RESEARCH ARTICLE

Phytochemical Analysis WILEY

Anticoagulant activity analysis and origin identification of *Panax notoginseng* using HPLC and ATR-FTIR spectroscopy

Zhi-Ying Cui[1] | Chun-Lu Liu[1,2] | Dan-Dan Li[1] | Yuan-Zhong Wang[2] | Fu-Rong Xu[1]

[1]College of Traditional Chinese Medicine, Yunnan University of Chinese Medicine, Kunming, Yunnan, China

[2]Medicinal Plants Research Institute, Yunnan Academy of Agricultural Sciences, Yunnan, Kunming, China

Correspondence
Yuan-Zhong Wang, Medicinal Plants Research Institute, Yunnan Academy of Agricultural Sciences, 2238, Beijing Road, Panlong District, Yunnan, Kunming 650200, PR China.
Email: boletus@126.com

Fu-Rong Xu, College of Traditional Chinese Medicine, Yunnan University of Chinese Medicine, Yunnan, Kunming, 650500, China.
Email: xfrong99@163.com

Abstract

Introduction: *Panax notoginseng* is one of the traditional precious and bulk-traded medicinal materials in China. Its anticoagulant activity is related to its saponin composition. However, the correlation between saponins and anticoagulant activities in *P. notoginseng* from different origins and identification of the origins have been rarely reported.

Objectives: We aimed to analyze the correlation of components and activities of *P. notoginseng* from different origins and develop a rapid *P. notoginseng* origin identification method.

Materials and methods: Pharmacological experiments, HPLC, and ATR-FTIR spectroscopy (variable selection) combined with chemometrics methods of *P. notoginseng* main roots from four different origins (359 individuals) in Yunnan Province were conducted.

Results: The pharmacological experiments and HPLC showed that the saponin content of *P. notoginseng* main roots was not significantly different. It was the highest in main roots from Wenshan Prefecture (9.86%). The coagulation time was prolonged to observe the strongest effect (4.99 s), and the anticoagulant activity was positively correlated with the contents of the three saponins. The content of ginsenoside Rg_1 had the greatest influence on the anticoagulant effect. The results of spectroscopy combined with chemometrics show that the variable selection method could extract a small number of variables containing valid information and improve the performance of the model. The variable importance in projection has the best ability to identify the origins of *P. notoginseng*; the accuracy of the training set and the test set was 0.975 and 0.984, respectively.

Conclusion: This method is a powerful analytical tool for the activity analysis and identification of Chinese medicinal materials from different origins.

KEYWORDS
anticoagulation activity, ATR-FTIR, HPLC, origin identification, *Panax notoginseng*

1 | INTRODUCTION

Panax notoginseng (Burk.) F.H. Chen is one of the traditional precious and bulk-traded medicinal materials in China. It is a famous Chinese traditional medicinal herb in Yunnan Province. At present, it has a medicinal history of more than 400 years, and it is often sold as herbal medicine or made into herbal preparations, occupying a unique position in the regional market.[1,2] It was usually applied in China as a hemostatic medicine to control internal and external bleeding, which could activate blood circulation throughout the whole body, and it was used to stop wound bleeding on the battlefield without causing blood stasis. In addition, due to its good vasodilatory and antihypertensive actions, it was used worldwide to treat cardiovascular disease.[3–6] It also has myocardial protective effects, particularly in improving ischemia/reperfusion-induced injury after percutaneous coronary intervention therapy.[3,7] Furthermore, with the increasing attention to herbal medicines, the cytotoxic effect of *P. notoginseng* on cancer cells and the underlying mechanisms have been studied recently.[8–11] As a medicine and food, it has the effect of dispersing blood stasis and promoting blood circulation, expanding blood vessels, eliminating swelling, regulating blood lipid levels, and exerting antithrombotic effects, and therefore it is popular among consumers.[4,12] Saponins are important secondary metabolites and form the pharmacodynamic basis of *P. notoginseng*.[13,14] Saponin components are often used as quality control indicators for *P. notoginseng* medicinal materials or preparations. Modern pharmacological studies have shown that the anticoagulation effects of *P. notoginseng* are mostly related to saponin components. For example, Ren et al.[15] evaluated the anticoagulation activity of extracts of *P. notoginseng* and different medicinal *Panax* species and analyzed the saponin compositions to verify and clarify the mechanisms underlying the anticoagulation effects. Du et al.[16] showed that saponins have strong anticoagulation activity through ultra-performance liquid chromatography (UPLC) combined with hierarchical clustering analysis and multiple linear regression analysis. In addition, through isolation, identification, and quantification of 16 ginsenosides and two sapogenins, it was shown that some of these compounds were potential natural inhibitors of coagulation factor Xa.[17]

However, in recent years, the continuous increase market in demand, the influence of continuous implant disorder in *P. notoginseng*, and the shortage of land resources in Wenshan Prefecture in Yunnan Province have led to the gradual introduction of *P. notoginseng* from the surrounding areas, such as Honghe, Qujing, Kunming, and other origins.[18,19] However, differences in environmental factors such as temperature, humidity, altitude, and soil type could lead to significant differences in the internal composition and quality of *P. notoginseng*, resulting that appear the phenomenon of shoddy exceed and mixing the false in the market. One of the factors that have the greatest impact on the quality of *P. notoginseng* is the chemical composition (including saponins), and differences in composition will lead to changes in its pharmacological effects and clinical efficacy. Therefore, the origin of *P. notoginseng* batches has become a key issue.

Chemical fingerprinting combined with chemometrics methods is currently one of the most commonly used methods for evaluating the chemical characteristics and identifying the origin of Chinese medicinal materials. High-performance liquid chromatography (HPLC) dominates the field of drug analysis with its high analysis speed and its wide range of applications,[20,21] mainly through the separation and identification of the chemical components of Chinese medicinal materials, providing assessment tools for their quality.[22–24] HPLC is widely used for similarity and difference analysis, determination of components, and identification and pharmacological analysis of Chinese herbal medicines.[25–28] Infrared (IR) spectroscopy has been widely applied in the identification of Chinese herbal materials owing to its characteristics of being non-destructive and fast, and it can be used to obtain the main structural information of the sample's chemical composition.[29,30] It was developed mainly based on compounds' structural information to analyze the overall chemical composition, combined with chemometrics to identify and predict the Chinese herbal materials and to evaluate the quality by analyzing the model effect.[31–33] In qualitative and quantitative analysis based on IR spectroscopy, it is often combined with chemometric methods to explore the linear correlation between the overlapping spectra and the chemical composition of the sample.[34] Variable selection is one of the important steps in chemometric methods, which is mainly applied to eliminate redundant and collinear information to reduce the computing tasks and model dimensions, improving model performance.[35] As the most frequently applied variable selection methods, variable importance in projection (VIP) values, interval partial least squares, competitive adaptive reweighted sampling (CARS), and Monte Carlo uninformative variable elimination have been widely applied with IR spectroscopy as powerful tools to identify the quality of Chinese medicinal materials.[36–39] However, studies using feature extraction methods for variable selection from IR spectroscopy data to identify the origin of *P. notoginseng* are rare.

Therefore, the purpose of this work was to apply HPLC and attenuated total reflectance Fourier transform IR spectroscopy (ATR-FTIR) to analyze the influence of saponin components of *P. notoginseng* from different origins on the anticoagulant activity and to identify their origins, which comprehensively evaluate the quality. The saponin contents of *P. notoginseng* from different origins were determined by HPLC, and the correlation between anticoagulation activity and saponin content was analyzed. We analyzed the potential and feasibility of using VIP and CARS variable selection methods based on ATR-FTIR data for identifying *P. notoginseng* from different origins. In addition, in order to find the best identification method, different preprocessing methods were applied to ATR-FTIR, and the most influential spectral wavelengths in the classification process were manually selected.

2 | MATERIALS AND METHODS

2.1 | Materials and reagents

A total of 359 main root specimens of *P. notoginseng* samples were collected from four different origins in Yunnan province, Honghe

(HH), Kunming (KM), Qujing (QJ), and Wenshan (WS). Detailed information of *P. notoginseng* samples is shown in Figure 1. All collected fresh samples were cleaned and dried at room temperature, and the main roots, rhizome, and fibrous root parts were separated and sealed for storage. Before the experiments, the main root sample was ground by a pulverizer (FW-100, Tianjin Huaxin Instrument Factory) and powder was filtered through 90-μm mesh sieves, dried in an oven at 50°C, and kept at constant temperature until further analysis.

The standards were purchased from Beijing Soledad Bao Technology Co. (Beijing, P.R. China), including notoginsenoside R_1 and ginsenosides Rg_1 and Rb_1 (batch number 1028C022, purity ≥ 98%). Methanol and acetonitrile of chromatography grade for HPLC were purchased from Sigma-Aldrich (Shanghai-China) and Merck & Co., Inc. (Kenilworth, NJ, USA), respectively. Analytic grade ethanol reagent was purchased from Tianjin Zhiyuan Chemical Reagent Co., Ltd. Deionized (ultra-pure) water for HPLC was prepared by using a UPTL-II-40 L system (Chengdu, China).

Phosphate-buffered saline (PBS) was purchased from Beijing Soledad Bao Technology Co. (Beijing, China). Rivaroxaban was purchased from Bayer Pharma AG (Berlin, Germany). The prothrombin time (PT), activated partial thromboplastin time (APTT), thrombin time (TT), fibrinogen (FIB) assay kit was purchased from Sysmex Corporation (Kobe, Japan).

Male rats (220–260 g) were obtained from Hunan Slake Jingda Experimental Animal Co., Ltd. (License No. SCXK [Xiang]2019-0004). The animal experiments were approved by Yunnan University of Chinese Traditional Medicine. Animals were housed at the experimental animal center of Yunnan University of Traditional Chinese Medicine in a dark air-conditioned room at a temperature of 23 ± 1°C and a relative humidity of 30–70% with free access to food and water. The animals were acclimated at least one week before any experiments.

2.2 | Pharmacological analysis

2.2.1 | HPLC determination and saponin content analysis

All samples were analyzed using an Agilent 1260 liquid chromatograph (Agilent Corporation, USA) coupled with a Zorbax Eclipse Plus C18 column (4.6 mm × 250 mm, 5 μm, Agilent Corporation, USA). Mobile phase A was water and mobile phase B was acetonitrile. The gradient conditions were as follows: 0–12 min, 19% B; 12–60 min, 19–36% B; 60–77 min, 36% B; 77–80 min, 36–19% B; 80 min, 19% B. The flow rate was 0.6 mL/min, the injection volume was 10 μL, the column temperature was 30°C, and the samples were detected by absorption at 203 nm. The method was validated by measuring precision, repeatability, and stability according to the ICH Harmonized Tripartite Guideline (2006). A Similarity Evaluation System for Chromatographic Fingerprinting of TCM (Version 2004 A, Chinese Pharmacopoeia Committee, Beijing, China) to analyze the similarity and fingerprint was used.

The sampling amount of each origin was based on the amount of crude drug. First, 0.2 g was placed into a 10-mL centrifuge tube, to which 5 mL of methanol was added. After filtering through a 0.22-μm nylon membrane filter, the test solution was put into a 2-mL sample bottle. The mixed reference solution of notoginsenoside R_1, ginsenoside Rg_1, and ginsenoside Rb_1 was prepared according to the Chinese Pharmacopoeia 2020 Edition (Commission C.P., 2020) and identified through HPLC. Based on the standard curve of the control solution concentration, the test solution content was calculated.
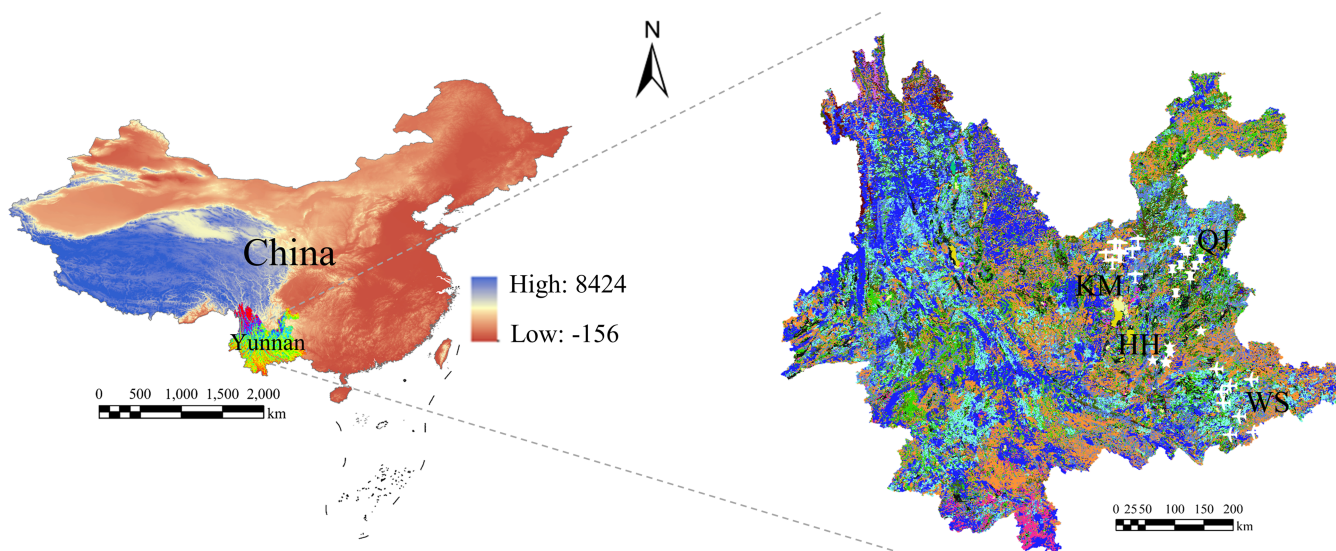
FIGURE 1　Geographical information of *P. notoginseng*. (Left) A map of China. (Right) A map of Yunnan province. HH: Honghe, Yunnan province; KM: Kunming, Yunnan province; QJ: Qujing, Yunnan province, WS: Wenshan, Yunnan province

2.2.2 | Preparation methods

The powder was obtained from 10 g of each sample, and 80 mL of 80% ethanol was added. Then, they were ultrasonically extracted for 30 min, followed by vacuum filtration. This procedure was repeated, and the filtrates were combined. Samples were distilled under reduced pressure until the solvent had evaporated, so we could obtain the total extract of *P. notoginseng* main roots. PBS (pH = 7.4) was prepared with 0.01 M dibasic sodium phosphate. The amount of sample of each origin was converted to 0.2 g raw drug, and 2 mL PBS was added to the total extract. Male 220–260-g Sprague Dawley rats were anesthetized with intraperitoneal injection, and 2 mL of blood was taken from the intraperitoneal vein (0.2 mL of sodium citrate + 1.8 mL of venous blood), gently reversed, and separated within 1 h. Plasma of rats was prepared by centrifugation at 3,000 r/min for 15 min. The positive control was prepared by diluting 0.01-g/tablet of rivaroxaban to $1*10-6$ g·mL^{-1}.

2.2.3 | Determination of the four indicators of blood coagulation

The four indicators of blood coagulation were determined by applying a fully automatic bolld coagulation analyzer (CA-600, Sysmex, Japan). First, 200 μL of plasma was placed in 2 mL collection of centrifuge tube and 100 μL *P. notoginseng* samples was added At the same time,100 μL of PBS was added to the blank control group and 100 μL of diluted rivaroxaban was added to the positive control group. The indicators were determined three times in parallel per sample.

2.2.4 | Correlation coefficient analysis

SPSS 21.0 was used to process and analyze experimental data and to determine whether saponin content was significantly different in *P. notoginseng* main roots of different origins. Data were examined for normality and homogeneity of variances, and one-way analysis of variance was conducted. In the univariate homogeneity test, variance was considered homogeneous if $P > 0.05$ and a significant difference was considered if $P < 0.05$.

2.3 | Spectral analysis

2.3.1 | ATR-FTIR spectral acquisition

FTIR spectra were collected with a Frontier FT spectrometer (PerkinElmer, USA) equipped with a deuterated triglycine sulfate detector, which coupled with ZnSe attenuated total reflectance accessory (Perkin Elmer, Norwalk, CT, USA). The instrument was preheated at 65% relative humidity for 30 min prior to the analysis. Furthermore, it was required that the indoor temperature at 25°C and the relative humidity of under 45% be sustained throughout the whole scanning procedure to reduce the absorption of CO_2 and H_2O when scanning the sample in order to reduce the background signal. Spectra were scanned in the range of 4,000–400 cm^{-1}. Each sample was scanned 16 times at a resolution of 4 cm^{-1} and tested three times in parallel. Finally, the average spectrum was calculated to establish a model for the next analysis.

2.3.2 | Data preprocessing

The original spectrum reveals a lot of chemical information, but there were peak overlaps and interferences such as stray light, noise, and baseline drift.[40] Therefore, the derivative,[41] multiplicative scatter correction (MSC) and standard normal variable (SNV),[42] Savitzky–Golay filtering,[43] and their combinations were applied for preprocessing to eliminate the unnecessary signal variations. Before preprocessing, MATLAB 2017a (the MathWorks) was applied to divide the data into the test set and the training set based on the classic Kennard–Stone algorithm to eliminate human interference. Among them, 238 samples were used as the training set, and the other 121 samples were used as the test set. All preprocessing methods were carried out through SIMCA-P+ 14.0 software (Umetrics, Umea, Sweden).

2.3.3 | Selection of feature variables

The main purpose of variable selection is to select relevant and information-rich data to reduce the dimensionality and redundancy of the data. Besides, it could decrease irrelevant information's interference with the model and improve the efficiency of the model and the accuracy and reliability of the predictions.[44] The VIP method is the most commonly applied variable selection method in partial least squares (PLS) models and has been widely used to select important variables in PLS/partial least squares discriminant analysis (PLS-DA) models to reduce model dimensions and enhance interpretability.[45] The confidence interval of VIP was 95%. When the VIP value is >1, this indicated that a variable is important; if it lies between 0.5 and 1, the importance of the variable should be analyzed according to the specific problem; if the VIP value is <0.5, this means that the variable is not important.[33] Therefore, the variables with a VIP value of >1 in the 4,000–400 bands were screened through SIMCA-P+ 14.0 software (Umetrics, Umea, Sweden) for further modeling analysis.

The CARS algorithm is a newly established method for extracting wavelengths of characteristic variables. Foremost, Monte Carlo sampling or random sampling was applied to select a part of the samples in the calibration set, and then the maximum absolute value of the regression coefficient that is based on the PLS model through the adaptive weighted sampling method was retained to evaluate each variable's importance. Finally, models were established for each subset through cross-validation, selecting the wavelength variable subset for the smallest value of the root mean squares error of cross-validation as the optimal subset.[46] The method was performed using MATLAB software (version R2017b, MathWorks, USA). In this study,

the number of Monte Carlo simulation was set to 500, the group number for cross-validation was set as 7-fold, and the pretreatment method was determined as center.

2.3.4 | Pattern recognition technology

PLS-DA is widely used as a method for linear multivariate data discrimination in chemometrics, with applications in food and herb areas, such as the authentication of origin, cultivation models, processing methods, fraud, etc.[47–50] It could be used to calculate the probability of each class and select the class associated with the highest probability for sample classification, which is widely applied to cope with complicated data matrices through dimensionality reduction. Therefore, PLS-DA was performed to establish a discrimination model to identify the origin of *P. notoginseng* in this study. The evaluation of classification performance parameters was performed based on a confusion matrix. The total numbers of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) samples were summarized to calculate the sensitivity, specificity, accuracy, and performance of the model. A good model must have high sensitivity and specificity coefficients. The closer the value is to 1, the better the model effect. PLS-DA models were established through SIMAC-P+ 14.0 (Umetrics, Sweden) software. The related equations are as follows:

$$Sensitivity = \frac{TP}{TP + FN} \times 100\%, \qquad (1)$$

$$Specificity = \frac{TN}{TN + FP} \times 100\%, \qquad (2)$$

$$Accuracy = \frac{TP + TN}{Total\ samples} \times 100\%. \qquad (3)$$

3 | RESULTS AND DISCUSSION

3.1 | Saponin content and the four indicators of blood coagulation

The content of three saponins in *P. notoginseng* samples from different origins was calculated according to the content determination method. The resuts are shown in Table 1. The saponin content of *P. notoginseng* main roots was not significantly different. Their saponin content was ordered as followed: WS (9.86%) > HH (7.60%) > KM (7.56%) > QJ (7.53%).

The results of the four indicators of blood coagulation are shown in Table 2. The correlation analysis of saponin content and the four indicators of blood coagulation was performed using SPSS 21.0 software, the results are shown in Table 3. There was an extremely significant correlation between PT and APTT and the content of notoginsenoside R_1, ginsenosides Rg_1 and Rb_1, and the total content of the three saponins. TT had moderate positive correlations with the contents of notoginsenoside R_1 and ginsenoside Rg_1 and a weak positive correlation with ginsenoside Rb_1 and the total content of the three saponins. FIB had a weak negative correlation with the contents of notoginsenoside R_1 and ginsenoside Rg_1 and a significant moderate

TABLE 3 Correlation coefficients of saponins

Saponin component	PT	APTT	TT	FIB
Notoginsenoside R_1	0.517**	0.523**	0.358*	−0.255
Ginsenoside Rg_1	0.851**	0.680**	0.309	−0.273
Ginsenoside Rb_1	0.802**	0.762**	0.211	−0.321*
Total of saponins	0.858**	0.744**	0.298	−0.309

*$p < 0.05$. **$p < 0.001$.

TABLE 1 Contents of three saponins in *P. notoginseng*

No.	State/origin	Notoginsenoside R_1 (%)	Ginsenoside Rg_1 (%)	Ginsenoside Rb_1 (%)	Total saponins (%)
1	WS	1.13 ± 0.40a	4.99 ± 2.15a	3.89 ± 1.58a	9.86 ± 3.72a
2	HH	0.93 ± 0.21a	3.74 ± 0.49a	2.92 ± 0.32a	7.60 ± 0.82a
3	QJ	0.98 ± 0.17a	3.79 ± 0.80a	2.75 ± 0.51a	7.53 ± 1.29a
4	KM	0.99 ± 0.15a	3.79 ± 1.01a	2.78 ± 0.44a	7.56 ± 1.31a

TABLE 2 The four indicators of blood coagulation

No.	State/origin	PT (s)	APTT (s)	TT (s)	FIB (%)
1	WS	4.99 ± 2.30a	30.48 ± 12.83a	20.87 ± 8.49a	0.23 ± 0.13b
2	HH	3.35 ± 0.36b	23.49 ± 5.12a	17.93 ± 6.33a	0.30 ± 0.05ab
3	QJ	3.97 ± 0.64ab	24.00 ± 5.43a	20.64 ± 6.99a	0.30 ± 0.10ab
4	KM	3.95 ± 1.43ab	25.53 ± 6.31a	21.27 ± 5.66a	0.32 ± 0.04a

negative correlation with ginsenoside Rb_1 and the total contents of the three saponins. The results show that higher contents of the three saponins were associated with a longer prolongation time of PT, APTT, and TT and lower FIB content. The animal experiments showed that the effects of *P. notoginseng* from different origins on rat plasma could prolong PT, APTT, and TT and reduce the concentration of FIB. The results of prolongation of PT of *P. notoginseng* in four different origins are as follows: WS (4.99 s)>QJ (3.97 s)>KM (3.95 s)>HH (3.35 s). The correlation analysis results showed that PT, APTT, and the contents of four saponins had different degrees of positive correlation and FIB had different degrees of negative correlation with the contents of four saponins. That is to say, the higher the saponin content, the better the anticoagulant effect of *P. notoginseng*. Among them, the content of ginsenoside Rg_1 has the greatest influence on the anticoagulant effect.

3.2 | ATR-FTIR spectral analysis and preprocessing

Figure 2 shows the average spectrum and preprocessing based on the different origins in the range of 4,000 to 400 cm^{-1}. As seen from Figure 2A, different origins' sample absorption peaks are the same, and the peak intensity is marginally different. Bands around 3,285 cm^{-1} were attributable to the symmetric and asymmetric stretching vibration of O-H. The peak at 2,922 cm^{-1} was mainly assigned to the asymmetric stretching vibration of CH_2.[51] The peak at 1,753 cm^{-1} was assigned to the carbonyl peak, and the peak at 1,630 cm^{-1} was assigned to the bending vibration of O-H. The broad bands around 1,411 cm^{-1} were mainly assigned to the transformation vibration of C-H. The peak at 1,368 cm^{-1} corresponded to NO_3^-, indicating the signature of nitrate in the samples.[52] The peak at 1,236 cm^{-1} showed the CH_2OH mode and the bands at 1,145 and
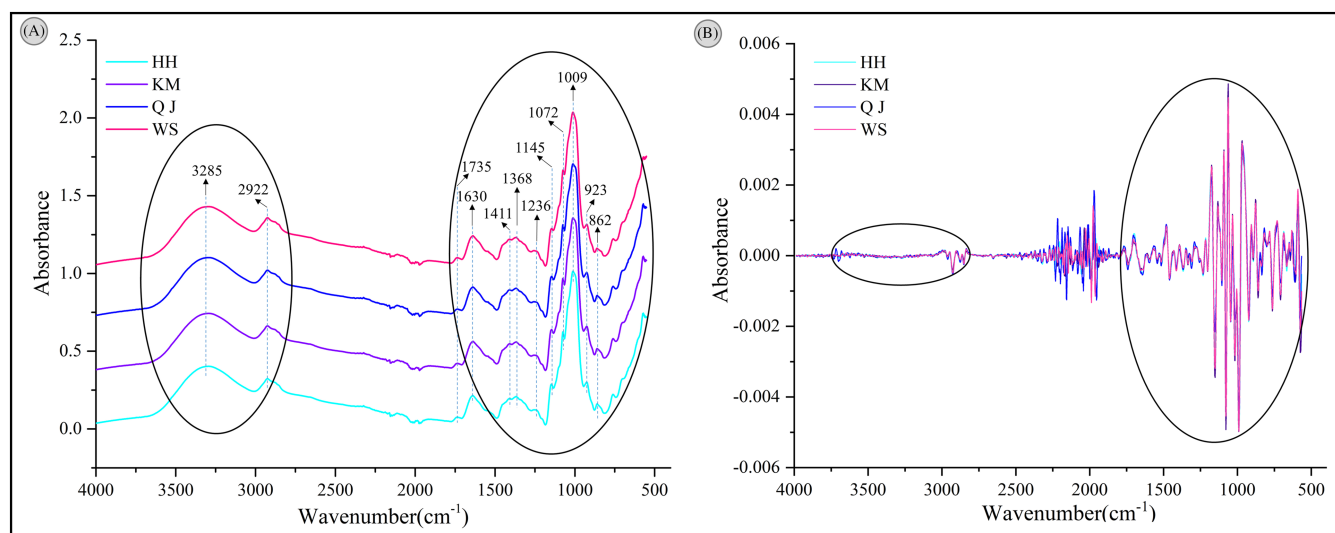
FIGURE 2 Original whole wavelength range ATR-FTIR spectra (A) and second derivative whole wavelength range ATR-FTIR spectra (B) of *P. notoginseng* from different origins. The circles represent bands of manually selected wavelengths. HH: Honghe Yunnan; KM: Kunming Yunnan; QJ: Qujing Yunnan; WS: Wenshan Yunnan

TABLE 4 The classification accuracy for each class and total accuracy

Pretreatment method	Training set					Test set				
	1	2	3	4	ACC	1	2	3	4	ACC
None	0.800	0.384	0.717	0.550	0.613	0.600	0.621	0.867	0.733	0.705
1st	0.817	0.850	0.883	0.800	0.838	0.933	0.897	0.967	0.900	0.924
2nd	**0.933**	**0.914**	**0.933**	**0.901**	**0.920**	**0.966**	**1.000**	**0.967**	**0.933**	**0.967**
MSC	0.767	0.833	0.850	0.684	0.783	0.867	0.966	0.933	0.933	0.924
SNV	0.800	0.850	0.850	0.650	0.786	0.833	0.931	0.967	0.933	0.916
2nd + SMC	0.967	0.983	0.917	0.900	0.942	0.967	1.000	1.000	0.933	0.975
2nd + SNV	0.966	0.950	0.900	0.883	0.925	0.967	1.000	1.000	1.000	0.992
2nd + SMC + SG	0.900	0.900	0.816	0.850	0.867	1.000	0.966	0.933	0.866	0.941
2nd + SNV + SG	0.800	0.883	0.783	0.800	0.817	0.833	0.897	0.867	0.833	0.857

1,072 cm^{-1} were due to symmetric vibration of C-C and C-O. The peak at 1,009 cm^{-1} was mainly assigned to the bending vibration of C-O-H.[53] The peaks at 923 and 862 cm^{-1} represented the stretching vibration of C-C and the bending vibration of C-H, respectively.[54] According to the original spectrum, the different origins could not be directly discriminated from the spectra only by the naked eye; this was owing to the similar chemical composition of *P. notoginseng* samples.

Each ATR-FTIR spectrum dataset was preprocessed by different methods. The results show that the second derivative (2nd) spectrum has the best effect (Table 4). It improved the change rate of the whole spectrum and revealed slight differentiation. As shown by a circle in Figure 2B, the preprocessing 2nd spectra were located in two spectral regions, which was consistent with the original spectral absorption peak (Figure 2A) results. The characteristic peaks in the 1,800–500 cm^{-1} and 3,700–2,800 cm^{-1} regions were more obvious, and the absorption intensities of the characteristic peaks for KM and WS were higher than those of other origins. That is to say, some variance existed in the types and components of saponins, flavonoids, sugars, and others of different origins. Therefore, the supervised PLS-DA model was applied based on the 2nd spectra for further analysis.

3.3 | PLS-DA analysis of the whole wavelength and manually selected wavelength range

First, a PLS-DA model of the whole wavelength range was established. In addition, some wavelength data of the whole spectral region could contain useless information, which may influence the establishment of discrimination models. Therefore, wavelengths with a high correlation with target sample features were selected, which has the ability to eliminate redundant information and increase model performance.[55] To eliminate regions and noises without sample information, as shown in Section 3.2, the regions of the main absorption peak distribution (3,700–2,800 cm^{-1} and 1,800–500 cm^{-1}) were manually selected to establish the PLS-DA model.

The classification parameters and accuracy of the classification models established through the whole wavelength and manually selected wavelengths are shown in Tables 4 and 5. The models with high classification accuracy were considered as the best discriminant equation. The model accuracy for the training set and the test set of the whole wavelength range was up to 0.920 and 0.967, respectively, demonstrating a high accuracy in the model. Relatively speaking, the model performance of the manual selection decreased, as the accuracy was 0.912 and 0.934, respectively. In addition, the confusion matrix of PLS-DA models showed that for the whole wavelength, 19 out of 238 training set samples and 4 out of 121 test set samples were misclassified (Figure 3A). Relatively speaking, the PLS-DA models with manually selected wavelengths have more misclassifications than those for the whole wavelength. The PLS-DA model of the whole wavelength could more effectively classify *P. notoginseng* than the manually selected wavelengths. The explanation could be owing to the manual selection of the main absorption peak region, while ignoring some spectral information that was important for multivariate analysis.

3.4 | PLS-DA analysis based on variable selection

As mentioned earlier, variable selection strategies could be used to select the important variables that influence the model, allowing for

TABLE 5 The classification parameters of PLS-DA models based on original data and variable selection methods

| Selection method	Class	Training set			Test set		
		Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
Whole	HH	0.933	0.961	0.954	0.996	1	0.992
	KM	0.914	0.972	0.958	1	1	1
	QJ	0.933	0.978	0.966	0.967	0.967	0.967
	WS	0.900	0.983	0.962	0.933	0.990	0.922
Select	HH	0.950	0.961	0.958	1	0.967	0.975
	KM	0.862	0.967	0.941	0.871	0.989	0.959
	QJ	0.933	0.972	0.962	0.935	0.978	0.975
	WS	0.900	0.983	0.962	0.931	0.978	0.959
VIP	HH	0.967	0.983	0.979	1	1	1
	KM	0.966	0.989	0.983	1	1	1
	QJ	0.983	0.994	0.992	0.983	1	0.983
	WS	1	1	1	1	0.978	0.983
CARS	HH	0.950	0.972	0.966	0.967	0.978	0.975
	KM	0.897	0.983	0.962	0.968	0.978	0.975
	QJ	0.933	0.972	0.962	0.933	1	0.933
	WS	0.933	0.978	0.966	1	1	1
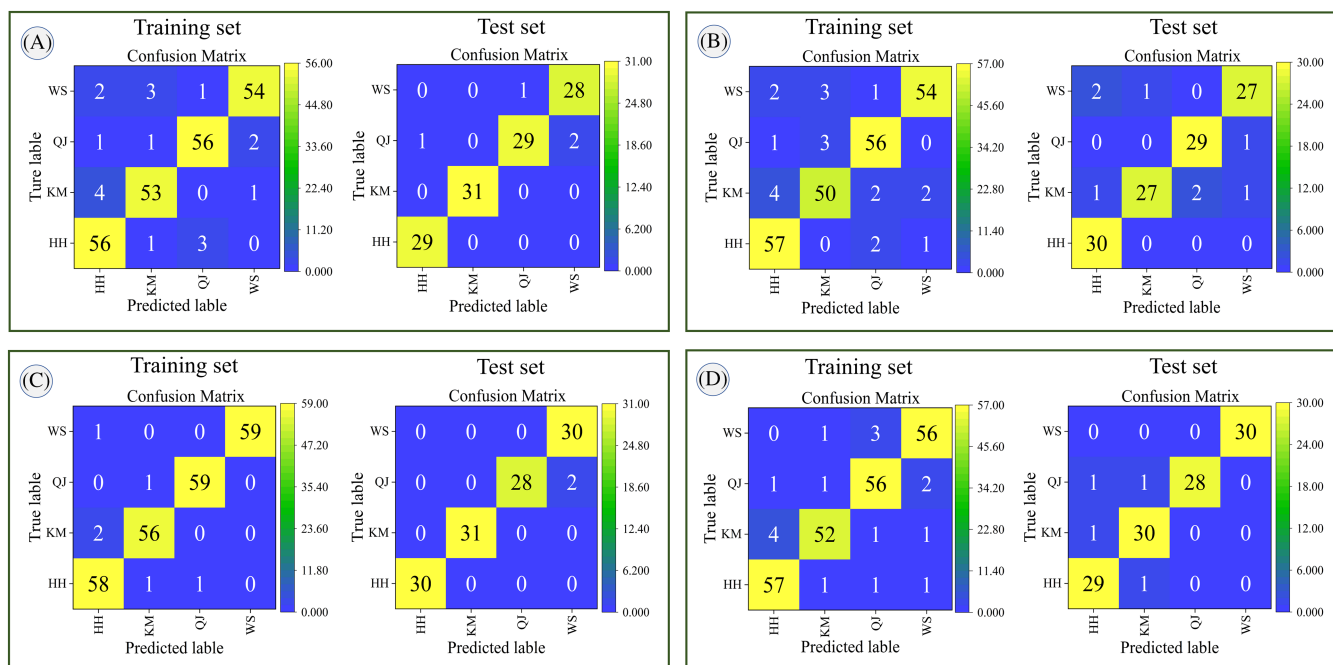
FIGURE 3 The confusion matrix of PLS-DA models. (A) Original second derivative data. (B) Manual selection data. (C) The feature of VIP. (D) The feature of CARS. The abscissa corresponds to the predicted samples of *P. notoginseng*, and the ordinate corresponds to the actual samples of *P. notoginseng*. The samples that appear on the diagonal line show that the predicted value was consistent with the actual value. The data that appear outside the diagonal line are considered as misclassified. The darker the color, the more samples are correctly classified

TABLE 6 The classification accuracy and total accuracy for different variable selection methods

Variable selection method	Training set					Test set				
	1	2	3	4	ACC	1	2	3	4	ACC
Whole	0.933	0.914	0.933	0.901	0.920	0.966	1.000	0.967	0.933	0.967
Select	0.950	0.862	0.933	0.900	0.912	1	0.871	0.967	0.900	0.934
VIP	**0.967**	**0.967**	**0.983**	**0.983**	**0.975**	**1**	**1**	**0.933**	**1**	**0.984**
CARS	0.950	0.897	0.933	0.933	0.929	0.967	0.968	0.933	1	0.967

the analysis of the question posed. We selected variables with VIP value > 1 as the important wavelengths (Figure 4). Based on the ATR-FTIR spectra of the *P. notoginseng* samples, 760 variables' wavelengths were selected. Then, the selected variables' wavelengths were reintegrated into 359 samples, and a data matrix of 359 samples × 760 variables was established, which was analyzed by the PLS-DA model. Subsequently, the CARS algorithm was used to extract the important variables in the ATR-FTIR spectra. One of the great advantages of the CARS algorithm is that the variable selection of each algorithm's strategy could offer the analogous probability of being sampled for each variable, which raises selected variables' opportunities greatly. In the present work, the ATR-FTIR spectral data were input into the script of the CARS algorithm, and 125 variables' wavelengths were finally screened. Therefore, a 359 samples ×125 variables data matrix was built and further analyzed.

The classification parameters, classification accuracy, and confusion data matrix of the PLS-DA model that was established through the two variable selection methods are demonstrated in Table 5, Table 6, and Figure 3, respectively. The variable selection methods of VIP have the highest accuracy, with the training set and test set reaching 0.975 and 0.984, respectively, being greater than 0.95. In addition, for the VIP, the confusion matrix results of the PLS-DA model showed that 4 samples out of 238 samples in the training set were misclassified, and 2 samples out of 121 samples in the test set were misclassified (Figure 3C), which indicated that the model could identify the origins of *P. notoginseng* samples with high accuracy. Relatively speaking, the PLS-DA model results of the CARS algorithm are slightly worse, with the training set and test set reaching 0.929 and 0.967, respectively. The confusion matrix results show that 17 samples among 238 samples (training set) were misclassified, and 4 samples among 124 samples (test set) were misclassified (Figure 3D). From the
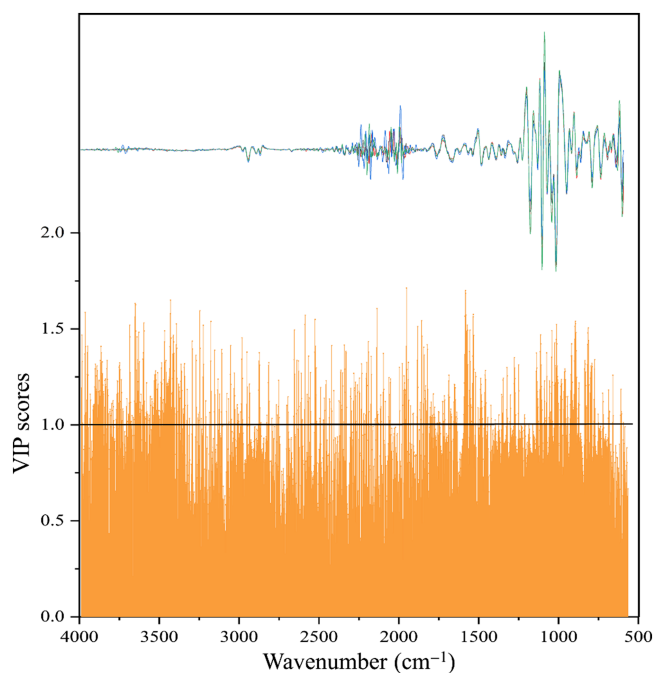
FIGURE 4 Selection of feature variables based on the VIP strategy. The wavelengths with VIP values greater than 1 were selected as feature variables

perspective of the discriminant models based on the variable selection methods using the PLS-DA model, the prediction accuracy of the models based on the selected features through the VIP was better than that of the models based on the whole wavelength, manually selected wavelengths, and CARS. No significant improvement was noticed in the models that were constructed with the variables selected by CARS with regard to the models that were established with the whole wavelength. The main reason for this may be that with the feature variables extracted using the algorithm, redundant information still exists. In addition, an interesting finding is that compared to the 760 variables selected based on VIP, CARS selected only 125 variables, indicating that it provided a more concise, convenient, and fast model than VIP. However, unfortunately it reduces the classification capacity of the model. All in all, in this article, the PLS-DA model of variable selection methods performed better than the whole wavelength and manually selected wavelengths, and the VIP was the best.

4 | CONCLUSIONS

The inherent quality of Chinese herb medicines is related to their safety and clinical efficacy. In recent years, the short supply of *P. notoginseng* in the market has become increasingly serious. The introduction from different places is an inevitable trend for the future of the economic development of *P. notoginseng*. However, the accumulation of chemical components in medicinal parts is easily affected

by artificial and external factors, which result in differences in the quality of *P. notoginseng* from different origins. Dencichine and quercitrin, which were isolated from *P. notoginseng*, exert hemostatic effects, and the components that exert anticoagulant activity are mainly saponins. Modern pharmacological studies have shown that *P. notoginseng* has powerful blood-activating effects, but there is no clear report on its main components, which are mostly believed to be saponins. The four indicators of blood coagulation in this study can be used to judge whether there exists abnormality of coagulation function. Analysis was performed based on the contents of three saponins and their anticoagulant effects in the notoginseng extract from four different origins, which revealed the relationship between the saponin content and anticoagulant effects in *P. notoginseng* from different origins. Meanwhile, the different origins were identified.

In the present work, the three saponins in *P. notoginseng* from different origins were determined by HPLC, and the influence of origin on the anticoagulant activity of *P. notoginseng* was analyzed. The ATR-FTIR spectral data were analyzed by manual selection and two variable selection strategies to identify the different origins, which demonstrated the feasibility of variable selection strategies in identifying *P. notoginseng* from different origins. The results showed that the saponin content in Wenshan Prefecture was the highest (9.86%), the coagulation time was prolonged for the best effect (4.99 s), and the PT was positively correlated with contents of the three saponins and total saponin content, with correlation coefficients of 0.517**, 0.851**, 0.802**, and 0.858**, respectively. The PLS-DA model using variable selection provided a reliable technique for the identification of *P. notoginseng* from different origins. However, VIP achieved the best performance with accuracy values of 0.975 and 0.984 for the training and test sets, respectively, so VIP was slightly better than CARS. To sum up, in terms of saponin content, the quality of *P. notoginseng* from Wenshan Prefecture was better. The higher the saponin content, the better the anticoagulant effect of *P. notoginseng*. The content of ginsenoside Rg_1 has the greatest influence on the anticoagulant effect. In addition, the variable selection strategy results are better than the original spectrum results, and a small number of variables containing effective information can be extracted. This study provides valuable insights into the main anticoagulant components of *P. notoginseng* from different origins, better results for the identification of *P. notoginseng* from different origins, and a scientific basis for the quality control and rational utilization of *P. notoginseng*.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Yuan-Zhong Wang https://orcid.org/0000-0001-5376-757X

Fu-Rong Xu https://orcid.org/0000-0001-9588-9444

REFERENCES

1. Li J, Wang RF, Zhou Y, et al. Dammarane-type triterpene oligoglycosides from the leaves and stems of *Panax notoginseng* and their antiinflammatory activities. *J Ginseng Res*. 2019;46:377-384. doi:10.1016/j.jgr.2017.11.008

2. Xiong Y, Chen LJ, Man JH, Hu Y, Cui X. Chemical and bioactive comparison of *Panax notoginseng* root and rhizome in raw and steamed forms. *J Ginseng Res*. 2019;43:385-393. doi:10.1016/j.jgr.2017.11.004

3. Yang XC, Xiong XJ, Wang HR, Wang J. Protective effects of *panax notoginseng* saponins on cardiovascular diseases: a comprehensive overview of experimental studies. *Evid-Based Compl Alt*. 2014;2014:204840.

4. Wang T, Guo RX, Zhou GH, et al. Traditional uses, botany, phytochemistry, pharmacology and toxicology of Panax notoginseng (Burk.) FH Chen: a review. *J Ethnopharmacol*. 2016;188:234-258. doi:10.1016/j.jep.2016.05.005

5. Peng M, Yi YX, Zhang T, Ding Y, Le J. Stereoisomers of saponins in *Panax notoginseng* (Sanqi): a review. *Front Pharmacol*. 2018;9:188. doi:10.3389/fphar.2018.00188

6. Fung FY, Wong WH, Ang SK, et al. A randomized, double-blind, placebo-controlled study on the anti-haemostatic effects of *Curcuma longa*, *Angelica sinensis* and *Panax ginseng*. *Phytomedicine*. 2017;32:88-96. doi:10.1016/j.phymed.2017.04.004

7. Chan P, Thomas GN, Tomlinson B. Protective effects of trilinolein extrated from *Panax notoginseng* against cardiovascular disease. *Acta Pharmacol Sin*. 2002;23:1157-1162.

8. Liu HB, Lu XY, Hu Y, Fan XH. Chemical constituents of Panax ginseng and Panax notoginseng explain why they differ in therapeutic efficacy. *Pharmacol Res*. 2020;161:105263.

9. Zhao HP, Han ZP, Li GW, Zhang S, Luo Y. Therapeutic potential and cellular mechanisms of *Panax notoginseng* on prevention of aging and cell senescence-associated diseases. *Aging Dis*. 2017;8(6):721-739. doi:10.14336/AD.2017.0724

10. Hawthorne B, Lund K, Freggiaro S, Kaga R, Meng J. The mechanism of the cytotoxic effect of *Panax notoginseng* extracts on prostate cancer cells. *Biomed Pharmacother*. 2022;149:112887. doi:10.1016/j.biopha.2022.112887

11. Yang QB, Wang PW, Cui JG, Wang W, Chen Y, Zhang T. *Panax notoginseng* saponins attenuate lung cancer growth in part through modulating the level of Met/miR-222 axis. *J Ethnopharmacol*. 2016;193:255-265. doi:10.1016/j.jep.2016.08.040

12. Shi XW, Yu WJ, Li LX, et al. *Panax notoginseng* saponins administration modulates pro- /anti-inflammatory factor expression and improves neurologic outcome following permanent MCAO in rats. *Metab Brain Dis*. 2017;32:221-233. doi:10.1007/s11011-016-9901-3

13. Wang JR, Yau LF, Gao WN, et al. Quantitative comparison and metabolite profiling of saponins in different parts of the root of *Panax notoginseng*. *J Agric Food Chem*. 2014;62(36):9024-9034. doi:10.1021/jf502214x

14. Kim DH. Chemical Diversity of *Panax ginseng*, *Panax quinquifolium*, and *Panax notoginseng*. *J Ginseng Res*. 2012;36:1-15. doi:10.5142/jgr.2012.36.1.1

15. Ren YS, Ai J, Liu XQ, et al. Anticoagulant active ingredients identification of total saponin extraction of different panax medicinal plants based on grey relational analysis combined with UPLC-MS and molecular docking. *J Ethnopharmacol*. 2020;260:11295.

16. Du XH, Zhao YL, Yang DF, et al. A correlation model of UPLC fingerprints and anticoagulant activity for quality assessment of *Panax notoginseng* by hierarchical clustering analysis and multiple linear regression analysis. *Anal Methods-UK*. 2015;7:2985-2992.

17. Xiong LX, Qi Z, Zheng BZ, et al. Inhibitory Effect of Triterpenoids from *Panax ginseng* on Coagulation Factor X. *Molecules*. 2017;22(4):649. doi:10.3390/molecules22040649

18. Dong JE, Wang Y, Zuo ZT, Wang YZ. Deep learning for geographical discrimination of *Panax notoginseng* with directly near-infrared spectra image. *Chemometr Intell Lab*. 2020;197:103913.

19. Li Y, Zhang J, Xu FR, et al. Rapid prediction study of total flavonids content in *panax notoginseng* using infrared spectroscopy combined with chemometrics. *Spectrosc Spect Anal*. 2017;37:70-74.

20. Luo JY, Chen GS, Liu DH, et al. Study on the material basis of houpo wenzhong decoction by HPLC fingerprint, UHPLC-ESI-LTQ-Orbitrap-MS, and network pharmacology. *Molecules*. 2019;24(14):2561. doi:10.3390/molecules24142561

21. Li X, Wang YR, Ma L, Cui JX, Hong WX. Application of the quality evaluation of traditional Chinese herbal medicines using chromatography of fingerprint. *J Biomed Eng*. 2012;29:192-196.

22. Zhang XF, Zhang SJ, Gao BB, et al. Identification and quantitative analysis of phenolic glycosides with antioxidant activity in methanolic extract of *Dendrobium catenatum* flowers and selection of quality control herb-markers. *Food Res Int*. 2019;123:732-745.

23. Guo L, Gong MX, Wu S, Qiu F, Ma L. Identification and quantification of the quality markers and anti-migraine active components in Chuanxiong Rhizoma and Cyperi Rhizoma herbal pair based on chemometric analysis between chemical constituents and pharmacological effects. *J Ethnopharmacol*. 2020;246:112228. doi:10.1016/j.jep.2019.112228

24. Xu J, Zhou RR, Luo L, Dai Y, Feng YR, Dou ZH. Quality evaluation of decoction pieces of gardeniae fructus based on qualitative analysis of the HPLC fingerprint and Triple-Q-TOF-MS/MS combined with quantitative analysis of 12 representative components. *J Anal Methods Chem*. 2022;2022:1-13.

25. Peng C, Zhu YL, Yan FL, et al. The difference of origin and extraction method significantly affects the intrinsic quality of licorice: A new method for quality evaluation of homologous materials of medicine and food. *Food Chem*. 2021;340:127907. doi:10.1016/j.foodchem.2020.127907

26. Sun SS, Li YC, Zhu LJ, Ma HY, Li LP, Liu YF. Accurate discrimination of *Gastrodia elata* from different geographical origins using high-performance liquid chromatography fingerprint combined with boosting partial least-squares discriminant analysis. *J Sep Sci*. 2019;42:2875-2882.

27. Wang Y, Shen T, Zhang J, Huang HY, Wang YZ. Geographical authentication of *gentiana rigescens* by High-Performance liquid chromatography and infrared spectroscopy. *Anal Lett*. 2018;51:2173-2191.

28. Nie H, Zhang H, Zhang XQ, et al. Relationship between HPLC fingerprints and in vivo pharmacological effects of a traditional Chinese medicine: Radix Angelicae Dahuricae. *Nat Prod Res*. 2011;25:53-61.

29. Casale M, Bagnasco L, Zotti M, Piazza SD, Sitta N, Oliveri P. A NIR spectroscopy-based efficient approach to detect fraudulent additions within mixtures of dried *porcini* mushrooms. *Talanta*. 2016;160:729-734.

30. Biancolillo A, Marini F, Ruckebusc C, Ruckebusch C, Vitale R. Chemometric strategies for Spectroscopy-Based food authentication. *Applied Sciences*. 2020;10(18):6544. doi:10.3390/app10186544

31. Yue JQ, Li ZM, Zuo ZT, Zhao YL, Zhang J, Wang YZ. Study on the identification and evaluation of growth years for *Paris polyphylla* var. *Yunnanensis* using deep learning combined with 2DCOS. *Spectrochim Acta A*. 2021;261:120033. doi:10.1016/j.saa.2021.120033

32. Chen H, Lin Z, Tan C. Fast discrimination of the geographical origins of notoginseng by near-infrared spectroscopy and chemometrics. *J Pharmaceut Biomed*. 2018;161:239-245.

33. Liu L, Zuo ZT, Wang YZ, Xu FR. A fast multi-source information fusion strategy based on FTIR spectroscopy for geographical authentication of wild *Gentiana rigescens*. *Microchem J*. 2020;159:105360.

34. Du QW, Zhu M, Shi T, et al. Adulteration detection of corn oil, rapeseed oil and sunflower oil in camellia oil by *in situ* diffuse reflectance near-infrared spectroscopy and chemometrics. *Food Control*. 2021; 121:107577. doi:10.1016/j.foodcont.2020.107577

35. Liu CL, Zuo ZT, Xu FR, Wang YZ. Authentication of herbal medicines based on modern analytical technology combined with chemometrics approach: A review. *Crit Rev Anal Chem*. 2022;1-26. doi:10.1080/10408347.2021.2023460

36. Zhou XH, Xiang BR, Wang ZM, Zhang M. Determination of quercetin in extracts of ginkgo biloba l. Leaves by near-infrared reflectance spectroscopy based on interval partial least-squares (iPLS) model. *Anal Lett*. 2007;40:3383-3391.

37. Pan W, Wu M, Zheng ZZ, Guo L, Lin Z, Qiu B. Rapid authentication of *Pseudostellaria heterophylla* (Taizishen) from different regions by near-infrared spectroscopy combined with chemometric methods. *J Food Sci*. 2020;85(7):2004-2009. doi:10.1111/1750-3841.15171

38. Liu ZM, Yang SB, Wang YZ, Zhang J. Multi-platform integration based on NIR and UV-Vis spectroscopies for the geographical traceability of the fruits of *Amomum tsao-ko*. *Spectrochim Acta A*. 2021;258:119872. doi:10.1016/j.saa.2021.119872

39. Li SL, Xing BC, Lin D, Yi HJ, Shao QS. Rapid detection of saffron (*Crocus sativus* L.) Adulterated with lotus stamens and corn stigmas by near-infrared spectroscopy and chemometrics. *Ind Crop Prod*. 2020; 152:112539.

40. Li Y, Shen Y, Yao CL, Guo DA. Quality assessment of herbal medicines based on chemical fingerprints combined with chemometrics approach: A review. *J Pharmaceut Biomed*. 2020;185:113215.

41. Arndt M, Drees A, Ahlers C, Fischer M. Determination of the Geographical Origin of Walnuts (*Juglans regia* L.) Using Near-Infrared Spectroscopy and Chemometrics. *Foods*. 2020;9(12):1860. doi:10.3390/foods9121860

42. Dhanoa MS, Lister SJ, Sanderson R, Barnes RJ. The link between multiplicative scatter correction (MSC) and standard normal variate (SNV) transformations of NIR spectra. *J near Infrared Spec*. 1994;2:43-47. doi:10.1255/jnirs.30

43. Savitzky A, Golay MJE. Smoothing and differentiation of data by simplified least squares procedure. *Anal Chem*. 1964;36(8):1627-1639. doi:10.1021/ac60214a047

44. Pei YF, Zhang QZ, Wang YZ. Application of authentication evaluation techniques of ethnobotanical medicinal plant genus *paris*: A review. *Crit Rev Anal Chem*. 2020;50:405-423.

45. Galindo-Prieto B, Trygg J, Geladi P. A new approach for variable influence on projection (VIP) in O2PLS models. *Chemometr Intell Lab*. 2017;160:110-124.

46. Li HD, Liang YZ, Xu QS, Cao DS. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Anal Chim Acta*. 2009;648:77-84.

47. Lu XH, Xia ZY, Qu FF, Zhu ZM, Li SW. Identification of authenticity, quality and origin of saffron using hyperspectral imaging and multivariate spectral analysis. *Spectrosc Lett*. 2020;53:76-85.

48. Ballabi D, Robotti E, Grisoni F, et al. Chemical profiling and multivariate data fusion methods for the identification of the botanical origin of honey. *Food Chem*. 2018;266:79-89. doi:10.1016/j.foodchem.2018.05.084

49. Walkowiak A, Ledzińsk A, Zapadka M, Ledziński Ł, Kupcewicz B. Detection of adulterants in dietary supplements with *Ginkgo biloba* extract by attenuated total reflectance Fourier transform infrared spectroscopy and multivariate methods PLS-DA and PCA. *Spectrochim Acta A*. 2019;208:222-228. doi:10.1016/j.saa.2018.10.008

50. Górski L, Kowalcze M, Jakubowska M. Classification of six herbal bioactive compositions employing LAPV and PLS-DA. *J Chemometr*. 2019;33(4):e3112. doi:10.1002/cem.3112

51. Li Y, Zhang JY, Wang YZ. FT-MIR and NIR spectral data fusion: A synergetic strategy for the geographical traceability of *Panax notoginseng*. *Anal Bioanal Chem*. 2018;410:91-103.

52. Yang XD, Li GL, Song J, Gao M, Zhou S. Rapid discrimination of Notoginseng powder adulteration of different grades using FT-MIR spectroscopy combined with chemometrics. *Spectrochim Acta A*. 2018;205:457-464. doi:10.1016/j.saa.2018.07.056

53. Ma F, Chen JB, Wu XX, Zhou Q, Sun SQ. Rapid discrimination of *Panax notogeinseng* of different grades by FT-IR and 2DCOS-IR. *J Mol Struct*. 2016;1124:131-137.

54. Yang XD, Song J, Wu X, Xie L, Liu X, Li G. Identification of unhealthy *Panax notoginseng* from different geographical origins by means of multi-label classification. *Spectrochim Acta A*. 2019;222:117243. doi:10.1016/j.saa.2019.117243

55. Moros J, Garrigue S, Guardia MDL. Vibrational spectroscopy provides a green tool for multi-component analysis. *TrAC-Trend Anal Chem*. 2010;29(7):578-591. doi:10.1016/j.trac.2009.12.012