Elemental and isotopic fingerprints combined with chemometrics: A robust strategy for geographical authentication of Zanthoxylum bungeanum Maxim from Gansu GI regions

Chunqiu Chen
Chinese Academy of Forestry
Runhong Mo
Chinese Academy of Forestry
Fubin Tang
Chinese Academy of Forestry
Yihua Liu
Chinese Academy of Forestry
Ling Yang
Sichuan Academy of Forestry
Wenting Liu
Chinese Academy of Forestry
Minghua Qu
Chinese Academy of Forestry
Zhanglin Ni
nizhanglin@aliyun.com
Chinese Academy of Forestry

Article

Keywords: Zanthoxylum bungeanum, geographical origin, multi-element analysis, chemometrics

Posted Date: May 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9709154/v1

Additional Declarations: No competing interests reported.

Abstract

Zanthoxylum bungeanum is the widely favored condiment in both the agricultural industry and the gourmet market. This study established a robust method for identifying the geographical origin of *Zanthoxylum bungeanum* from Gansu province combining stable isotopes ($\delta^{13}C$, $\delta^{15}N$), 43 multi-element fingerprints, and chemometric techniques. One-way analysis of variance (ANOVA) showed that 26 variables (including $\delta^{13}C$ and 25 elements) had significant regional differences in *Zanthoxylum bungeanum* among Gansu GI, Gansu non-GI, and other provinces ($p < 0.05$). The performance of linear discriminant analysis (LDA), random forest (RF), and orthogonal partial least squares discriminant analysis (OPLS-DA) was compared. LDA achieved an accuracy of 89.3% in classifying GI and non-GI products from Gansu province, and an accuracy ranging from 82% to 87% in classifying different GI county areas from Gansu province (Wudu, Qinan, Gangu, and Qingshui). This method demonstrates promising potential as an analytical tool for precise determination of the geographic origin of *Zanthoxylum bungeanum*.

1. Introduction

Zanthoxylum bungeanum, originally from China, has both a seasoning function and medicinal value[1,2]. *Zanthoxylum bungeanum* is an indispensable seasoning in Chinese cuisine, particularly serving as a fundamental component in Sichuan cuisine, where it features in over 90% of regional dishes[3]. Due to the unique aroma and numbing flavor, *Zanthoxylum bungeanum* has become one of the most popular spices in the Asian market[4]. China maintains global dominance in both production and consumption of *Zanthoxylum bungeanum*, with an estimated annual consumption of 450,000 tons, of which *Zanthoxylum bungeanum* Maxim accounts for more than 70%. Gansu province, recognized as the primary cultivation zone for premium red-husked Z. *bungeaum*, produces specimens cultivated in specific arid mountainous regions at elevations between 1,200 and 1,800 meters above sea level. These unique edaphic and altitudinal conditions promote the biosynthesis and accumulation of key flavor compounds, particularly those responsible for the characteristic aroma and numbing qualities[5]. *Zanthoxylum Bungeaum* cultivated in Gansu province boasts the highest number of Geographical Indication (GI) certified production areas officially designated by the People's Republic of China. Products bearing this certification command premium market prices, typically 40–60% higher than non-certified counterparts, reflecting their guaranteed quality and regional authenticity[6].

With the rapid growth of *Zanthoxylum bandeau* planting area in China, accurately determining the geographical origins of *Zanthoxylum bungeanum* products in the domestic market has become a significant challenge. To ensure the authenticity of GI products, it is imperative to implement robust and scientifically sound traceability technologies. There are many common methods for the origin tracing of agricultural products, including elemental fingerprinting analysis[7,8], stable isotope techniques[9,10], organic component fingerprinting[11,12], and near-infrared (NIR) spectroscopy[13,14]. Due to its advantages in metabolomic identification and volatile organic compound analysis, mass spectrometry has demonstrated significant potential for geographical origin authentication[15]. However, as a typical agricultural product rich in volatile substances, *Zanthoxylum bungeanum* has characteristic aroma components and organic components that demonstrate significant chemical instability. These compounds are prone to oxidative degradation during postharvest handling, extended storage, and transportation, which pose great challenges to the traceability of origin[16]. Therefore, it is essential to establish a practical and rapid method to verify the geographic origin of *Zanthoxylum bungeanum*. Extensive empirical research has confirmed that the comprehensive discrimination system for the geographical traceability of agricultural products, constructed based on stable isotopes and multi-element integration, exhibits exceptional reliability[17,18]. In comparison with conventional traceability methods, both stable isotopes and multi-element analysis have the characteristics of high sensitivity and accuracy, as well as fast analysis speed, and have been widely used in various fields[19–21]. For example, the $\delta^{13}C$ isotopic ratio indicates variations in photosynthetic pathways influenced by environmental factors throughout plant growth stages[22], while $\delta^{15}N$ values serve as indications of nitrogen cycling processes affected by different fertilization regimes and soil properties[23]. The multi-element contenttrations in plants is related to environmental factors in different regions[24]. In summary, these parameters create a fingerprint that demonstrates the exclusive chemical characteristics of the plant.

In addition, to systematically evaluate discriminative ability and develop robust prediction models, chemometric techniques is usually used to analyze the data. Commonly used methods include random forests (RF), orthogonal partial least squares discriminant analysis (OPLS-DA), and linear discriminant analysis (LDA). These techniques demonstrate efficacy in elemental fingerprint-based classification, with LDA achieving a high degree of interpretability through discriminant analysis and has demonstrated high discriminant accuracy in traceability studies of agricultural products such as cherries, potatoes and rice, effectively proving the applicability of this method to plant-based products[25–27]. The RF method effectively improves classification stability by integrating and aggregating decision trees. It has been successfully applied to the geographical origin identification of green tea and tobacco, strongly confirming the significant advantages of combining elemental characteristics with machine learning in origin tracing[28,29]. Orthogonal partial least squares discriminant analysis (OPLS-DA) is a supervised discriminant method that can handle multi-dimensional variable data. Chung et al. found that the geographical classification error rate of OPLS-DA model for unknown samples was less than 1.2%, which verified its effectiveness and reliability as a tool for identifying the origin of dried mushrooms[30].

In order to determine the origin of *Zanthoxylum bungeanum* in Gansu province, stable isotopes ($\delta^{13}C$ and $\delta^{15}N$) and 43 multi-element variables were used as indicators, and linear discriminant analysis and random forest model were combined to screen for the best discriminant indicators. The primary goal of this study is to establish a traceability method of the geographic indication *Zanthoxylum bungeanum* in Gansu province to provide a reference for the establishment of the traceability system for *Zanthoxylum bungeanum*.

2. Results

2.1 Stable isotope ratios and multi-element contents of *Zanthoxylum bungeanum*

Table 1 displays the $\delta^{15}N$ and $\delta^{13}C$ isotope ratios and 43 multi-element contents of *Zanthoxylum bungeanum* samples from Gansu GI, Gansu non-GI and other provinces. The mean $\delta^{13}C$ values of 420 samples are generally similar in the rang from − 27.98‰ to − 26.18‰, the result showed that Gansu GI samples were significantly higher than those from Gansu non-GI and other provinces ($p < 0.05$). The production of Gansu GI products is primarily concentrated in the southeastern region of Gansu province, which is characterized by a semi-arid climate and limited annual precipitation. This arid environment may be one of the reasons for the increase in $\delta^{13}C$ values of *Zanthoxylum bungeanum* in the process of long-term natural selection or artificial breeding. Conversely, there were no significant differences in mean $\delta^{15}N$ values ($p > 0.05$). It is hypothesized that the $\delta^{15}N$ value is closely linked to soil nutrient availability, nitrogenous fertilizers and local agricultural practices[31], rather than directly reflecting the geographical origin of plants.

Table 1
Stable isotope ratios and multi-element content in *Zanthoxylum bungeanum* from different geographic origins (mg/kg).

Variables	Elements	Gansu (GI)	Gansu (non-GI)	Other provinces
Isotope ratios	$\delta^{15}N$(‰)	−0.11 ± 1.37	−0.11 ± 1.50	−0.14 ± 1.56
	$\delta^{13}C$(‰)	-26.18 ± 0.55^{a}	-26.52 ± 0.34^{b}	-27.98 ± 0.07^{b}
Macro-elements	K	16076.65 ± 3008.47	15481.50 ± 3224.03	15848.57 ± 4372.22
	Ca	9184.03 ± 1647.33^{b}	9654.65 ± 1598.01^{a}	9840.63 ± 3148.82^{a}
	Mg	2102.01 ± 565.82^{A}	2043.64 ± 425.85^{A}	1503.54 ± 435.14^{B}
	Al	112.16 ± 47.39	100.35 ± 42.43	98.35 ± 70.40
	P	2917.83 ± 482.03	2902.29 ± 651.85	3007.93 ± 898.81
	S	1248.33 ± 154.07	1283.45 ± 166.26	1295.03 ± 243.04
Micro-elements	B	39.38 ± 9.02^{A}	$37.22 \pm 7.28A^{B}$	34.19 ± 10.43^{B}
	Ti	4.70 ± 3.37^{B}	3.98 ± 2.50^{B}	9.63 ± 4.80^{A}
	V	0.32 ± 0.16	0.28 ± 0.16	0.31 ± 0.16
	Mn	31.00 ± 8.43^{b}	25.78 ± 6.23^{b}	45.15 ± 3.79^{a}
	Fe	122.38 ± 57.17^{A}	115.09 ± 53.60^{AB}	104.02 ± 57.38^{B}
	Co	0.11 ± 0.025^{B}	0.099 ± 0.031^{B}	0.13 ± 0.082^{A}
	Ni	1.96 ± 0.69^{a}	1.45 ± 0.56^{b}	1.80 ± 0.95^{ab}
	Cu	6.37 ± 2.02^{A}	5.49 ± 1.09^{B}	5.87 ± 2.39^{AB}
	Zn	12.10 ± 2.27^{B}	11.60 ± 1.91^{B}	15.33 ± 5.21^{A}
	As	0.076 ± 0.039	0.097 ± 0.014	0.078 ± 0.053
	Rb	18.51 ± 7.52	12.68 ± 4.76	17.22 ± 9.26
	Sr	112.21 ± 52.31^{A}	88.40 ± 38.28^{B}	35.42 ± 28.85^{B}
	Zr	0.099 ± 0.038	0.10 ± 0.052	0.11 ± 0.053

Note: The data in the table are presented as the mean ± standard deviation; different capital letters in the same column indicate highly significant differences ($p < 0.01$), and different lowercase letters indicate significant differences ($p < 0.05$).

Variables	Elements	Gansu (GI)	Gansu (non-GI)	Other provinces
	Nb	0.018 ± 0.0084	0.018 ± 0.0098	0.018 ± 0.010
	Mo	0.49 ± 0.21^{AB}	0.53 ± 0.31^{A}	0.40 ± 0.36^{B}
	Pd	0.17 ± 0.077^{A}	0.14 ± 0.064^{B}	0.068 ± 0.049^{C}
	Ag	0.0026 ± 0.0015^{B}	0.0027 ± 0.0032^{B}	0.0043 ± 0.0019^{A}
	Sn	2.19 ± 0.46^{B}	1.63 ± 0.73^{C}	2.53 ± 0.83^{A}
	Te	0.0024 ± 0.0015^{B}	0.0021 ± 0.0015^{B}	0.0031 ± 0.0028^{A}
	Cs	0.11 ± 0.035^{A}	0.091 ± 0.035^{B}	0.067 ± 0.032^{C}
	Ba	10.08 ± 2.84	9.05 ± 3.09	10.75 ± 7.72
	Tl	0.012 ± 0.007^{A}	0.0093 ± 0.004^{B}	0.0089 ± 0.0048^{B}
	Pb	0.19 ± 0.08^{B}	0.21 ± 0.09^{B}	0.49 ± 0.35^{A}
	U	0.012 ± 0.0056^{A}	0.0119 ± 0.0060^{A}	0.0083 ± 0.0062^{B}
Trace-elements	La	0.26 ± 0.047	0.21 ± 0.14	0.26 ± 0.033
	Ce	0.48 ± 0.35	0.39 ± 0.26	0.37 ± 0.13
	Pr	0.046 ± 0.021	0.048 ± 0.03	0.047 ± 0.025
	Nd	0.17 ± 0.075	0.17 ± 0.11	0.18 ± 0.09
	Sm	0.032 ± 0.014	0.031 ± 0.017	0.029 ± 0.007
	Eu	0.011 ± 0.003	0.011 ± 0.005	0.012 ± 0.0069
	Gd	0.033 ± 0.015	0.032 ± 0.018	0.030 ± 0.019
	Dy	0.018 ± 0.008	0.018 ± 0.009	0.016 ± 0.003
	Ho	0.0037 ± 0.0015^{A}	0.0035 ± 0.0019^{AB}	0.0030 ± 0.0024^{B}
	Er	0.011 ± 0.0046^{A}	0.010 ± 0.0056^{A}	0.0081 ± 0.0063^{B}
	Tm	0.0014 ± 0.0006^{A}	0.0013 ± 0.00073^{A}	0.0010 ± 0.0008^{B}
	Yb	0.0071 ± 0.0030^{A}	0.0067 ± 0.0035^{AB}	0.0057 ± 0.0044^{B}
	Y	0.087 ± 0.036^{A}	0.081 ± 0.043^{B}	0.081 ± 0.079^{C}

Note: The data in the table are presented as the mean ± standard deviation; different capital letters in the same column indicate highly significant differences ($p < 0.01$), and different lowercase letters indicate significant differences ($p < 0.05$).

In *Zanthoxylum bungeanum* samples, the content of K exhibited the highest abundance among macro-elements, ranging from 15835 to 16076 mg/kg. Al exhibited the lowest concentration among macro-elements, ranging from 98.35 to 112.16 mg/kg. Among trace elements, Fe demonstrated the highest content, ranging from 104.02 to 122.38 mg/kg. For rare earth elements, Ce displayed the highest concentration while Tm exhibited the lowest concentration, consistent with the findings reported by Ma et al[32]. There were statistical differences in the contents of 26 elements (including $\delta^{13}C$ and 25 elements) among the three sample groups ($p < 0.05$). Among these significant elements, the contents of B, Sr, Pd, Tl, and Cs in Gansu GI samples were significantly higher than those in Gansu non-GI samples and other provinces ($p < 0.05$), forming a unique fingerprint. Among trace elements, samples from other provinces had higher levels of Ti, Mn, Co, Zn, Sn, Pb, Ag, and Te ($p < 0.05$), while among the rare earth elements, Ho, Er, Yb, Y, and Tm were the lowest in other provinces ($p < 0.01$). The heat map generated by Pearson's hierarchical clustering analysis revealed that $\delta^{13}C$ was positively correlated with Mg, Fe, B, Sr, Pd, and Cs (Fig. 1). Certain rare earth elements (e.g., Ho, Er, Yb, Tm) exhibited strong positive correlations with each other ($r > 0.8$), indicating that these rare earth elements were controlled by a common geological background.

2.2 Multivariate modeling for the origin traceability of *Zanthoxylum bungeanum*

Given that LDA, OPLS-DA, and RF have been extensively validated as robust analytical methodologies for the authentication and traceability of plant-derived products, this study aims to conduct a systematic comparative evaluation of the performance metrics of these three modeling approaches to identify the optimal analytical framework. The proposed model system is designed to achieve two primary objectives: (1) discriminating *Zanthoxylum bungeanum* samples originating from Gansu province GI regions from non-GI counterparts; and (2) differentiating *Zanthoxylum bungeanum* produced in four GI producing counties (Wudu, Qinan, Gangu, and Qingshui Counties) in Gansu province.

2.2.1 Development and validation of a random forest classification model for *Zanthoxylum bungeanum*

In this study, 70% of the samples served as training datasets to establish models. The remaining 30% of samples formed the testing set, utilized to evaluate the prediction performance of the training models. Confusion matrix thermogram shows the classification effect of random forest classification model in various producing areas (Fig. 2). Figure 2a shows the classification results of the test data of *Zanthoxylum bungeanum* from Gansu GI, Gansu non-GI, and other provinces, where most samples were accurately classified. Forty-seven samples in Gansu GI area were accurately classified, of which 4 samples were misclassified as Gansu non-GI, indicating that the model has strong recognition ability for this area. The overall accuracy of origin discrimination was above 90%.

The classification results of the test sets of four GI producing areas in Gansu are shown in Fig. 2b, and most of the samples in the test set were accurately classified, among which the classification effect in

Qingshui area was poor, and only 6 of 11 samples were accurately classified. The classification results for the Wudu region were the best, with 14 of the 16 samples correctly classified and only 2 misclassified as Qinan. The importance of variables indicates the contribution ratio of each variable in the construction of RF model (Fig. 2c). Elements with a random forest importance score above a threshold of 5 were selected. The RF importance plots of variables indicate that Pb, Co and Ni were important classification variables of *Zanthoxylum bungeanum* in Gansu GI.

2.2.2 Development and validation of orthogonal partial least squares discriminant analysis model for *Zanthoxylum bungeanum*

As shown in Fig. 3, the OPLS-DA model can clearly distinguish sample groups from different geographical sources. In the results of inter-provincial scale classification (Fig. 3a), samples from Gansu GI, Gansu non-GI and other provinces exhibit an obvious aggregation trend with some overlap. The OPLS-DA classification results showed that the GI regions in Gansu overlap substantially and the separation trend is not obvious (Fig. 3b). In detail, the plot can be subdivided into three regions: samples from Gangu and Qinan are located on the left, samples from Wudu are located on the right, samples from Qingshui are located on the upper center. Combined with Fig. 5, this suggests that classification may be based on geographical proximity which places them under similar environmental conditions, so that the multi-element profiles of *Zanthoxylum bungeanum* samples from two neighboring regions are close to each other. The Variable Importance in Projection (VIP) value is commonly employed to quantify the contribution of individual variables to the model's predictive performance. As illustrated in the VIP plot (Fig. 3c), the variables exhibiting the highest discriminatory power for distinguishing the geographical origin of Gansu GI samples (VIP > 1.5) include Sr, Pd, and Mg.

2.2.3 Development and validation of a linear discriminant analysis model for *Zanthoxylum bungeanum*

Figure 4a shows discriminant analysis results samples based on Gansu GI, Gansu non-GI and other provinces *Zanthoxylum bungeanum*. The figure demonstrates a relatively distinct separation between samples from Gansu province and those from other provinces, with only minor overlapping regions observed between Gansu GI and non-GI samples. A further analysis of the geographical indication (GI) production areas in Gansu revealed that samples from four GI regions (Wudu, Gangu, Qingshui, and Qinan) each formed distinct clusters (Fig. 4b). The samples collected from Wudu demonstrated the minimal degree of overlap with those from the other three regions, which is consistent with their geographical distribution, as Wudu is comparatively more geographically isolated from the remaining three areas. This spatial separation phenomenon strongly validates the effectiveness of the linear discriminant analysis (LDA) model in distinguishing key production regions. The significance of the selected elements in the LDA model is evaluated based on the Area Under the Curve (AUC) values derived from the Receiver Operating Characteristic (ROC) analysis. As illustrated in the ROC and AUC plot of the variables (Fig. 4c), Sr, Pd, Cs, and $\delta^{13}C$ emerge as the primary contributing variables in the LDA model. These variables exhibit AUC values exceeding 0.8, thereby qualifying them as key discriminative indicators for distinguishing *Zanthoxylum bungeanum* originating from the Gansu geographical indication (GI) region.

2.3. Machine learning model comparison

To evaluate the predictability and stability of the classification models, the discrimination accuracy of the LDA, OPLS-DA, and RF traceability models was verified. Table 2 presents the classification results of *Zanthoxylum bungeanum* from Gansu GI, Gansu non-GI, and other province regions using LDA, OPLS-DA, and RF. LDA has the highest discrimination accuracy for Gansu GI samples, with an accuracy rate of 89.3%, which is significantly better than OPLS-DA and RF models. Additionally, in the discriminant analysis of four GI-producing areas in Gansu province, LDA maintained a discrimination accuracy exceeding 80%. In contrast, OPLS-DA exhibited a markedly low accuracy of only 22.5% in distinguishing Qinan production area, which RF achieved accuracies of 37.8% for Wudu and 50.0% for Qingshui (Table 3). Samples from Qinan and Gangu were misidentified, likely due to their shared geographical characteristics as both are located within Tianshui City and situated in the Weihe River Valley. These areas exhibit similar environmental conditions, including a temperate semi-humid climate and loess plateau topography. Additionally, the two locations are in closer proximity to each other compared to the other two production regions, with the shortest straight-line distance among them. In contrast, Wudu is geographically isolated from the other three areas by the Qinling Mountain range, creating a distinct spatial separation. This spatial contiguity between Qinan and Gangu contributes to the classification confusion observed in *Zanthoxylum bungeanum* samples from these two producing areas. Overall, LDA is significantly superior to both the OPLS-DA and RF models.

Table 2
Classification of *Zanthoxylum bungeanum* from Gansu GI, Gansu non-GI, and other provinces regions using LDA, OPLS-DA, and RF.

Modeling	Regions	Accuracy (%)	Fredicated category		
			Gansu GI	Gansu non-GI	Other provinces
LDA	Gansu GI	89.3	134	15	1
	Gansu non-GI	80.0	13	68	4
	Other provinces	89.2	8	12	165
OPLS-DA	Gansu GI	84.7	127	1	22
	Gansu non-GI	23.8	54	20	10
	Other provinces	92.61	12	1	163
RF	Gansu GI	81.3	122	17	11
	Gansu non-GI	57.1	0	12	9
	Other provinces	70.3	43	12	130

Table 3
Classification of *Zanthoxylum bungeanum* from Gansu GI regions in province using LDA, OPLS-DA, and RF.

Modeling	Regions	Accuracy (%)	Fredicated category			
			Wudu	Qinan	Gangu	Qingshui
LDA	Wudu	82.2	44	0	0	1
	Qinan	87.5	0	39	1	0
	Gangu	85.7	0	1	33	1
	Qingshui	86.7	1	0	1	28
OPLS-DA	Wudu	84.4	38	0	2	5
	Qinan	22.5	11	9	14	6
	Gangu	68.6	6	5	24	0
	Qingshui	66.7	6	3	1	20
RF	Wudu	37.8	17	17	11	0
	Qinan	57.5	0	23	9	8
	Gangu	85.7	1	2	30	2
	Qingshui	50.0	0	12	3	15

3. Conclusion

This study successfully developed a geographical origin model for *Zanthoxylum bungeanum* products in Gansu province through the analysis of stable isotopes ($\delta^{15}N$ and $\delta^{13}C$), 43 multi-element fingerprinting characteristics, and the application of chemometric methods. The $\delta^{13}C$ values of samples from Gansu's geographical indication regions were significantly higher than those from non-geographical indication regions in Gansu and other provinces. This finding establishes $\delta^{13}C$ as a reliable indicator for distinguishing geographical indication products. Furthermore, analysis of variance revealed significant differences in the concentrations of 26 elements across different regions, highlighting a distinct geochemical fingerprint associated with the Gansu GI samples. Specifically, the concentrations of B, Sr, Pd, Tl and Cs were significantly higher in Gansu GI samples, whereas elements such as Ti, Mn, Co, and Zn were found to be more abundant in non-GI samples. Additionally, a strong positive correlation among certain rare earth elements (Ho, Er, Yb, and Tm) was observed, underscoring the regional specificity of the elemental fingerprint. RF, OPLS-DA, and LDA were employed to classify the *Zanthoxylum bungeanum* samples and develop multivariate models. An analysis of model accuracy showed that the LDA model surpassed both RF and OPLS-DA in its ability to discriminate, effectively differentiating Gansu GI *Zanthoxylum bungeanum* from non-GI samples and pinpointing the four specific GI counties in Gansu. On the other hand, misclassification rates were higher for RF and OPLS-DA, especially for spatially close locations. In order to protect brand integrity and prevent fraudulent labeling in the spice market, this study offers a scientific tracing tool for validating Gansu GI *Zanthoxylum bungeanum*.

4. Methods

4.1 Sample collection and preparation

This study involved the collection of 420 *Zanthoxylum bungeanum* Maxim. samples from nine provinces in China, including 150 GI samples from Gansu (Wudu, Qinan, Gangu, and Qingshui Counties) and 85 non-GI samples from Gansu (Xihe, Li, Tanchang, and Linxia Counties). In addition, another 185 samples were collected from eight different Chinese provinces: Shanxi, Sichuan, Shandong, Yunnan, Hebei, Shaanxi, Henan, and Guizhou, considering the national distribution of *Zanthoxylum bungeanum* production. To minimize the impact of seasonal variations on the elemental composition, all samples were collected between August and September 2023. The collected samples were subjected to natural air-drying, followed by the removal of seeds and impurities. Finally, they were heated and dried at 80°C for 4 hours.

Afterward, the dry samples were milled to a fine powder using a pulverizing machine, passed through a 60-mesh sieve, then stored at 4°C for measurement. Detailed sampling information is provided in Fig. 5.

4.2 Chemicals and instrumentation

Nitric acid (HNO_3, GR, 65%) and hydrogen peroxide (H_2O_2, GR, 30%) were purchased from Sinopharm Chemical Reagent Factory (China) for sample digestion. Multi-element standard solutions of 100 mg L^{-1} (containing 65 elements) were purchased from the National Analysis Center of Iron and Steel (NCS, Beijing, China). 10 μg mL^{-1} rare earth element standard solution was purchased from the National Iron and Steel Materials Testing Center (Beijing, China). A certified reference material (Garlic powder, GBW10022, Institute of Geophysical and Geochemical Exploration, Tianjin, China) was used to verify the accuracy of the results of element quantification. Isotopic reference material including USGS40 ($\delta^{15}N$ = − 4.50‰ vs Air and $\delta^{13}C$ = − 26.38‰ vs V-PDB, Thermo Fisher Scientific, USA) was used to verify the accuracy of the isotopic results. Ultrapure water (18.2 MΩ cm^{-1}) was obtained from Genie U12e water purification system (RephiLe Bioscience, China).

Isotope analysis was performed using an automatic stable isotope mass spectrometry (IRMS, Integra 2, SerCon20, Britain). Microwave digestion system (TOPEX+, Preekem, Shanghai) was used to decompose the sample. Inductively coupled plasma tandem mass spectrometers (ICP-MS/MS, Agilent 8900, Agilent, USA) were used for the elemental content analysis. The instrument working conditions are shown in Table S1.

4.3 Stable isotope ratios analysis

About 5.0 mg sample was weighed into a tin capsule for $\delta^{15}N$ and $\delta^{13}C$ analysis. Then the capsule was closed with tweezers and placed into the automatic sampler of the elemental analyzer. In the C/N model, the temperatures of oxidation furnace and reduction furnace were set to 1000°C and 600°C respectively. The carrier gas was helium (purity > 99.9999%), and the flow rate is 230 ml min^{-1}. Stable isotope ratio was calculated according to Eq. (1) referring to the literature published by Coplen[33]:

$$\delta\,(‰) = \left(\frac{R_{sample}}{R_{standard}}\text{-}1\right)\times 1000$$

1

where δ represents either $\delta^{13}C$, $\delta^{15}N$; R_{sample} denotes the abundance ratio of heavy isotope against light isotope, e.g., $^{13}C/^{12}C$, $^{15}N/^{14}N$; $R_{standard}$ is the isotope ratio of reference standard versus primary standards, Pee Dee Belemnite (V-PDB) for carbon, Air gas (AIR) for nitrogen. Each sample was analyzed in triplicate and the average was calculated. The analytical precision of the above method was evaluated to be better than ± 0.2‰ for C and N.

4.4 Multi-element analysis

Approximately 0.2 g of sample was transferred to a digestion vessel, and then 3 mL HNO_3 and 2 mL H_2O_2 were added. Then the capped vessels were placed in the microwave digestion system and digested. The digestion temperature was raised from room temperature to 130°C within 10 min and maintained for 5 min. Then the temperature was ramped to 200°C within 10 min and maintained for 30 min. After cooling, the digests were subsequently diluted to 25.0 mL with deionized water for ICP-MS/MS analysis. The measurements of all samples and standard reference materials were performed in triplicate.

According to the concentration levels observed in *Zanthoxylum bungeanum* samples, the established linear ranges were found to be 0.01 ~ 2 μg L^{-1} for micro-elements, 5 ~ 500 μg L^{-1} for trace-elements, 0.5 ~ 50 mg L^{-1} for macro-elements, respectively. The limits of detection (LODs) and the limits of quantitation (LOQs) of ICP-MS analysis were calculated and summarized in the supporting information Table S2. The determination results of some elements were below the detection limit. In such cases, a value of 1/2 times the LOD was assigned for statistical analysis to ensure data integrity[34].

Accuracy validation of the ICP-MS/MS procedure was made by analyzing certified reference material (CRM) of Garlic powder GBW10022 (Table. S2). Since the concentrations of Pr, Ho, and Tm in the CRM were below the LOQ, their recoveries were correspondingly below 80% (77.6%, 76.9%, and 78.5%, respectively). The recovery rates for most elements in the certified reference material (CRM) ranged from 80.4% to 102%, indicating the reliability of the analytical method.

4.5 Statistical analysis and software

Mathematical statistics with chemometric methods was crucial for utilizing multivariate data of origin traceability. One-way analysis of variance (ANOVA) was used to preliminarily evaluate whether there were significant differences in element variables in *Zanthoxylum bungeanum* samples. A $p < 0.05$ value was regarded as significant. The classification modeling of *Zanthoxylum bungeanum* samples from different regions was performed using random forest (RF), orthogonal partial least squares discriminant analysis (OPLS-DA), and linear discriminant analysis (LDA), and to exclude overfitting, the accuracy of the final models was verified by leave-one-out method. To ensure the intuitiveness of the sampling map, ArcGIS 9.0 was used for geographic information processing and visualization. ANOVA, LDA modeling, calculations of receiver operating characteristic curves (ROC) and area under the curve (AUC) of variables were performed using SPSS software (26.0, IBM, USA). OPLS-DA modeling was conducted on SIMCA-p software (14.1, Umetrics, Sweden). RF modeling was implemented using R (Version 4.1.0). The element correlation analysis was completed by chiplot software (website address: https://www.chiplot.online/).

Declarations

Competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Contribution

C.Q.C.: Writing-original draft, Conceptualization, Data Curation; Software; R.H.M.,: Conceptualization, Methodology, Validation, Resources; F.B.T.: Writing-review & editing, Conceptualization, Project administration, Resources; Y.H. L.: Writing-review & editing, Conceptualization, Methodology; L.Y.: Investigation, Resources; W.T.L.: Writing-review & editing, Methodology; M.H.Q.: Methodology, Validation, Investigation; Z.L.N.: Writing-review & editing, Conceptualization, Data curation.

Acknowledgement

This study was supported by the Fundamental Research Funds for the Central Nonprofit Research of Chinese Academy of Forestry (CAFYBB2025QB002).

Data Availability

All data generated or analysed during this study are included in this published article and its Supplementary Information files.

References

1. Cui, C. *et al.* ^{1}H NMR-based metabolomic approach combined with machine learning algorithm to distinguish the geographic origin of huajiao (*Zanthoxylum bungeanum* Maxim.). *Food Control* 145 (2023a). https://doi.org/10.1016/j.foodcont.2022.109476.
2. Liu, S. *et al*. Characteristic differences in essential oil composition of six *Zanthoxylum bungeanum* Maxim. (Rutaceae) cultivars and their biological significance. *Journal of Zhejiang University-Science B* 18, 917–920 (2017). https://doi.org/10.1631/jzus.B1700232.
3. Ji, Y., Li, S. & Ho, C. Chemical composition, sensory properties and application of Sichuan pepper (*Zanthoxylum* genus). *Food Science and Human Wellness* 8, 115–125 (2019). https://doi.org/10.1016/j.fshw.2019.03.008.
4. Ke, J. *et al.* Application of HPLC fingerprint based on acid amide components in Chinese prickly ash (*Zanthoxylum*). *Industrial Crops and Products* 119, 267–276 (2018). https://doi.org/10.1016/j.indcrop.2018.04.018.
5. Liang, W. *et al.* Phytochemistry and health functions of *Zanthoxylum bungeanum* Maxim and *Zanthoxylum schinifolium* Sieb. et Zucc as pharma-foods: A systematic review. *Trends in Food Science & Technology* 143 (2024). https://doi.org/10.1016/j.tifs.2023.104225.
6. Zhang, W. *et al.* Comparison of volatile components in fresh and dried *Zanthoxylum bungeanum* Maxim. *Food Science and Biotechnology* 28, 1083–1092 (2019b). https://doi.org/10.1007/s10068-019-00562-3.
7. Mamede, R. *et al.* Spatial variability of elemental fingerprints of sea lettuce (*Ulva* spp.) and its potential use to trace geographic origin. *Algal Research-Biomass Biofuels and Bioproducts* 59 (2021). https://doi.org/10.1016/j.algal.2021.102451.
8. Wang, J., Wang, S., Ge, X. & Zhang, M. Authentication of American ginseng (*Panax quinquefolius* L.) from different origins by linear discriminant analysis of multi-elements. *European Food Research and Technology* 247, 2657–2666 (2021). https://doi.org/10.1007/s00217-021-03816-9.
9. Camin, F. *et al.* Stable isotope techniques for verifying the declared geographical origin of food in legal cases. *Trends in Food Science & Technology* 61, 176–187 (2017). https://doi.org/10.1016/j.tifs.2016.12.007.
10. Suzuki, Y. Achieving food authenticity and traceability using an analytical method focusing on stable isotope analysis. *Analytical Sciences* 37, 189–199 (2021). https://doi.org/10.2116/analsci.20SAR14.
11. Qie, M., Zhang, B., Li, Z., Zhao, S. & Zhao, Y. Data fusion by ratio modulation of stable isotope, multi-element, and fatty acids to improve geographical traceability of lamb. *Food Control* 120 (2021). https://doi.org/10.1016/j.foodcont.2020.107549.
12. Zhang, J. *et al.* Origin identification of the sauce-flavor Chinese Baijiu by organic acids, trace elements, and the stable carbon isotope ratio. *Journal of Food Quality* 2019 (2019a). https://doi.org/10.1155/2019/7525201.

13. Ji, Q. *et al.* Protected geographical indication discrimination of Zhejiang and non-Zhejiang *Ophiopogonis japonicus* by Near-Infrared (NIR) spectroscopy combined with chemometrics: The influence of different stoichiometric and spectrogram pretreatment methods. *Molecules* 28 (2023). https://doi.org/10.3390/molecules28062803.
14. Li, Y., Zhang, J. & Wang, Y. FT-MIR and NIR spectral data fusion: a synergetic strategy for the geographical traceability of *Panax notoginseng*. *Analytical and Bioanalytical Chemistry* 410, 91–103 (2018). https://doi.org/10.1007/s00216-017-0692-0.
15. Fan, L. *et al.* Geographical-origin discrimination and volatile oil quantitative analysis of *Zanthoxylum bungeanum* Maxim. with a portable near-infrared spectrometer. *Analytical Methods* 11, 5301–5310 (2019). https://doi.org/10.1039/c9ay01520e.
16. Zhu, X. *et al.* Unveiling the flavor and quality variations in dried *Zanthoxylum bungeanum* maxim from China's diverse regions. *Food Chemistry-X* 25 (2025). https://doi.org/10.1016/j.fochx.2024.102017.
17. Chung, I. *et al.* C/N/O/S stable isotopic and chemometric analyses for determining the geographical origin of *Panax ginseng* cultivated in Korea. *Journal of Ginseng Research* 42, 485–495 (2018a). https://doi.org/10.1016/j.jgr.2017.06.001.
18. Wadood, S. A., Guo, B., Zhang, X., Hussain, I. & Wei, Y. Recent development in the application of analytical techniques for the traceability and authenticity of food of plant origin. *Microchemical Journal* 152 (2020). https://doi.org/10.1016/j.microc.2019.104295.
19. Cengiz, M. F. *et al.* Geographical origin of imported and domestic teas (*Camellia sinensis*) from Turkey as determined by stable isotope signatures. *International Journal of Food Properties* 20, 3234–3243 (2017). https://doi.org/10.1080/10942912.2017.1283327.
20. Wang, Y. *et al.* Stable isotope and multi-element profiling of Cassiae Semen tea combined with chemometrics for geographical discrimination. *Journal of Food Composition and Analysis* 107 (2022). https://doi.org/10.1016/j.jfca.2021.104359.
21. Zhang, X., Liu, Y., Li, Y. & Zhao, X. Identification of the geographical origins of sea cucumber (*Apostichopus japonicus*) in northern China by using stable isotope ratios and fatty acid profiles. *Food Chemistry* 218, 269–276 (2017). https://doi.org/10.1016/j.foodchem.2016.08.083.
22. Wadood, S. A. *et al.* Geographical origin classification of peanuts and processed fractions using stable isotopes. *Food Chemistry-X* 16 (2022). https://doi.org/10.1016/j.fochx.2022.100456.
23. Liu, X. *et al.* Isotope chemometrics determines farming methods and geographical origin of vegetables from Yangtze River Delta Region, China. *Food Chemistry* 342 (2021). https://doi.org/10.1016/j.foodchem.2020.128379.
24. Mottese, A. F. *et al.* Traceability of *Opuntia ficus-indica* L. Miller by ICP-MS multi-element profile and chemometric approach. *Journal of the Science of Food and Agriculture* 98, 198–204 (2018). https://doi.org/10.1002/jsfa.8456.
25. Chung, I. *et al.* Geographic authentication of Asian rice (*Oryza sativa* L.) using multi-elemental and stable isotopic data combined with multivariate analysis. *Food Chemistry* 240, 840–849 (2018b). https://doi.org/10.1016/j.foodchem.2017.08.023.

26. Rodriguez Galdon, B. *et al.* Differentiation of potato cultivars experimentally cultivated based on their chemical composition and by applying linear discriminant analysis. *Food Chemistry* 133, 1241–1248 (2012). https://doi.org/10.1016/j.foodchem.2011.10.016.
27. Wang, S. *et al.* Geographical origin traceability of sweet cherry (*Prunus avium (L.) Moench*) in China using stable isotope and multi-element analysis with multivariate modeling. *Food Chemistry-X* 23 (2024). https://doi.org/10.1016/j.fochx.2024.101477.
28. Cui, L. *et al.* Geographical origin classification of tobacco by stable isotope and multi-elemental analysis in combination with chemometric methods. *Microchemical Journal* 193 (2023b). https://doi.org/10.1016/j.microc.2023.109163.
29. Deng, X. *et al.* Predictive geographical authentication of green tea with protected designation of origin using a random forest model. *Food Control* 107 (2020). https://doi.org/10.1016/j.foodcont.2019.106807.
30. Chung, I. *et al.* Potential geo-discriminative tools to trace the origins of the dried slices of shiitake (*Lentinula edodes*) using stable isotope ratios and OPLS-DA. *Food Chemistry* 295, 505–513 (2019). https://doi.org/10.1016/j.foodchem.2019.05.143.
31. Wang, Z. *et al.* Linking growing conditions to stable isotope ratios and elemental compositions of Costa Rican bananas (*Musa* spp.). *Food Research International* 129 (2020). https://doi.org/10.1016/j.foodres.2019.108882.
32. Ma, Y., Wang, X., Huang, C., Tian, M. & Wei, A. Use of mineral element profiling coupled with chemometric analysis to distinguish *Zanthoxylum bungeanum* cultivars and health risks of potentially toxic elements in pericarps. *Journal of the Science of Food and Agriculture* 102, 1823–1831 (2022). https://doi.org/10.1002/jsfa.11517.
33. Coplen, T. B. Guidelines and recommended terms for expression of stable-isotope-ratio and gas-ratio measurement results. *Rapid Communications in Mass Spectrometry* 25, 2538–2560 (2011). https://doi.org/10.1002/rcm.5129.
34. Li, A. *et al.* The effect of urinary essential and non-essential elements on serum albumin: Evidence from a community-based study of the elderly in Beijing. *Frontiers in Nutrition* 9 (2022). https://doi.org/10.3389/fnut.2022.946245.

Figures

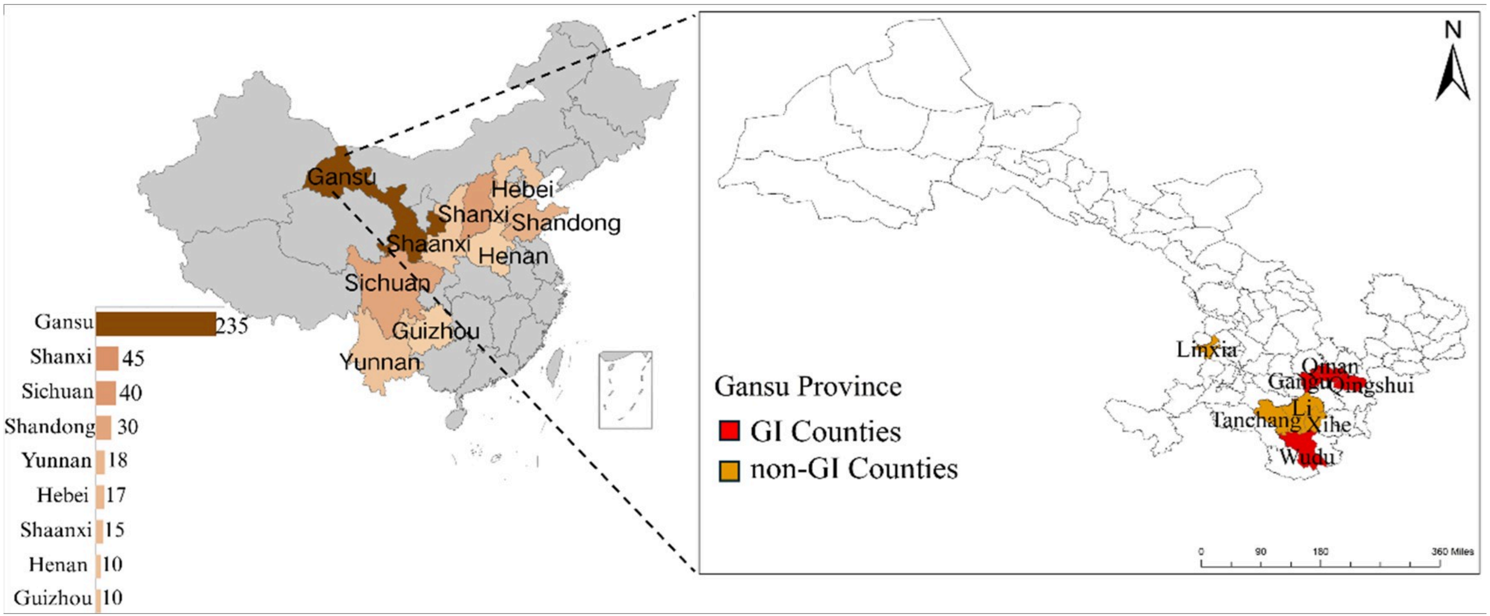

Figure 1

Heat map of elements correlation analysis.

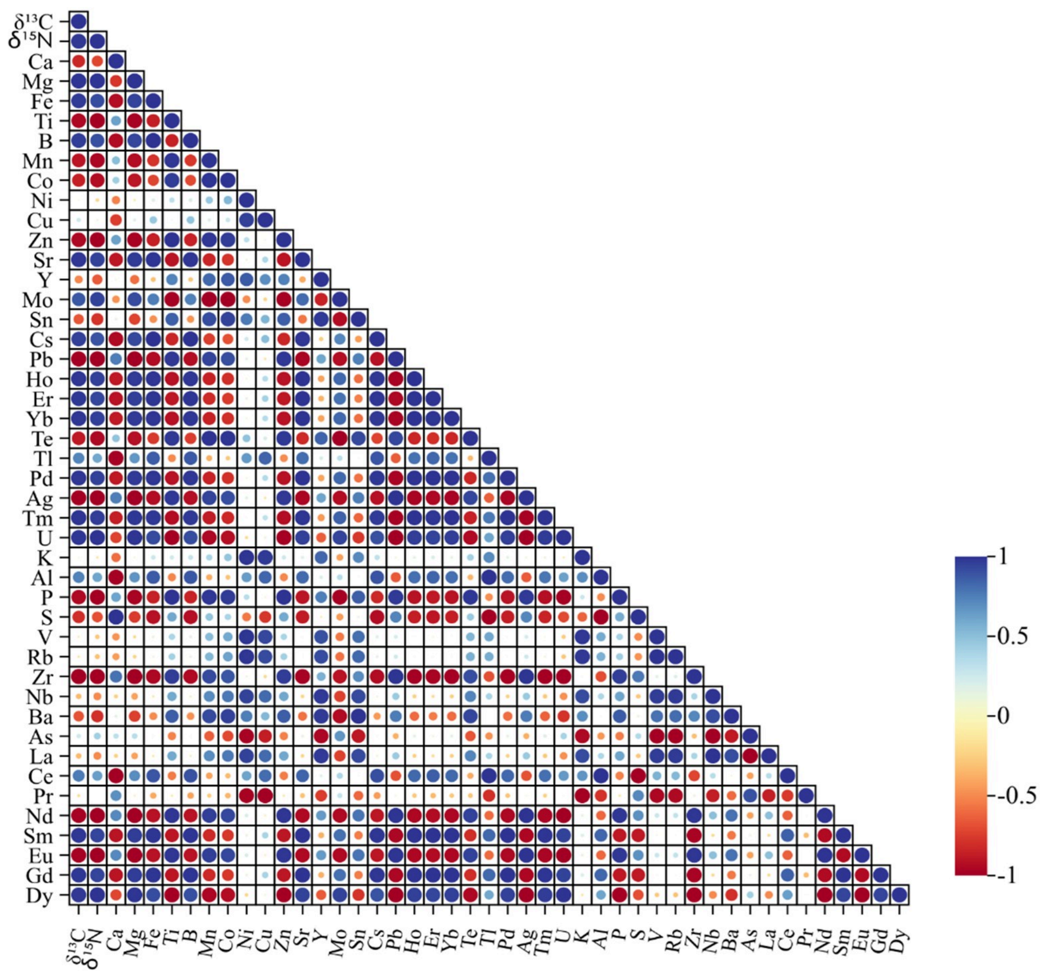

Figure 2

(a) RF model classification map of three types of *Zanthoxylum bungeanum*. (b) Classification map of RF model of Gansu GI areas. (c) RF variable importance plots of *Zanthoxylum bungeanum* samples.

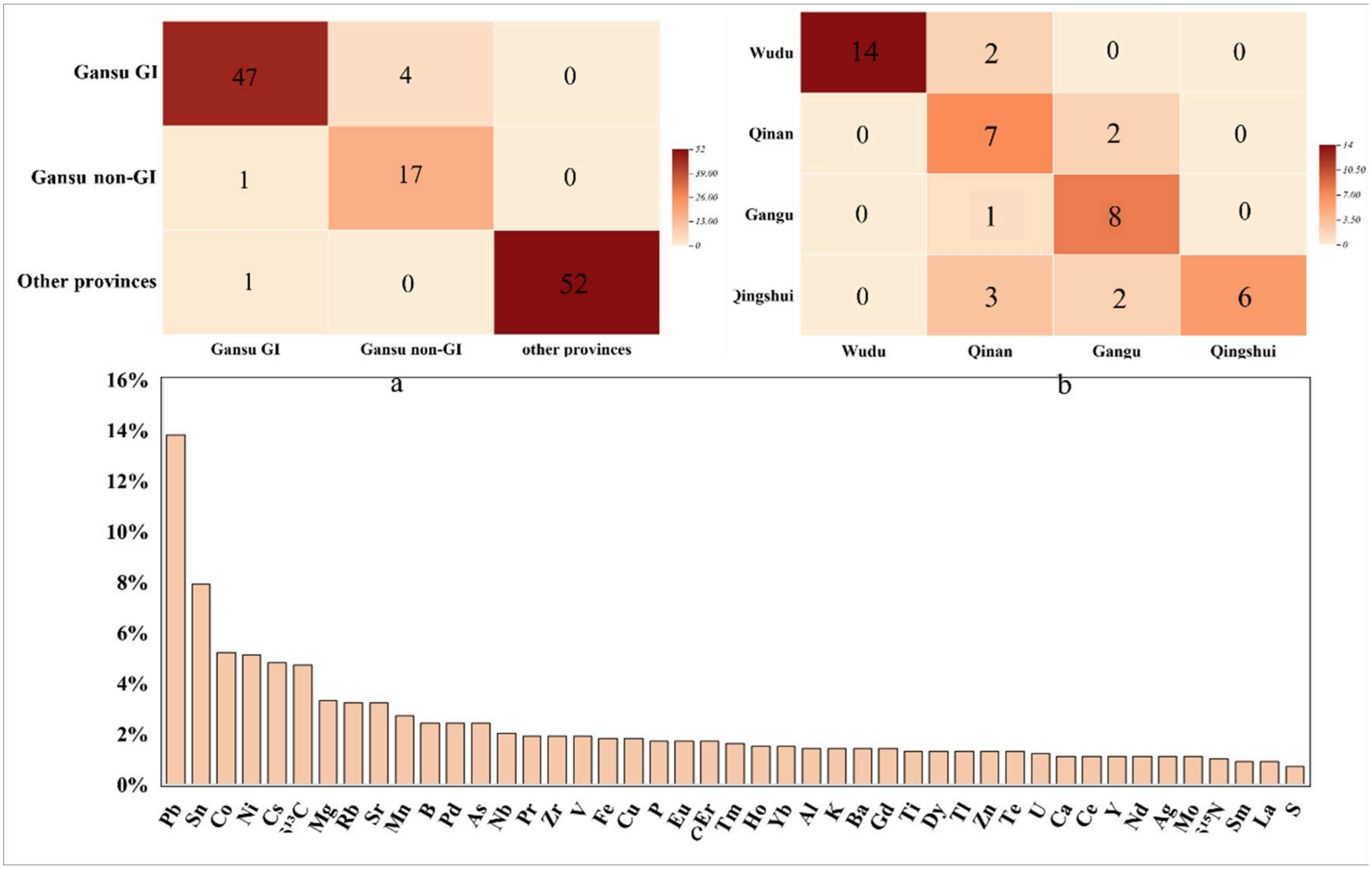

Figure 3

(a) Plots of OPLS-DA scores of three types of *Zanthoxylum bungeanum*. (b) Plots of OPLS-DA scores of Gansu GI areas. (c) VIP plot of OPLS-DA model variables.

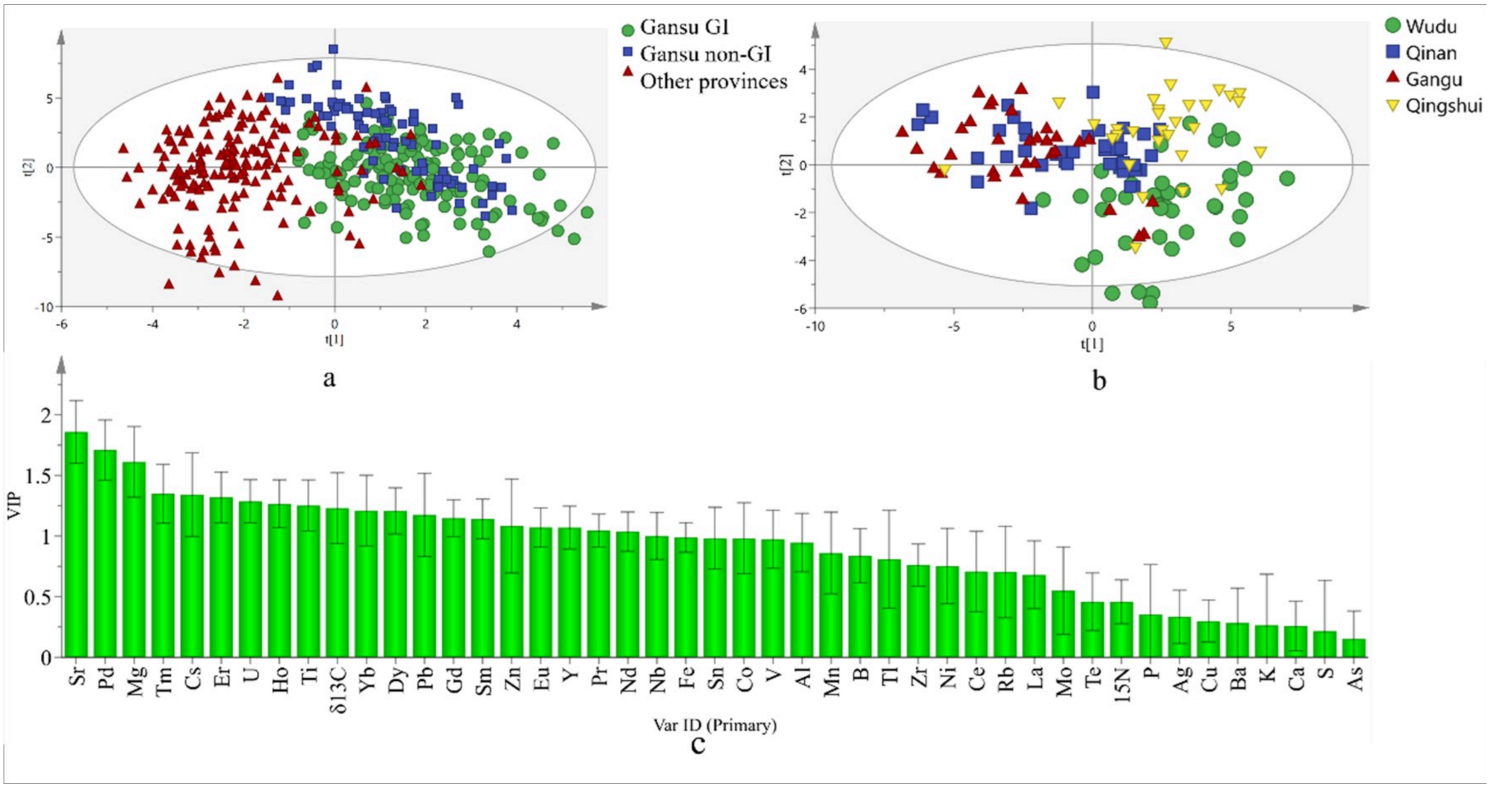

Figure 4

(a) Discriminant analysis of three types of *Zanthoxylum bungeanum*. (b) Discriminant analysis of Gansu GI samples. (c) ROC and AUC curve of variable in LDA model.

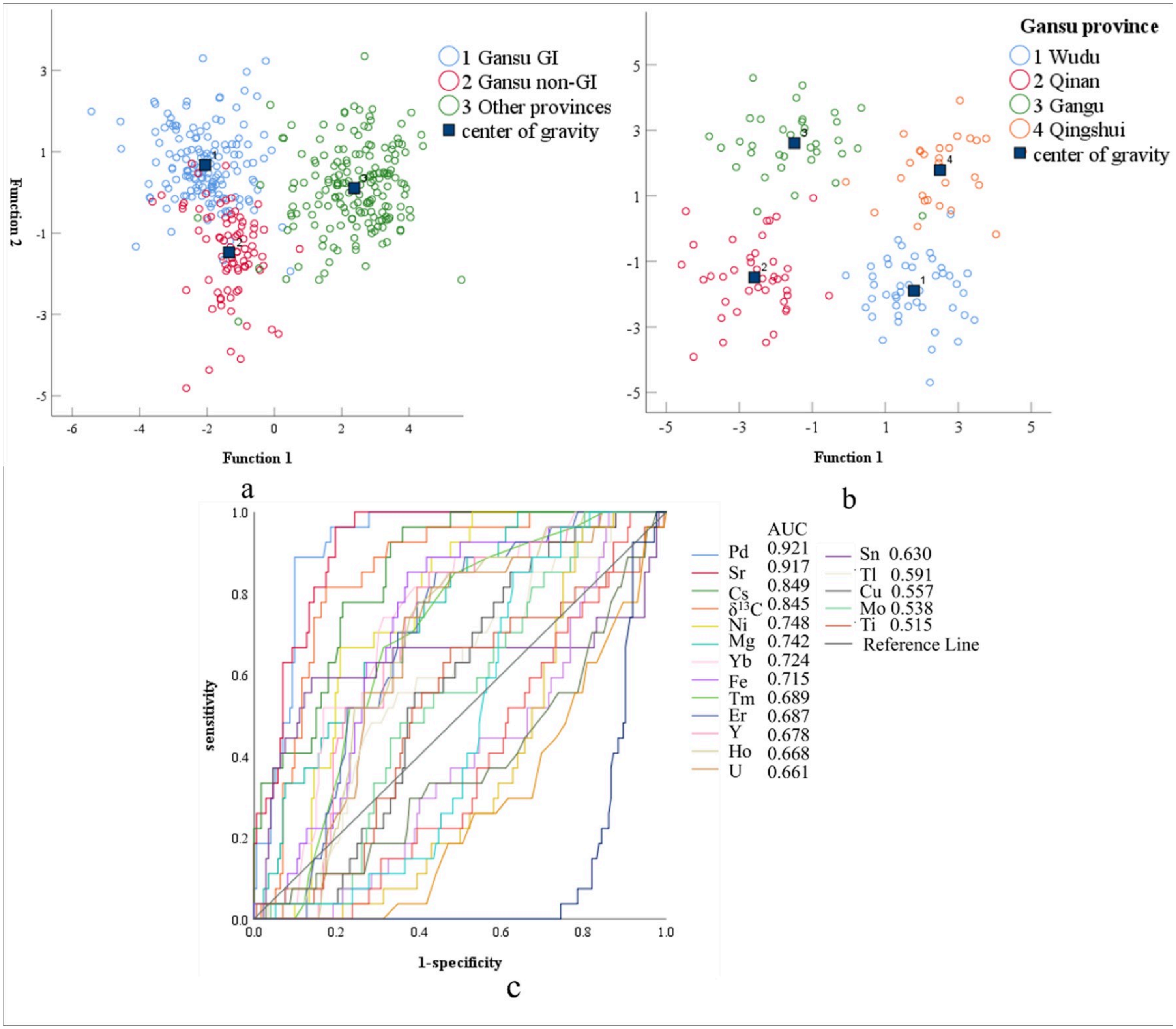

Figure 5

Spatial distribution of samples collected from nine provinces in China and the sampling diagram of different counties in Gansu province.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterials.docx