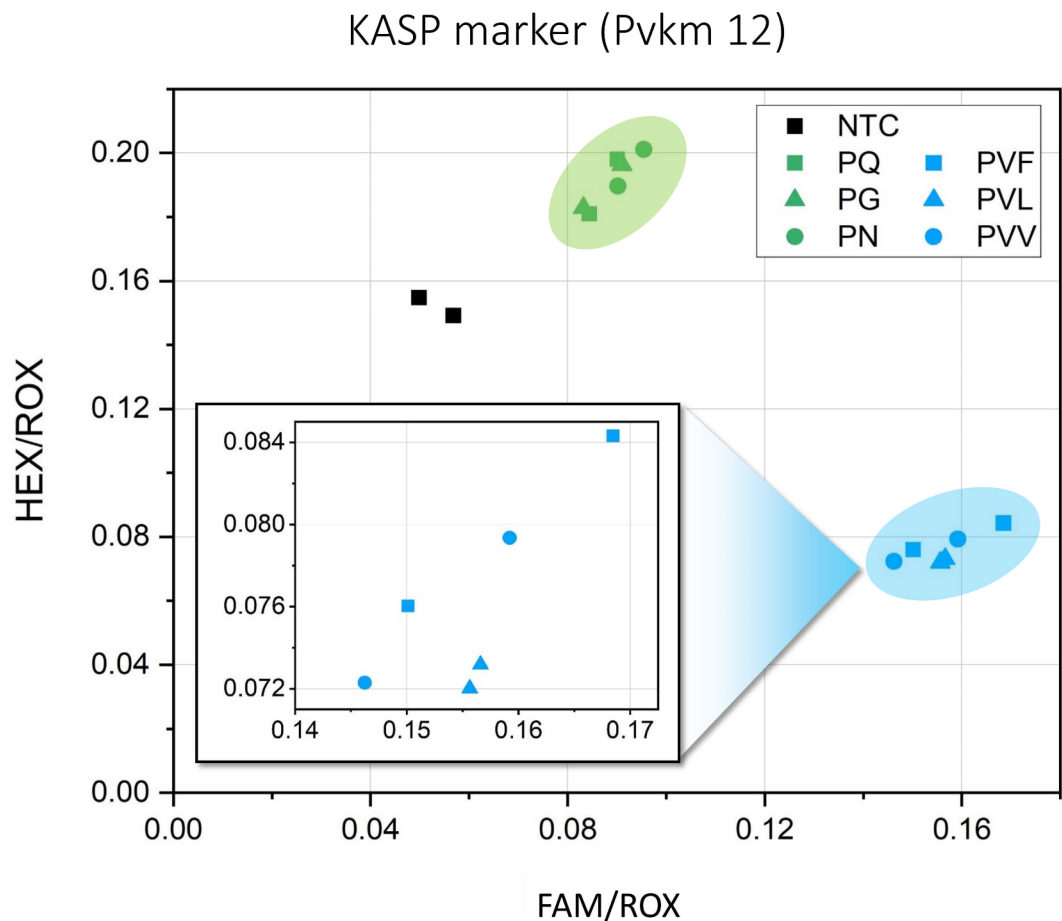
Fig. 3. Gene KASP marker results. NTC refers to the blank control, and PG, PN, and PQ denote *Panax ginseng*, *P. notoginseng*, and *P. quinquefolius*, respectively. PVV, PVF, and PVL refer to three distinct varieties of *P. vietnamensis*.

(accuracy = 83.3%; 95% confidence interval: 51.6–97.9%) (Supplementary Fig. S9a). Similarly, ROC analysis demonstrated a respectable level of predictive accuracy, albeit one that was below that of the former model (Supplementary Fig. S9b).

Discussion

This is the first study to establish and present a discrimination method for PVV, PVF, and PVL, the three varieties of *P. vietnamensis*. Additionally, it is the first investigation of *P. vietnamensis* varieties that utilizes leaf samples for discrimination marker discovery. Given the morphological ambiguity observed among the varieties and the potential for unintentional misclassification during the cultivation process, it is imperative to develop a reliable and precise discrimination method that is also straightforward to implement. Furthermore, while the official market values of the three varieties remain undisclosed, the potential for adulteration among these varieties exists in the future. Here, we established a sequential discrimination algorithm by elucidating variety-specific phytochemicals as well as lipid enrichment characteristics, all while minimizing product damage.

As emphasized above in an earlier section, differentiation while preserving product integrity is crucial for the ginseng industry, where cultivation typically takes between 4 and 6 years. Considering that the TLC and MS analyses require less than 50 μL of sample extract (50 mg mL^{-1}), it is estimated that approximately 2.5 mg of dried leaf material, which corresponds to less than one-tenth of a single ginseng leaf, is sufficient for analysis. In this regard, leaf sampling has the advantage of providing a sufficient amount of sample while minimizing the damage.

Already, a few studies have identified leaf-metabolite markers for differentiating various *Panax* species[15,16]. Notably, these studies all selected ginsenosides as potential discrimination markers. In the current study, we observed that the enrichments of four ginsenosides, compound K, ginsenosides F1, Rg1, and Rh3, were quite significantly different when we compared PVV versus PVL, as shown in Supplementary Fig. S8. However, the fold changes (FCs) of these compounds were less pronounced compared to the markers identified in current study. The individual log$_2$FC values for compound K, ginsenosides F1, and Rh3 were below 2.00, while ginsenoside Rg1 showed a value of 2.83. In contrast, the log$_2$FC values for 'FA 18:3 + 2O_2', 'LPC 18:2', 'LPE 18:2', and 'LPE 16:0' were significantly higher, at 3.26, 3.18, 3.16, and 3.11, respectively. Consequently, the predictability of the
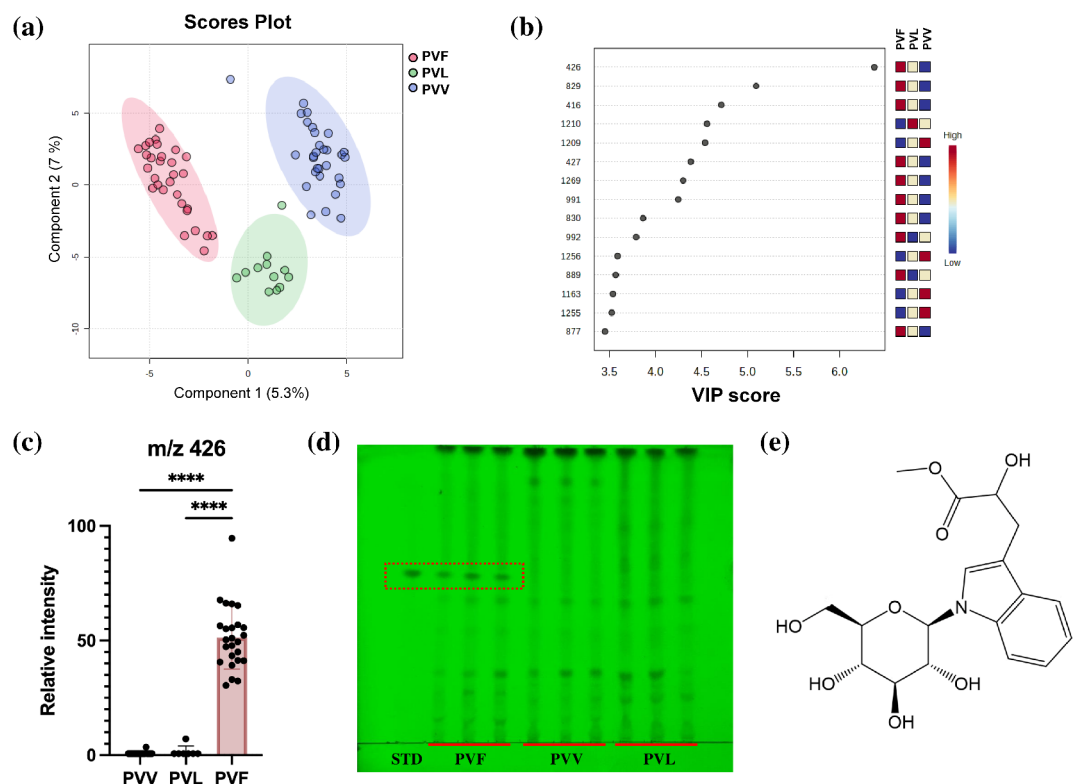
Fig. 4. Discovery of PVF-specific phytometabolite marker. (**a**) Partial least squares-discriminant analysis (PLS-DA) between PVV, PVF, and PVL; and (**b**) the VIP scores of the first component of the optimal PLS-DA model. The integer numbers in the row indicate the mass-to-charge ratio (*m/z*) of the features. (**c**) Relative intensity of *m/z* 426 between PVV, PVF, and PVL. The data are expressed as the mean ± standard deviation. **** indicates that the adjusted *P*-value is < 0.0001. (**d**) TLC image of ginseng leaves methanol extract under a UV 254 nm wavelength. Three random samples from three groups of *Panax vietnamensis* are developed. STD denotes the later isolated indole alkaloid glycoside. (**e**) Molecular structure of isolated indole alkaloid glycoside, panaxindole.

model trained with our markers (Fig. 5g) was notably superior to the model trained with the ginsenosides (Supplementary Fig. S9b). Regarding the role of ginsenosides in PVF differentiation, several ginsenosides exhibited significant differences compared to PVV or PVL. Nevertheless, none of the identified ginsenosides was unique to a specific variety. In contrast, panaxindole was exclusively present in PVF, demonstrating clear specificity.

Due to the relatively recent discovery and characterization of new varieties of *P. vietnamensis*, the chemical characteristics of individual varieties are less well understood. Nonetheless, one study attempted to establish a discrimination method among *P. vietnamensis* species[10]. The authors identified the unique expression patterns of ginsenosides in PVV and PVF. They applied chemometrics and machine learning modeling to assess the performance of these distinguishing factors. Our research strengthens the differentiation method of *P. vietnamensis* through the inclusion of an additional variety type, PVL, as well as the identification of a novel PVF-specific marker. As stated in the results section, this novel PVF-specific marker, panaxindole, was elucidated in a recent study but was initially misidentified as a PVV-derived compound[14]. This undoubtedly highlights the extreme difficulty in distinguishing among the varieties of *P. vietnamensis*, necessitating the implementation of a qualified differentiation method.

Although our study has not elucidated the biological function of panaxindole, several studies have investigated the function of indole alkaloid glycosides from various sources. As an example, it has been demonstrated that several indole alkaloid *N*-glycosides isolated from *Ginkgo biloba* possess anti-inflammatory and anti-aging properties[17,18]. Other indole alkaloid glycosides, including those identified in *Gardneria nutans*, *Isatis indigotica*, and *I. tinctoria*, exhibited consistent anti-inflammatory activity[19–21]. This suggests that PVF may possess an enhanced anti-inflammatory function, necessitating additional research in this area.

Intriguingly, the HFAs identified in this study are mostly mono- or polyhydroxylated forms of C18 fatty acids, which have been reported to correlate with reduced inflammatory processes. For instance, treatment of α-linolenic acid (FA 18:3) to THP-1 macrophages resulted in a decrease in the excretion of inflammatory cytokines such as IL-6, TNF-α, and IL-1β[22]. This effect was accompanied by increased levels of two hydroxy derivatives of FA 18:3 (FA 18:3 + 1O), namely 9-hydroxy-octadecatrienoic acid and 13-hydroxy-octadecatrienoic acid. Additionally, 13-hydroxy-octadecatrienoic acid has been shown to exhibit anti-inflammatory properties when applied to lipopolysaccharide-treated macrophages[23]. By inactivating NLRP3 inflammasome complex via
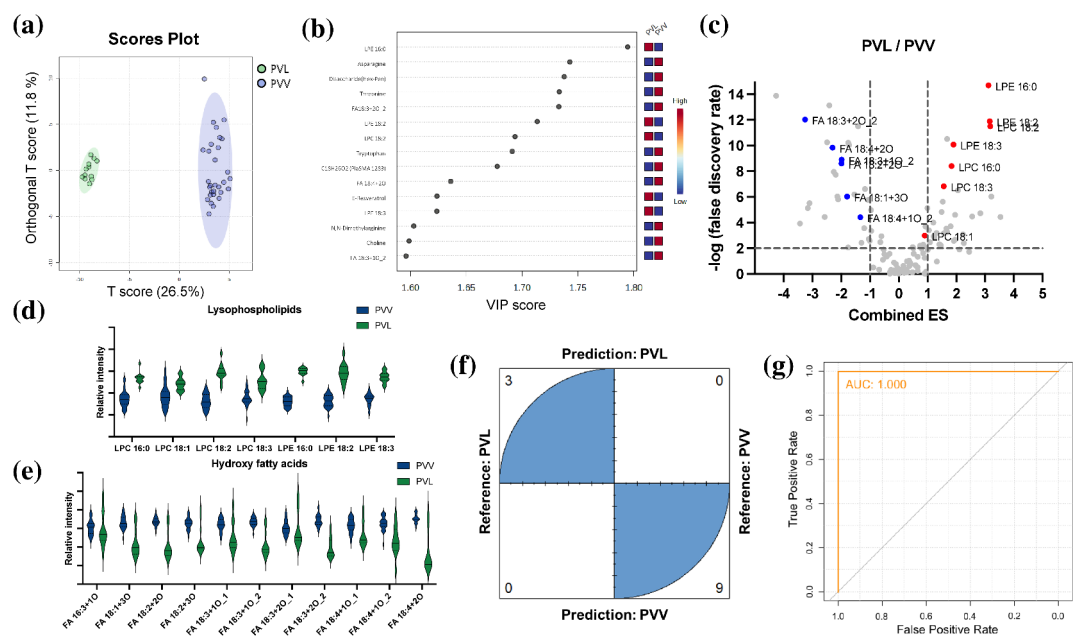
Fig. 5. The metabolomics-based discovery of PVV and PVL discriminating markers, followed by a model evaluation that is generated using the selected markers. (**a**) Orthogonal partial least squares-discriminant analysis (OPLS-DA) between PVV and PVL; and (**b**) the VIP scores of component 1 (T-score) of the OPLS-DA model. (**c**) Volcano plot with PVL over PVV. Lysophospholipids and hydroxy fatty acids (HFAs) are highlighted. (**d**) Difference in the expression of lysophospholipids. (**e**) Difference in the expression of HFAs. (**f**) Confusion matrix of machine learning-based discrimination using six lysophospholipids and four HFAs that were selected based on their VIP score and (**g**) receiver operating characteristic (ROC) curve discriminating PVL over PVV, with an area under the curve (AUC) value.

the PPAR-γ pathway, it inhibits the expression of iNOS and TNF-α while upregulating IL-10 expression. HFAs can also conjugate with other fatty acids to form cross-esterified compounds known as FAHFAs (fatty acid esters of hydroxy fatty acids), and their potential health benefits have been actively studied. These include reverting metabolic dysregulations by inhibiting hepatic glucose production, enhancing insulin-stimulated glucose uptake, and reducing adipose tissue inflammation[24,25]. Moreover, FAHFAs have been implicated in calcium-dependent signaling pathways during myocardial ischemia and related pathological events[26]. Consistent with these findings, plasma levels of FAHFAs were found to be significantly lower in obese children and adults with coronary artery disease[27]. Given these observations, further studies are warranted to investigate the biological functions of *P. vietnamensis* in light of these potential benefits, particularly for PVV, which exhibits high levels of HFAs.

Numerous studies have applied mass spectrometry to identify root-based phytochemical markers for the discrimination of ginseng species and ages. Consistently, ginsenosides or related triterpenoid saponins were found to be the most predictive markers when it comes to the comparison of *P. ginseng*, *P. quinquefolius*, *P. notoginseng*, *P. japonicus*, and *P. japonicus* var. *major*[28–33]. In some cases, primary metabolites and lipids were incorporated to enhance the predictability[31–33]. Notably, several HFAs have been described as effective discriminating markers. For example, (15Z)-9,12,13-trihydroxy-15-octadecenoic acid (FA 18:1 + 3O) demonstrated noteworthy efficacy in the differentiation of *P. quinquefolius* cultivated in various provinces across China[32]. *P. ginseng* is reported to be enriched in 12-hydroxyoctadec-9-enoic acid (FA 18:1 + 1O) and hydroxyhexadecanoic acid (FA 16:0 + 1O), and the potential role of these HFAs as discriminant markers was discussed[33]. It is expected that the variation in HFA expression levels could serve as a discrimination feature among distinct *Panax* species. HFA was one of the most significant chemical classes capable of differentiating PVV from PVL, as shown in Fig. 5e. Conversely, lysophospholipid has rarely been described as a discriminant marker among the *Panax* genus in any other publications. However, in the case of other agricultural products, lysophospholipid was able to function as a robust and accurate discrimination marker for finding geographical origins or detecting adulterated admixtures[34,35].

In summary, we established a benchmark of an intact sample distinguishing method for the three varieties of *P. vietnamensis* by profiling their phytochemical characteristics with multiple analytical platforms. It is our conviction that our research would contribute to the development of the Vietnamese ginseng industry through the provision of a framework for precise classification.

Methods
Chemicals and reagents
The derived cleaved amplified polymorphic sequence (dCAPS) marker Pvdm12 for *P. vietnamensis* (forward primer: TGCGCGAATCTCAGCAATCACTAG, reverse primer: AAATTCAATGAGGATTTGGTTCAT) was

purchased from GeneAll Biotechnology Co., Ltd. (Seoul, Korea). LC−MS grade solvents and formic acid were purchased from Merck (Darmstadt, Germany) and Fisher Scientific (Waltham, MA, USA). Exgene™ Plant SV was purchased from GeneAll Biotechnology Co., Ltd. TLC silica gel 60 F_{254} glass-backed plates (1.05715.0001) for TLC analysis were bought from Merck.

Plant material and sample extraction
The collection of plant materials and all investigations were conducted in accordance with relevant institutional, national, and international guidelines and legislation. Voucher specimens of three *P. vietnamensis* varieties have been deposited in the Global Biodiversity Information Facility and are classified as *Panax* species. The 3 to 4-year-old ginseng leaf samples were collected and identified by an expert (T.H. Van Le) from the farms that specialize in the cultivation of *P. vietnamensis*. The collected leaves were lyophilized and weighed prior to subsequent procedures. For DNA analysis, 25 mg of dried leaves was ground into powder using a TissueLyser II for 15 min and extracted by Exgene™ Plant SV. *P. vietnamensis*-specific KASP marker was converted by the dCAPS, Pvdm12. For the TLC and MS analysis, the ground samples were subjected to methanol extraction at 60 °C for 1 h at a concentration of 50 mg mL^{-1}.

Gene analysis by using KAPS markers
The purified DNA was quantified using a Nanodrop (Fisher Scientific) and diluted to a concentration of 10 μg μL^{-1}. Here, we used a published marker, Pvdm12, which can identify *P. vietnamensis* by leaves, and converted it to a KASP marker to obtain the fluorescence by using a LightCycler 96 from Roche Diagnostics (Rotkreuz, Switzerland)[36]. KASP enables bi-allelic discrimination through the competitive binding of two allele-specific forward primers and a common reverse primer to amplify the target DNA sequence through PCR. The resulting products, corresponding to the sequence of the target DNA, are fluorescently labeled. In the case of a homozygous sample, PCR products will be labeled with either HEX or FAM fluorescent dyes. However, for a heterozygous sample, products labeled with both HEX and FAM will be present. Unlabeled fluorescent dyes remain quenched. KASP can be conducted in microtiter plates and can be detected using a fluorescence microplate reader LightCycler 96[37,38].

Flow injection analysis coupled to mass spectrometry
Methanol extracts of PVV, PVF, and PVL were diluted 100-fold with 70% acetonitrile in water with 0.1% formic acid for LC−MS analysis. The diluted samples were introduced into the Agilent QQQ 6460 mass spectrometer via direct injection using a carrier solution composed of 70% acetonitrile in water containing 0.1% formic acid. The intensity of precursor ions was measured using the negative ion scan mode, covering a range of *m/z* 200 to 1700. The capillary voltage was set to 3000 V, and the gas flow rate was maintained at 5 L min^{-1}. Ionization was performed at a fragmentor voltage of 100 V.

Thin layer chromatography
Five microliters of methanol extract was deposited on the silica gel plate, l cm from the bottom. The analytes were subsequently separated using a mobile phase of chloroform: methanol: water (65:35:10, v/v/v). After it was air-dried, the developed TLC plate was visualized under the 254 nm UV source. A key compound of PVF was then collected using preparative TLC under identical conditions.

PVF key compound confirmation
The isolated compound was dissolved in deuterated methanol, followed by a filtration process. For the structural confirmation, the isolated compound was subjected to high-resolution MS and full spectrum examination, including^{1}H-NMR, ^{13}C-NMR, ^{1}H-^{1}H COSY, HMBC, and HSQC, using an 800 MHz NMR from Bruker Corporation (Billerica, MA, USA). The NMR results were analyzed by MestReNova from Mestrelab Research S.L. (Santiago de Compostela, Spain).

Global profiling of metabolites
A 1290 UPLC was coupled with a 6530 hybrid Q-TOF system from Agilent (Santa Clara, CA, USA) for the comprehensive profiling of metabolites of PVV and PVL. Metabolites were profiled by comparing retention time, *m/z*, and MS/MS spectrum to the two different LC−MS in-house libraries: a primary metabolite library constructed using MSMLS from Sigma Aldrich (Saint Louis, MO, USA) with amide-bonded hydrophilic interaction chromatographic separation[39], and a phytometabolite library constructed using PHYTOMLS (Sigma Aldrich) with reverse chromatographic separation[40]. MSDIAL version 4.90 was used for data pre-processing and peak annotation. The details of the parameters are presented in Supplementary Table S2.

Statistical analyses
For the pre-processing of MS data, features that exhibited greater than 50% overall missing values were excluded from the analysis, and the remaining missing values were approximated to one-fifth of the minimum positive value of the variable. The data were subsequently structured using total intensity normalization, logarithmic transformation, and Pareto scaling in order to prepare them for multivariate analysis. Multivariate analyses, including PLS-DA and OPLS-DA, were analyzed using MetaboAnalyst 6.0. Highly contributing predictive components were selected based on VIP scores. One-way analysis of variance and Tukey's multiple comparison test were conducted using GraphPad Prism 10 from Dotmatics (Boston, MA, USA).

Differentiation between PVV and PVL by machine learning algorithm

Marker candidates for PVV and PVL differentiation were selected among the highly contributing predictive components from OPLS-DA. The dataset containing quantitative values of marker candidates was divided into two sets: 70% for a training set and 30% for a test set. The classification was carried out with reliance on the Random Forest algorithm available in the caret package version 6.0–94 in R. On the training set, a repeated 10-fold cross-validation was performed to evaluate the performance of the model. The model's predictability was illustrated through the confusion matrix and ROC curves.

Data availability

The metabolomics datasets supporting the findings of this study are available from the corresponding author upon reasonable request.

References

1. Konoshima, T. et al. Cancer chemopreventive activity of majonoside-R2 from Vietnamese ginseng, Panax Vietnamensis. *Cancer Lett.* **147**, 11–16. https://doi.org/10.1016/S0304-3835(99)00257-8 (1999).
2. Chae, S. et al. Effect of compound K, a metabolite of Ginseng Saponin, combined with γ-Ray Radiation in Human Lung Cancer cells in Vitro and in vivo. *J. Agric. Food Chem.* **57**, 5777–5782. https://doi.org/10.1021/jf900331g (2009).
3. Kang, S. & Min, H. Ginseng, the 'Immunity Boost': the effects of Panax ginseng on Immune System. *J. Ginseng Res.* **36**, 354–368. https://doi.org/10.5142/jgr.2012.36.4.354 (2012).
4. Sievenpiper, J. L., Arnason, J. T., Leiter, L. A., 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.* **23**, 248–258. https://doi.org/10.1080/07315724.2004.10719368 (2004).
5. Park, S. Y. et al. Systems-level mechanisms of action of Panax ginseng: a network pharmacological approach. *J. Ginseng res.* **42**, 98–106. https://doi.org/10.1016/j.jgr.2017.09.001 (2018).
6. Dung, H. T. & Grushvitski, I. A new species of the genus Panax (Araliaceae) from Vietnam. *Botanicheskii Zhurnal* **70** (1985).
7. Komatsu, K. et al. Phylogenetic analysis based on 18S rRNA gene and matK gene sequences of Panax vietnamensis and five related species. *Planta Med.* **67**, 461–465. https://doi.org/10.1055/s-2001-15821 (2001).
8. Zhu, S., Fushimi, H., Cai, S., Chen, H. & Komatsu, K. A new variety of the genus Panax from southern Yunnan, China and its nucleotide sequences of 18S ribosomal RNA gene and matK gene. *J. Jap Bot.* **78**, 86–94 (2003).
9. Duy, N. V., Trieu, L. N., Chinh, N. & Tran, V. T. A new variety of Panax (Araliaceae) from Lam Vien Plateau, Vietnam and its molecular evidence. *Phytotaxa* **277**, 047–058. https://doi.org/10.11646/phytotaxa.277.1.4 (2016).
10. Nguyen, H. T. et al. Untargeted Metabolomics Approach for the differentiation between Panax vietnamensis var. Vietnamensis and Panax vietnamensis var. Fuscidiscus. *Metabolites* **13** (2023).
11. Zhang, G. H. et al. Transcriptome analysis of Panax vietnamensis var. Fuscidicus discovers putative ocotillol-type ginsenosides biosynthesis genes and genetic markers. *BMC Genom.* **16**, 159. https://doi.org/10.1186/s12864-015-1332-8 (2015).
12. Linh, N. N. et al. Species discrimination of novel chloroplast DNA barcodes and their application for identification of Panax (Aralioideae, Araliaceae). *PhytoKeys* **188**, 1–18; (2022). https://doi.org/10.3897/phytokeys.188.75937
13. Phuong Trang, N. T., Mai, H., Yury, N. T., Galina, D. & N, Z. & rbcL and rpoL gene sequences of Panax vietnamensis var. Fuscidiscus and Panax Vietnamensis, the background for identification and comparison. *Acad. J. Biology* **39**, 80–85. https://doi.org/10.15625/0866-7160/v39n1.7870 (2016).
14. Vu, V. T. et al. Panaxindole, a novel indole alkaloid N-glucoside from the leaves of Panax vietnamensis ha et grushv. (Vietnamese ginseng). *J. Nat. Med.* **77**, 972–977. https://doi.org/10.1007/s11418-023-01728-4 (2023).
15. Mao, Q. et al. Discrimination of leaves of Panax ginseng and P. Quinquefolius by Ultra high performance liquid chromatography quadrupole/time-of-flight mass spectrometry based metabolomics approach. *J. Pharm. Biomed. Anal.* **97**, 129–140 (2014).
16. Wang, X. et al. Pseudotargeted metabolomics approach enabling the classification-induced ginsenoside characterization and differentiation of ginseng and its compound formulation products. *J. Agric. Food Chem.* **71**, 1735–1747 (2023).
17. Cheng, J. T. et al. Isolation of two rare N-glycosides from Ginkgo biloba and their anti-inflammatory activities. *Sci. Rep.* **10**. https://doi.org/10.1038/s41598-020-62884-1 (2020).
18. Choi, Y. J. et al. Antiskin aging effects of indole alkaloid n-glycoside from ginkgo fruit (Ginkgo biloba fruit) on TNF-α-Exposed human dermal fibroblasts. *J. Agric. Food Chem.* **70**, 13651–13660 (2022).
19. Si, Y. Y. et al. Neuroinflammatory inhibitors from Gardneria Nutans Siebold & Zuccarini. *RSC Adv.* **11**, 27085–27091 (2021).
20. Yang, L. et al. Indole alkaloids from the roots of Isatis Indigotica and their inhibitory effects on nitric oxide production. *Fitoterapia* **95**, 175–181 (2014).
21. Zhang, D. et al. Indole alkaloid glycosides from Isatis tinctoria roots. *Nat. Prod. Res.* **35**, 244–250 (2021).
22. Pauls, S. D. et al. Anti-inflammatory effects of α-linolenic acid in M1-like macrophages are associated with enhanced production of oxylipins from α-linolenic and linoleic acid. *J. Nutr. Biochem.* **57**, 121–129. https://doi.org/10.1016/j.jnutbio.2018.03.020 (2018).
23. Kumar, N. et al. 15-Lipoxygenase metabolites of α-linolenic acid, [13-(S)-HPOTrE and 13-(S)-HOTrE], mediate anti-inflammatory effects by inactivating NLRP3 inflammasome. *Sci. Rep.* **6**, 31649. https://doi.org/10.1038/srep31649 (2016).
24. Yore, M. M. et al. Discovery of a class of endogenous mammalian lipids with anti-diabetic and anti-inflammatory effects. *Cell* **159**, 318–332. https://doi.org/10.1016/j.cell.2014.09.035 (2014).
25. Aryal, P. et al. Distinct biological activities of isomers from several families of branched fatty acid esters of hydroxy fatty acids (FAHFAs). *J. Lipid Res.* **62**, 100108. https://doi.org/10.1016/j.jlr.2021.100108 (2021).
26. Huang, J. M., Xian, H. & Bacaner, M. Long-chain fatty acids activate calcium channels in ventricular myocytes. *Proc. Natl. Acad. Sci.* **89**, 6452–6456. https://doi.org/10.1073/pnas.89.14.6452 (1992).
27. Ong, S. M. et al. Plasma fatty acid esters of hydroxy fatty acids and surrogate fatty acid esters of hydroxy fatty acids hydrolysis activity in children with or without obesity and in adults with or without coronary artery disease. *Diabetes Obes. Metab.* **26**, 3429–3438. https://doi.org/10.1111/dom.15686 (2024).
28. Chan, T. W. D. et al. Differentiation and authentication of Panax ginseng, Panax quinquefolius, and Ginseng products by using HPLC/MS. *Anal. Chem.* **72**, 1281–1287. https://doi.org/10.1021/ac990819z (2000).
29. Yang, W. et al. Identification and differentiation of Panax ginseng, Panax Quinquefolium, and Panax notoginseng by monitoring multiple diagnostic chemical markers. *Acta Pharm. Sin B* **6**, 568–575 (2016).
30. Du, Z., Li, J., Zhang, X., Pei, J. & Huang, L. An Integrated LC-MS-Based strategy for the Quality Assessment and discrimination of three Panax species. *Molecules* **23** (2018).

31. Yoon, D., Shin, W. C., Oh, S. M. & Choi, B. R. Young Lee, D. Integration of multiplatform metabolomics and multivariate analysis for geographical origin discrimination of Panax ginseng. *Food Res. Int.* **159**, 111610. https://doi.org/10.1016/j.foodres.2022.111610 (2022).
32. Pang, S. et al. Discrimination for geographical origin of Panax quinquefolius L. using UPLC Q-Orbitrap MS-based metabolomics approach. *Food Sci. Nutr.* **11**, 4843–4852. https://doi.org/10.1002/fsn3.3461 (2023).
33. Ji, R. et al. Untargeted metabolomic analysis and Chemometrics to identify potential marker compounds for the Chemical differentiation of Panax ginseng, P. Quinquefolius, P. notoginseng, P. japonicus, and P. japonicus var. Major. *Molecules* **28** (2023).
34. Long, N. P., Lim, D. K., Mo, C., Kim, G. & Kwon, S. W. Development and assessment of a lysophospholipid-based deep learning model to discriminate geographical origins of white rice. *Sci. Rep.* **7**, 8552 (2017).
35. Lim, D. K. et al. Combination of mass spectrometry-based targeted lipidomics and supervised machine learning algorithms in detecting adulterated admixtures of white rice. *Food Res. Int.* **100**, 814–821 (2017).
36. Nguyen, V. B. et al. Comprehensive comparative analysis of chloroplast genomes from seven Panax species and development of an authentication system based on species-unique single nucleotide polymorphism markers. *J. Ginseng Res.* **44**, 135–144. https://doi.org/10.1016/j.jgr.2018.06.003 (2020).
37. Patterson, E. L., Fleming, M. B., Kessler, K. C., Nissen, S. J. & Gaines, T. A. A KASP genotyping method to identify Northern Watermilfoil, Eurasian Watermilfoil, and their interspecific hybrids. *Front. Plant. Sci.* **8**. https://doi.org/10.3389/fpls.2017.00752 (2017).
38. Smith, S. M. & Maughan, P. J. in *Plant Genotyping: Methods and Protocols* (ed Jacqueline Batley) 243–256. (Springer New York, 2015).
39. Blaženović, I. et al. Structure annotation of all Mass Spectra in Untargeted Metabolomics. *Anal. Chem.* **91**, 2155–2162. https://doi.org/10.1021/acs.analchem.8b04698 (2019).
40. Tsugawa, H. et al. A cheminformatics approach to characterize metabolomes in stable-isotope-labeled organisms. *Nat. Methods.* **16**, 295–298. https://doi.org/10.1038/s41592-019-0358-2 (2019).

Acknowledgements

Author contributions

Cheng RR: Conceptualization, Investigation, Data Processing, Writing – original draft. Yoon YC: Software. Jung CW: Data curation. Kim TH: Methodology. Wang Q: Resources. Cho WH: Investigation. Yang TJ: Resources. Le THV: Resources. Cho CJ: Investigation. Kim JH: Investigation. Hyun GH: Visualization. Park JH: Conceptualization. Kwon SW: Supervision, Funding acquisition. Kim SJ: Conceptualization, Writing – original draft, Supervision.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-88321-9.

Correspondence and requests for materials should be addressed to S.W.K. or S.J.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.