



Replacing traditional coffee appraisers with inductively coupled plasma - mass spectrometry (ICP-MS): From manual sensory evaluation to scientific analysis

Xiaowei Pan^{a,b,1}, Wenjing Yan^{b,1}, Xiaopeng Wu^d, Jianzhi Ye^a, Yaohui Liang^a, Guoyan Zhan^a, Hao Dong^c, Wenzhen Liao^b, Xingfen Yang^b, Qi He^{b,*}

^a Agricultural Product Processing Research Institute, Chinese Academy of Tropical Agricultural Sciences, Zhanjiang, Guangdong Province 524001, China

^b School of Public Health/Food Safety and Health Research Center/Guangdong Provincial Key Laboratory of Tropical Disease Research /BSL-3 Laboratory, Southern Medical University, Guangzhou, Guangdong Province 510515, China

^c College of Light Industry and Food Sciences, Zhongkai University of Agriculture and Engineering, Guangzhou, Guangdong Province 510225, China

^d Analysis and Testing Center, Chinese Academy of Tropical Agriculture Sciences, Haikou, Hainan Province 571101, China

ARTICLE INFO

Keywords:

Coffee
Geographic origin
Plasma mass spectrometry (ICP-MS)
Fisher discriminant analysis (FDA)
Principal component analysis

ABSTRACT

Scientific development provides opportunities to replace many traditional manual methods to achieve more accurate results and higher efficiency. To scientifically ascertain the geographical origin of coffee, this study develops a method for the rapid determination of 16 inorganic elements in coffee using microwave digestion combined with the ICP-MS internal standard method. Principal component analysis (PCA), Fisher discriminant analysis (FDA), and Partial least squares discriminant analysis (PLS-DA) are employed to analyze 40 coffee samples from three production areas of Pu'er, Baoshan, and Wanning. The results show that the linear correlation coefficients of the 16 elements in this method are above 0.999, the detection limits are in the range of 0.0004–0.63 mg/kg, the RSD of the precision experiments are 4.5 %–13.5 %, the recovery rate of the peak experiment is 86.0 %–96.3 % with the RSD of 1.1 %–8.8 %, and the results of the standard substances are within the range of standard values. Using the discriminant analysis of inorganic elements in coffee (FDA and PLS-DA), coffee origin discrimination was realized, and six key elements (Al, Mn, Fe, Cu, Na, and Ba) are identified as effective discriminatory indexes. Accordingly, a coffee origin discrimination model is established, and the overall accuracy discrimination rate of the discrimination model are all more than 90.0 %, and FDA > PLS-DA. The findings indicated that the method has good accuracy and reliability, is suitable for analyzing and determining multiple elements in sample components as targets, and may have a positive impact on the development of related industries.

1. Introduction

Coffee is one of the three major non-alcoholic beverages (cocoa, coffee, and tea) in the world, with a rich history dating back to the 9th century (Dos et al., 2024; Haile & Kang, 2019). Over the past millennium, coffee has captivated billions of enthusiasts due to its unique aroma, flavor, and recognized potential health benefits (Liu, Huang, Song, Li, & Mao, 2023). Coffee is cherished not only as a cup of fragrant and refreshing liquid, but rather abundant energy and a perfect working state, or even a symbol of comfortable and stable life (Van Dam, Hu, & Willett, 2020). Therefore, the related industries of coffee, including

cultivation, processing, grading, and distribution, are increasingly becoming a mainstay in many cultures and economies (Barrea et al., 2023; Velasquez & Banchon, 2023).

With the improvement of living standards and changes in consumption habits, more and more Chinese people are discovering and beginning to experience the allure of coffee. China has emerged as the most promising and rapidly expanding coffee market in the world. In the past year alone, nearly 100,000 new coffee shops have been established across China. Over 300,000 tons of coffee beans have been consumed, and the consumer market has surpassed a value of over 100 billion of US dollars. The preferences of this vast market are set to shape the

* Corresponding author.

E-mail address: heqi18@163.com (Q. He).

¹ Xiaowei Pan and Wenjing Yan are both first authors.

development of the coffee industry for a considerable period of time in the future. In 2020, three distinguished Chinese coffee varieties, including Baoshan small-grain coffee from Baoshan city in Yunnan province, Xinglong coffee from Wanning city in Hainan province, and Pu'er coffee from Pu'er city in Yunnan province were selected as the first batch of China-Europe protected geographical indications, representing the best quality, most popular, and most consumer-recognized coffee in China (Dong, Tan, Zhao, Hu, & Lu, 2015; Ma et al., 2022).

The meticulous and rigorous pursuit of coffee quality is paramount for growers, distributors, and consumers alike. However, for a long time, due to the limitations of scientific and technological level, the identification and quality assessment of various coffee products have primarily relied on the sensory evaluation of the flavor and taste led by well-trained coffee appraisers (Schaneberg, 2020). These appraisers use their exceptionally sensitive noses and tongues, honed through extensive and complex professional training, to make their judgments. Despite their expertise, the subjective nature of their assessments introduces a degree of bias. Additionally, the physical and sensory limitations of humans mean that appraisers can become fatigued or distracted, impacting the reliability of their judgments. Therefore, developing faster, more accurate, and simpler means to replace manual evaluations, and determine the difference in coffee products, has become an urgent and significant challenge in the current coffee industry.

When discussing the issue of food authenticity (especially in the case of coffee), it is important to use a variety of analytical techniques to add depth and breadth to the study. Common additional analytical techniques include gas chromatography–mass spectrometry (GC–MS), high-performance liquid chromatography (HPLC), and near infrared spectroscopy (NIR). These techniques can provide chemical information about the composition of food products and help identify samples from different sources. However, ICP-MS (Inductively Coupled Plasma Mass Spectrometry) has demonstrated unique advantages in elemental analysis with its ability to detect a wide range of elements, including trace elements, with great sensitivity and accuracy, providing a wealth of chemical information for research (Moor, Devos, Guecheva, & Kobler, 2000; Nguyen et al., 2023; Zhao et al., 2011).

The introduction of ICP-MS has injected novelty into the study of food authenticity, mainly in its ability to analyze multiple elements in a sample simultaneously and provide isotopic ratios of the elements. This data not only reveals the origin and processing history of the sample but also helps scientists to identify cases of forgery or adulteration. For example, by analyzing metallic elements in coffee, it is possible to identify the growing region and soil composition, thus verifying its authenticity.

Further, machine learning methods such as Principal Component Analysis (PCA), Partial Least Squares Discriminant Analysis (PLS-DA) and Linear Discriminant Analysis (LDA) can effectively process and model chemical data obtained from techniques such as ICP-MS. These methods can reveal underlying structures and patterns in the data and effectively differentiate between samples from different sources. By analyzing multiple samples of multidimensional data on elemental content, machine learning models are able to identify characteristic differences between authentic and counterfeit products, providing a scientific basis for authenticity testing.

In this study, the changing pattern of multi-element composition of coffee samples from three production areas, namely, Pu'er, Yunnan; Baoshan, Yunnan; and Wanning, Hainan, was investigated by utilizing ICP-MS instead of the traditional coffee characterization method. Based on the fingerprints of 16 inorganic elements, combined with machine learning methods such as PCA, PLS-DA, and LDA, the coffee samples from the three different production areas were characterized. The novelty of this study is the comparison of PLS-DA and LDA machine learning methods for geographic traceability of coffee identification with selected markers. This approach helps to screen the optimal machine learning model for coffee geographic traceability, as well as the most

effective chemical markers among inorganic elements for tracing the origin of coffee globally.

2. Materials & methods

2.1. Reagents

Analytical grade concentrated nitric acid (HNO₃, 65 %) and hydrogen peroxide (H₂O₂, 30 %) are sourced from Tianjin Komio Co, China. Tuning liquid Li, Co, In, and U (10 mg/L) are from o2si Co, US; Standard solution Al, Ba, Ca, Cd, Co, Cu, Cr, Na, K, Fe, Mg, Mn, Zn, Ni, Pb, Se, Sc, Ge, Rh, Re (1000 mg/L), from National Nonferrous Metals and Electronic Materials Analysis and Testing Center, China. CRMs including GBW10016a-Tea, GBW10021-Beans were purchased from the National Institute of Metrology of China. The glassware is soaked overnight in 20 % v/v HNO₃ and thoroughly rinsed with pure water during use.

2.2. Sample collection and preparation

Totally 40 individual coffee samples were collected between October 2022 and April 2023 from three production areas, including: Pu'er city (22° 02' to 24° 50' N, 99° 09' to 102° 19' E, 317 m to 3370 m above sea level, 19.5 °C average annual temperature, 1172 to 2780 mm annual rainfall), in Yunnan province (12 samples), Baoshan city (98° 05' to 100° 02' N, 24° 08' to 25°51' E, 535 m to 3781 m above sea level, 15.5 °C average annual temperature, 922 to 2112 mm annual rainfall) in Yunnan province (15 samples), and Wanning city (18° 35' to 19° 06' N, 110° 00' to 110° 34' E, 9 m to 1288 m above sea level, 24.0 °C average annual temperature, 1646 to 2691 mm annual rainfall) in Hainan province (13 samples) (Fig. 1). Subsequently, all coffee beans were fully dried in the air and ground into powder using a high-speed grinder (FW100, Tianjin Taist Co, China) according to traditional methods.

2.3. Samples digestion

The prepared coffee powder (0.5000 g) was added into 65 % of concentrated nitric acid (6 mL) and 30 % of hydrogen peroxide (2 mL) (Khan et al., 2014). The mixture was incubated at 100 °C for 2 h, and then digested using a microwave digestion instrument (MARS, CEM Co, US). At a power of 1600 W, the sample was firstly heated to 120 °C within 10 min and maintained for 5 min, followed by heating to 150 °C within 5 min and maintaining for 5 min, and then heated to 180 °C within 5 min and maintain for 20 min. After this, the resulting sample was slightly cooled and treated using an acid catcher (VB24 UP, Beijing Laibertek Co., China) at 160 °C until almost dry. It was then cooled again and diluted with ultrapure water in a 25 mL colorimetric tube for testing (Khan et al., 2013).

2.4. Test methods

The test solution was determined using an ICP-MS (X2, Thermo Fisher Co, US) under an optimized condition (Collision gas: high purity He; 1350 W; Cooling air flow rate 15.0 L/min; Atomizing airflow speed 0.75 L/min; Sampling depth 7.5 mm; Ni sampling cone; Ni intercepts the cone) (Chen et al., 2011). The quantitative analysis was performed using external standard method. Sc, Ge, Rh, and Re are used as internal standard elements, analyzed by injection at the same time with the samples, and the determination was repeated three times for each sample.

2.5. Method validation criteria

Method linear relationship validation. Under the set instrument conditions, the prepared mixed standard solution was measured, and the linear correlation coefficient *r* of the solution was calculated based on

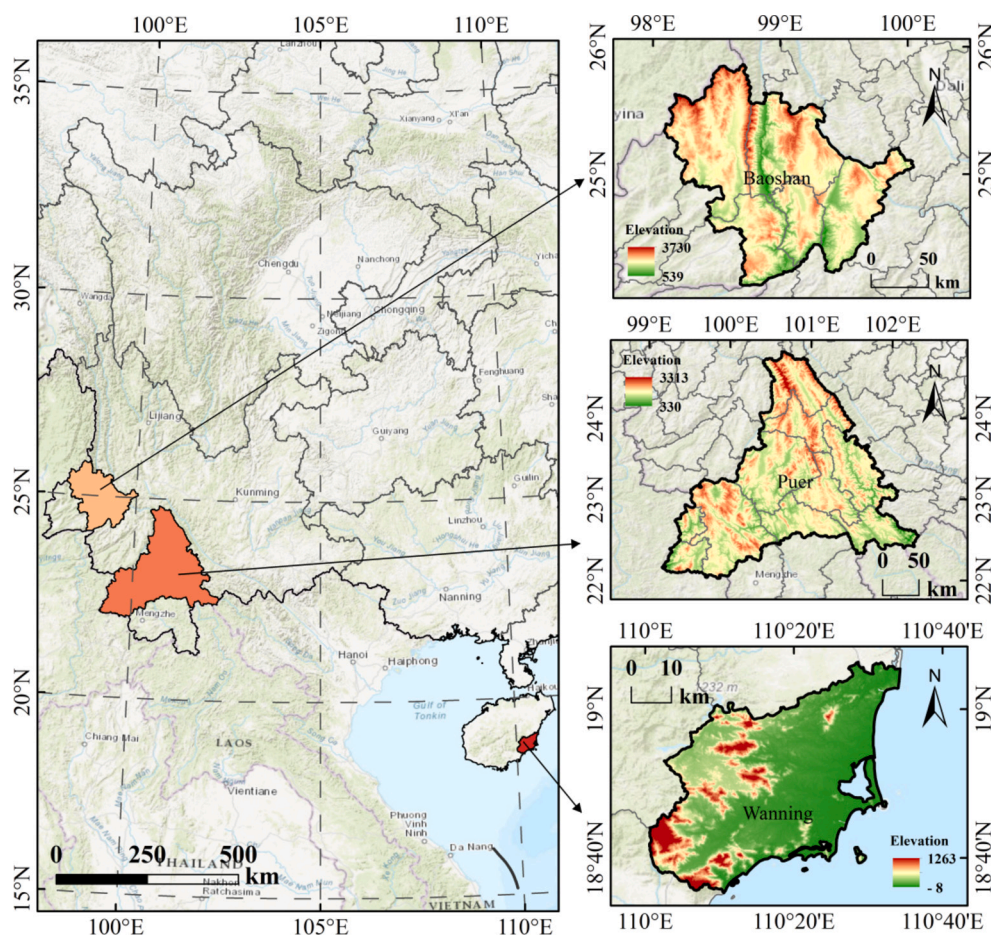


Fig. 1. Sampling locations in major coffee producing provinces in China.

the experimental results to obtain the linear relationship of the method.

Method detection limit and limit of quantification validation. Under the set instrument conditions, measure 11 blank solutions and calculate the standard deviation of the blank solution based on the results to obtain the detection limit of the method. The detection limit of the blank solution is calculated as three times the standard deviation. The limit of quantification for blank solutions is calculated as ten times the standard deviation.

Method precision experiment. According to the established experimental method for detecting coffee elements, the results were processed and measured six times in parallel. The relative standard deviation (RSD) of the data was calculated based on the experimental results to evaluate the accuracy of the method.

Method accuracy experiment. One is the standard substance validation experiment, which uses the established experimental method to determine the standard substances of tea (GBW10016a) and beans (GBW10021). Another method is the peak recovery experiment, which measures the recovery rates of 16 selected elements and generates peaks at two selected concentration levels of 0.1 and 1.0 $\mu\text{g}\cdot\text{g}^{-1}$ based on the concentration in the sample.

2.6. Statistical analysis

SPSS 26.0 software was used for statistical analysis, applying statistical methods including PCA, FDA, PLS-DA. Significance differences were evaluated by comparing the effects of different sample ($p < 0.05$).

3. Results & discussion

3.1. Method validation

The linearity, detection limit, limit of quantification, precision, accuracy, and peak recovery experiments of the applied analytical method demonstrate that the method meets the requirements and standards of the complies with the Chinese national standard - GB/T 27417-2017 Conformity Assessment Guidelines for the Confirmation and Validation of Chemical Analytical Methods. The linear correlation coefficients r of each element are all above 0.999, indicating a good linear relationship between the elements within the concentration range of the standard curve. The detection limit of 16 elements in coffee is 0.0004–0.63 mg/kg, and the limits of quantification is 0.0012–1.9 mg/kg, which meet the requirements of analytical experiments. The relative standard deviation (RSD) of 16 elements is 4.5 % ~ 13.5 %, indicating that the measurement method used in this experiment has good repeatability. The verification experiment of standard substances shows that the experimental results are within the standard value range of standard substances and meet the relevant accuracy requirements. The recovery rate of the peak experiment is 86.0 % ~ 96.3 %. The RSD ranges from 1.1 % to 8.8 %, indicating that the established detection method is accurate and reliable and can be used for rapid analysis of multiple elements in coffee.

3.2. Inorganic element content in coffee

Follow the established experimental method for pre-treatment and machine testing. According to the results in Fig. 2, coffee from Yunnan Pu'er, Yunnan Baoshan, and Hainan Wanning regions all contain rich inorganic elements such as K, Mg, Ca, Mn, Na, Fe, and there are also

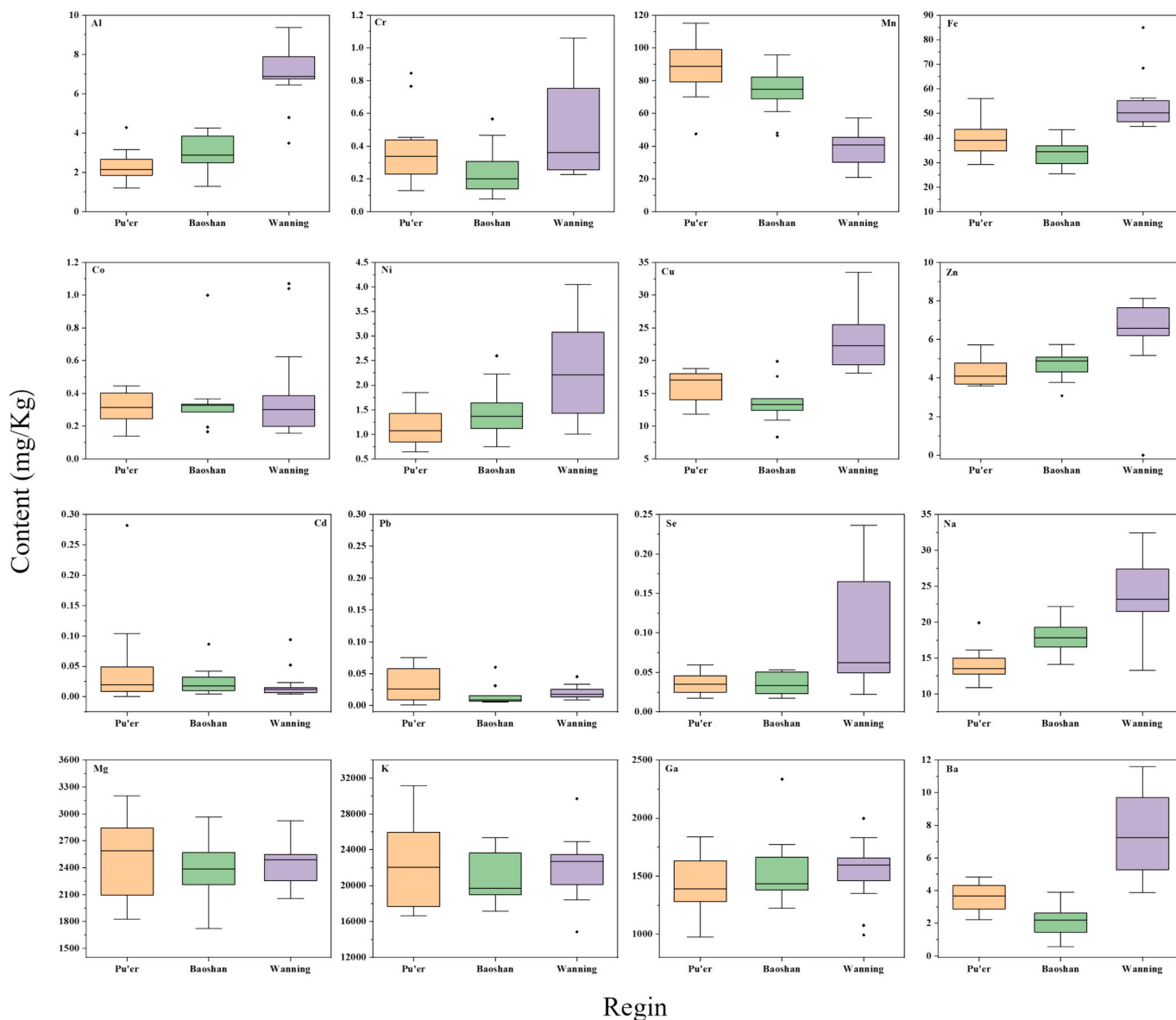


Fig. 2. Box plots of 16 element contents in coffee from Yunnan Pu'er, Yunnan Baoshan, and Hainan Wanning regions.

significant differences in the content of various elements. The content of K element in coffee from three producing areas is the highest, with an average of over 20 g/kg and a maximum value of 29.8 g/kg; The content of Mg element reaches 1720–3200 mg/kg; The content of Ca element reaches 978–2335 mg/kg; The content of Mn and Ni is also above 10 mg/kg. Similar to the findings of Worku et al. in terms of trends, especially in the concentrations of key elements such as Cu and Mn (Worku et al., 2019). The harmful heavy metals such as Cd and Pb in coffee meet the limit requirements of the Codex Alimentarius Commission's Codex General Standard for Contaminants and Toxins in Food, but there are still significant differences in the elemental content of coffee from different origins. Therefore, in response to the complex situation of coffee quality and pollution status in different regions, it is necessary to analyze the common main factors of coffee from different regions as quality control indicators, and timely understand the pollution risk of heavy metals in coffee from different regions.

An analysis of variance was conducted on 16 elements in coffee from three different producing areas. The results showed that there were significant differences ($P < 0.05$) in the content of Al, Ba, Ca, Cd, Co, Cu, Cr, Na, K, Fe, Mg, Mn, Zn, Ni, Pb, and Se in coffee samples from different

producing areas. The element content in coffee samples from different producing areas had their own regional characteristics. From Table 1, it can be found that the Al, Cr, Fe, Co, Ni, Cu, Zn, Se, Na, Ba elements in coffee from Hainan are generally significantly higher than those from Yunnan ($P < 0.05$). The content of various elements in coffee from Pu'er and Baoshan in Yunnan is relatively similar, but the difference is not significant; The manganese content in the Wanning area of Hainan was significantly lower than that in the other two areas ($P < 0.05$); The content of K, Mg, and Ca in coffee from the three producing areas is similar, and K element is significantly higher than the other two elements; The content of Cd and Pb is relatively low, and the production areas of Pu'er in Yunnan are significantly higher than those of Baoshan in Yunnan and Wanning in Hainan ($P < 0.05$). Therefore, from the differences in the content of multiple elements in coffee, it can be seen that its inorganic element content is influenced by environmental factors such as different growing soils, climatic conditions, and hydrology. It means that specific geographic characteristics of coffee, such as soil composition, climatic conditions, and agricultural practices, can affect the way inorganic elements are taken up and accumulated by coffee plants. For example, higher abundance of specific elements in the soil

Table 1
Determination results of different areas (mg/kg, n = 2).

element	Pu'er, Yunnan (n = 12)		Baoshan, Yunnan (n = 15)		Wanning, Hainan (n = 13)	
	mean value	Relative deviation	mean value	Relative deviation	mean value	Relative deviation
Al	2.32 ^c	0.83	3.06 ^b	0.83	7.06 ^a	1.57
Cr	0.382 ^b	0.22	0.239 ^c	0.13	0.531 ^a	0.32
Mn	88.05 ^a	18.46	73.03 ^b	13.17	38.6 ^c	11.41
Fe	39.63 ^b	7.45	34.20 ^b	5.05	53.7 ^a	11.31
Co	0.321 ^b	0.11	0.347 ^b	0.18	0.406 ^a	0.31
Ni	1.07 ^c	0.31	1.44 ^b	0.48	2.27 ^a	0.97
Cu	16.18 ^b	2.43	13.68 ^b	2.85	23.4 ^a	4.97
Zn	4.31 ^b	0.81	4.69 ^b	0.67	6.38 ^a	2.14
Cd	0.049 ^a	0.079	0.0230 ^b	0.021	0.0201 ^b	0.025
Pb	0.078 ^a	0.17	0.016 ^c	0.015	0.0205 ^b	0.01
Se	0.036 ^b	0.014	0.037 ^b	0.012	0.093 ^a	0.068
Na	14.02 ^c	2.37	17.93 ^b	2.28	23.6 ^a	5.16
Mg	2469.92 ^a	444.61	2421.27 ^a	315.02	2437 ^a	231.5
K	22,198.08 ^a	4889.66	20,846.07 ^b	2723.85	21961 ^a	3635.4
Ca	1442.67 ^b	266.64	1538.47 ^a	272.57	1532 ^a	273.91
Ba	3.60 ^b	0.86	2.11 ^c	0.90	7.46 ^a	2.35

may result in higher concentrations of these elements in the coffee plant.

3.3. Correlation analysis of element content in coffee from different origins

Using 16 elements as analysis indicators, correlation analysis was conducted on coffee samples from three producing areas. The larger the absolute value of the correlation coefficient, the stronger the correlation, the closer the correlation coefficient is to 1 or -1, the stronger the correlation, the closer the correlation coefficient is to 0, and the weaker the correlation (Song et al., 2022). From the correlation coefficient in Fig. 3, it can be seen that there is a significant correlation ($|r| > 0.6$) between many elements in coffee, such as Al and elements such as Ba, Ni, Zn, Na, etc., which are positively correlated and negatively correlated with Mn; Fe is positively correlated with elements such as Cu and Ba, while Cu is positively correlated with elements such as Ba and Se. Related analysis indicates that there is a synergistic or antagonistic relationship between these elements to promote mutual absorption. The inorganic element content of coffee is influenced by various factors, such as soil, climate, and internal distribution of plants. Although the content of each element is influenced by different factors, there may be some inherent patterns in their changes, and tracing the origin of coffee has certain potential applications. The correlation results indicate that these element indicators can reflect the origin information of coffee to a certain extent, and there is a certain correlation between them.

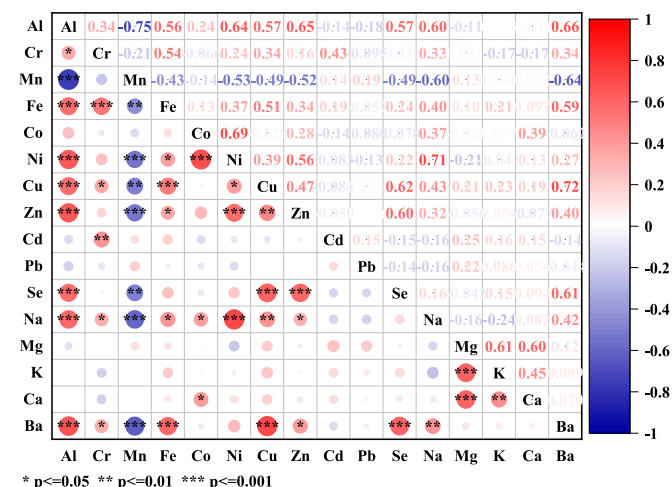


Fig. 3. Pearson correlation coefficients between elements in coffee.

3.4. Principal component analysis (PCA)

The principal component analysis method extracts a few independent comprehensive indicators (i.e. principal components) from a large amount of data that can fully reflect the information of the original variables by reducing the role of dimensions (Jolliffe & Cadima, 2016; Souza et al., 2024). Using the content of 16 elements in coffee from different regions as indicators, principal component analysis (PCA) was conducted using SPSS software. The results showed that the KMO statistic for this analysis was 0.660, indicating a significant correlation between all elements. The analysis obtained the characteristic values, contribution rates, and cumulative contribution rates of each principal component (Table 2a). The results of principal component analysis showed that the characteristic values of the first five principal components were 5.332, 2.461, 1.851, 1.664, and 1.008, respectively, with variance contribution rates of 33.324 %, 15.381 %, 11.569 %, 10.399 %, and 6.298 %, and cumulative contribution rates of 76.971 %, exceeding 70 % of the total variance. It can be concluded that the first five principal components can represent the characteristic information of inorganic elements in coffee.

From Fig. 4a, it can be seen that the first two principal components are positively correlated with some elements. The first principal component is positively correlated with Al, Ni, Cu, Zn, Ba, and negatively correlated with Mn, while the second principal component is positively correlated with Mg and K. Therefore, Al, Mn, Ni, Cu, Zn, Ba, Mg, and K are characteristic elements of coffee, which can effectively affect the composition of inorganic elements in coffee and fully reflect the information of sample data. By applying this scientific analysis technique, the elemental characteristics of coffee samples from different origins were analyzed in depth, thus providing reliable data support for coffee origin traceability.

Table 2a
The characteristic value, contribution rate and cumulative contribution rate of each principal component.

principal component	characteristic value	Variance contribution rate (%)	Cumulative variance contribution rate (%)
1	5.332	33.324	33.324
2	2.461	15.381	48.705
3	1.851	11.569	60.274
4	1.664	10.399	70.673
5	1.008	6.298	76.971

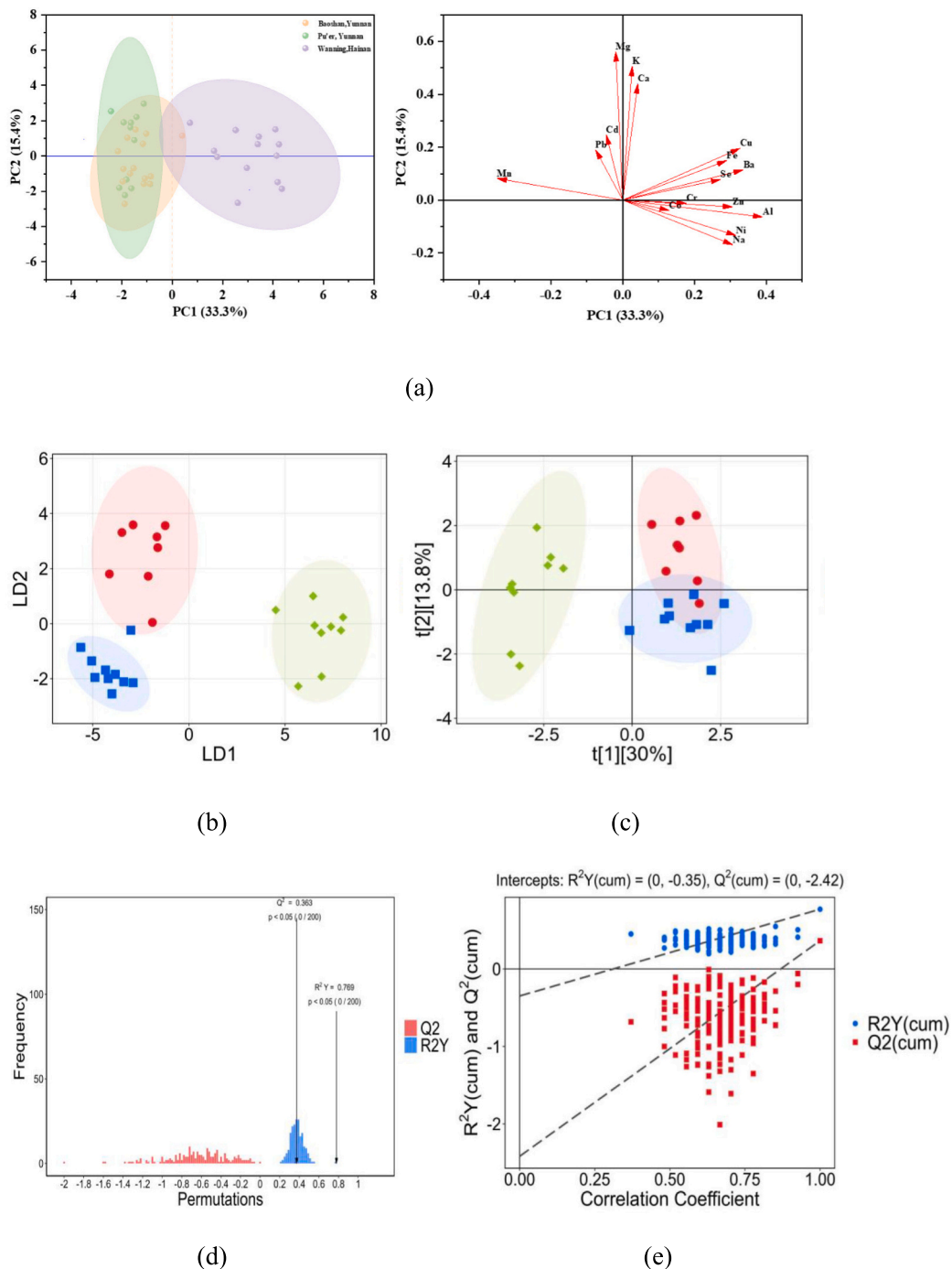


Fig. 4. a) Principal Component Analysis of 16 Element Contents in Coffee from Different Regions; b) Fisher Discriminant Analysis of 16 Elements in Coffee from Different Regions (The blue square represents Yunnan Pu'er coffee; The green rhomboid represents Hainan Wanning coffee; Red ball represents Yunnan Baoshan coffee); Comparison of the results of PLS-DA determination of 16 elements in coffee of different origins: c) Score plot (The blue square represents Yunnan Pu'er coffee; The green rhomboid represents Hainan Wanning coffee; Red ball represents Yunnan Baoshan coffee); d) Histogram of the replacement test; e) Plot figure of the replacement test. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.5. Fisher discriminant analysis (FDA)

Through the study of the differences in inorganic element content, it can be found that there are differences in the inorganic element content of coffee from different regions, but this cannot accurately distinguish the coffee production area. In order to determine the origin of coffee, Fisher discriminant analysis method was used to perform discriminant analysis on coffee samples from three production areas, including Fisher

function and cross validation.

In addition, in order to better understand the discrimination of various element content indicators on the origin of coffee, a general discrimination method based on Fisher discriminant function was established, and multivariate discriminant analysis was used to comprehensively analyze coffee samples. A total of 16 mineral elements were selected as independent variables in the analysis, and stepwise discriminant analysis was conducted to construct a discrimination

model. Stepwise discriminant analysis helps to construct a discriminant model that is able to classify the origin of coffee samples based on elemental content. By comparing the contribution of different elements in the model, six inorganic elements, Al, Mn, Fe, Cu, Na, and Ba, are identified as elements with significant effect in discriminating coffee samples of different origins. The contents of these elements vary significantly among different origins and play a key role in the construction of the discriminant model. The discriminant function model coefficients shown in Table 2b and the discriminant results demonstrated in Table 3 confirm that these elements have significant discriminatory power in distinguishing coffee samples of origin from Yunnan Pu'er, Yunnan Baoshan and Hainan Wanning. This finding provides a scientific basis for coffee origin traceability and lays the foundation for future research and application.

Fisher linear discriminant model for coffee producing areas is obtained from Table 3:

$$Y1 = 0.588 \times 1 - 0.026 \times 2 + 0.057 \times 3 + 0.135 \times 4 + 0.138 \times 5 + 0.251 \times 6 - 9.146;$$

$$Y2 = -0.126 \times 1 + 0.032 \times 2 + 0.04 \times 3 + 0.06 \times 4 - 0.126 \times 5 + 0.358 \times 6 - 3.57.$$

The classification results indicate that discriminant analysis has successfully distinguished three coffee producing areas in China, achieving traceability of the origins of Yunnan Pu'er, Yunnan Baoshan, and Hainan Wanning coffee. The correct discrimination rate of the back-generation test reached 92.5 %, while the overall correct discrimination rate of the cross-test was 90.0 %, showing the model's ability to discriminate in practical applications.

In the cross-examination, three samples from the Baoshan production area in Yunnan were mistakenly identified as the Yunnan Pu'er production area, and one sample from the Yunnan Pu'er production area was mistakenly identified as the Yunnan Baoshan production area. In contrast, the discrimination rate of the Wanning production area in Hainan was 100 %, which is consistent with the differences in element content mentioned earlier. Overall, the misjudgment rate of the cross-examination is less than 10 %, indicating that the model has practical application value for coffee origin traceability and that differences based on inorganic element fingerprints can effectively distinguish coffee from different origins.

The overall correct discrimination rate of the model reaches an acceptable standard (>90 %), which is based on the results of selecting 16 elements for detection. Although adding some elements can increase the discrimination rate to a certain extent, its effect is limited, and it is based on the increase in workload costs. In response to this, we have tried our best to eliminate redundant elements and achieve a suitable result.

Substituting the inorganic element content of coffee samples into the above discriminant traceability model, scatter plots were created based on the scores of Y1 and Y2, respectively. From Fig. 4b, the origins of Yunnan Pu'er, Yunnan Baoshan, and Hainan Wanning coffee can be more intuitively identified. Among them, there is a small overlap between Yunnan Pu'er and Baoshan coffee origins, which can easily cause model discrimination errors.

Table 2b
Distinctive coffee distinguish function model coefficient.

inorganic elements	function	
	1	2
Al(X1)	0.588	-0.126
Mn(X2)	-0.026	0.032
Fe(X3)	0.057	0.04
Cu(X4)	0.135	0.06
Na(X5)	0.138	-0.126
Ba(X6)	0.251	0.358
(Constant)	-9.146	-3.57

3.6. Partial least squares discriminant analysis (PLS-DA)

In contrast to PCA, a trend of increasing emphasis on differentiating samples from different production areas can be observed in the PLS-DA results (Fig. 4c). Unlike PCA, which relies only on the content of the elements themselves to distinguish different production areas, PLS-DA presupposes a matrix related to the production areas to assist in the distinction and is a more appropriate method to differentiate between different coffee geographic origins.

After the PLS-DA model is built, it is usually necessary to use some tests to verify the fit of the model and the accuracy of the prediction. The replacement test is currently the most commonly used evaluation method for PLS-DA models, and R^2Y and Q^2Y are the two most common parameters for the replacement test. As shown in Fig. 4d, the parameters of the fitted model are $R^2Y = 0.769$ and $Q^2 = 0.363$, indicating that the model has good explanatory ability for dependent variables Y ($R^2Y = 0.769$; usually the closer the R^2 is to 0.8, the more the model can explain most of the variance in the original data), and $Q^2 = 0.363$ also indicates that the model has good predictive ability. The $P < 0.05$ for the replacement test reinforces this point.

In addition, as shown in Fig. 4e, the horizontal axis of the figure shows the correlation between the dependent variable after data label substitution and before substitution, and the vertical coordinates are the values of R^2Y and Q^2 . Usually, to determine whether the model is overfitted based on the replacement test, one needs to focus on the following components: 1) all blue points on the left side should be lower than the original points on the rightmost side; 2) all red points on the left side should be lower than the original points on the rightmost side; and 3) the intercept of the regression line of the predicted value of Q^2 on the Y-axis should be less than 0. Based on the above criteria, the model is not overfitted and is reliable.

For PLS-DA, their biggest value in differentiating between coffees of different geographical origins is "providing VIP value." The VIP value reflects the contribution of each element to the interpretation of the model. In simple terms, the larger the VIP value of the element, the more important it is to distinguish between different coffee producing areas and the more worthy of attention. Usually, we think that $VIP > 1$ is one of the criteria for screening characteristic elements. The study found that the six most important elements, namely Al (VIP:1.398), Mn (VIP:1.321), Fe (VIP:1.152), Cu (VIP:1.142), Na (VIP:1.137), and Ba (VIP:1.546), are considered to be an important element characteristic to distinguish coffee samples from different origins, and this result is consistent with Fisher discriminant analysis.

3.7. Comparison of FDA and PLS-DA classification effectiveness

The Receiver Operating Characteristic Curve (ROC) is a visualization tool for evaluating classification methods and models. It is used to assess the effectiveness of certain metrics in distinguishing between samples from two or more different categories. By plotting the ROC curve and calculating the area under the curve (AUC), accuracy, and sensitivity, it is possible to compare the classification effectiveness of different methods. In general, the closer the accuracy, sensitivity, and AUC values are to 1, the better the classification performance is represented.

As shown in Fig. 5, when comparing the results of the two classification methods, FDA and PLS-DA, we find that both methods have the same accuracy (0.917) and sensitivity (0.889). However, the AUC value of FDA is 0.978, which is significantly higher than that of PLS-DA (0.833), which suggests that FDA is better than PLS-DA in terms of classification effectiveness, thus enabling more accurate traceability of coffee samples from different producing regions.

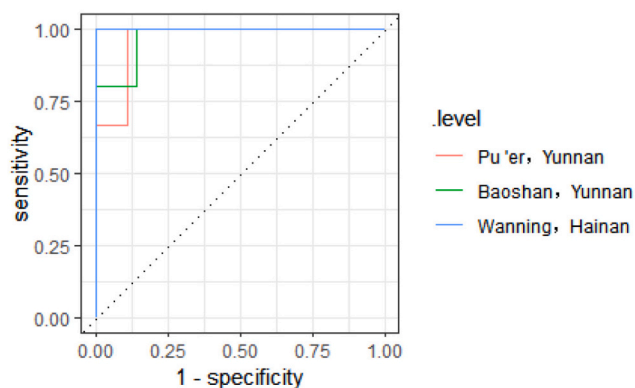
3.8. Comparison with similar studies

In the research field of food source identification, the combination of elemental fingerprints and chemometric methods has been proven to be

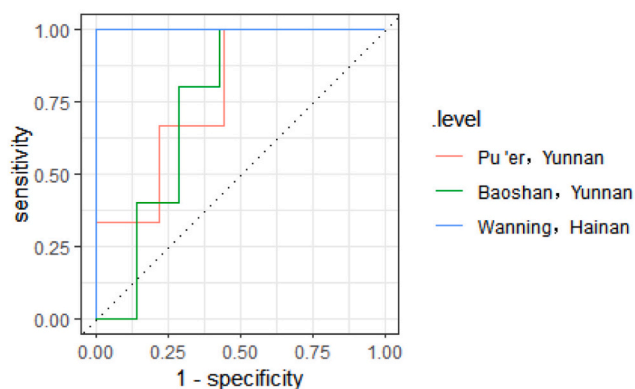
Table 3
General discriminant analysis of *Lycium barbarum* from different origins.

method	producing area	Prediction group members			total	Discrimination accuracy/ %	Overall correct discrimination rate/%
		Baoshan, Yunnan	Pu'er, Yunnan	Wanning, Hainan			
Initial ^a	Count	Baoshan, Yunnan	10	2	0	12	83.3
		Pu'er, Yunnan	1	14	0	15	93.3
		Wanning, Hainan	0	0	13	13	100
Cross validation ^{bc}	Count	Baoshan, Yunnan	9	3	0	12	75
		Pu'er, Yunnan	1	14	0	15	93.3
		Wanning, Hainan	0	0	13	13	100

Note: a) Correctly classified 92.5 % of the original grouped cases; b) Perform cross validation only for the individual cases in the analysis. In cross validation, each case is classified by functions derived from all cases outside of that case; c) Correctly classified 90.0 % of grouped cases that underwent cross validation.



(a)



(b)

Fig. 5. ROC curve: a) FDA; b) PLS-DA.

a powerful tool. ICP-MS technology is especially favored due to its high sensitivity and multi-element analysis ability. From Table 4, it can be seen that ICP-MS technology has played a key role in identifying the geographical sources of different foods. For example, Su et al. (Su et al., 2022) analyzed 39 mineral elements through ICP-MS in their study of wine and combined chemometric methods such as PCA, HCA, LDA, and ANN to achieve a correct classification rate of over 81.7 %. However, this paper achieved over 90 % accuracy in identifying the geographical origin of coffee using carefully selected 16 elements and ICP-MS

technology, combined with PCA and FDA. This result highlights the possibility of reducing the number of analytical elements while maintaining or improving classification accuracy.

Of particular note, Habte et al. (Habte et al., 2016) analyzed 45 elements in Ethiopian coffee using ICP-OES and ICP-MS and applied CA, LDA, and PCA, achieving a 100 % recognition rate. Although this result is impressive, this paper achieved similar high accuracy with fewer key elements, which not only reduces analysis costs but also improves analysis efficiency.

In addition, Liu et al. (Liu, Zeng, Zhao, & Tong, 2020) achieved a correct classification rate of more than 90 % by analyzing key elements, including 86Sr and 112Cd, using ICP-MS and ICP-OES techniques combined with PCA, LDA, and PLS-DA in the geographic origin identification of tea. This demonstrates that efficient geographic source identification can be achieved by selecting highly geographically characterized elements, even when the number of elements is limited.

These studies demonstrate the potential of ICP-MS technology in food authenticity analysis, especially in this paper, where higher classification accuracy was achieved using fewer elements by optimizing element selection and chemometric methods. This not only provides a new strategy for food source identification but also provides a valuable reference for future research.

4. Conclusion

This study demonstrates that ICP-MS combined with principal component analysis, Fisher discriminant analysis, and Partial least squares discriminant analysis effectively distinguishes the geographical origin of coffee based on its inorganic element content. By identifying the key elements—Al, Mn, Fe, Cu, Na, and Ba—and modeling based on them, both FDA and PLS-DA obtained more than 90 % correct discrimination rate, but FDA had better identification performance than PLS-DA. These findings offer a scientific approach to coffee origin traceability, providing a reliable alternative to traditional sensory evaluations. This advancement is pivotal for enhancing quality control, ensuring authenticity, and supporting the burgeoning coffee industry in China.

CRediT authorship contribution statement

Xiaowei Pan: Writing – original draft, Methodology, Funding acquisition, Data curation. **Wenjing Yan:** Writing – original draft, Visualization, Software, Formal analysis. **Xiaopeng Wu:** Data curation. **Jianzhi Ye:** Formal analysis. **Yaohui Liang:** Investigation. **Guoyan Zhan:** Data curation. **Hao Dong:** Methodology. **Wenzhen Liao:** Investigation. **Xingfen Yang:** Validation. **Qi He:** Writing – review & editing, Supervision, Conceptualization.

Table 4

Application of instrumental analysis and chemometric methods in the identification of the geographical origin of beverage.

Samples	Quantity	Key Variable	Method of Analysis	Chemometrics	Correct discrimination rate/%	Reference	
Tea	Longjing tea	150	$\delta^{13}\text{C}$, $\delta^{15}\text{N}$, $\delta^2\text{H}$, $\delta^{18}\text{O}$, K, Ca, Cu, Mg, Rb, Zn	EA/IRMS, ICP-MS	PCA, LDA, PLS-DA	>87.8 %	(Liu et al., 2019)
	Guizhou tea	73	P, Sr, U, Pb, Cd, Cr	ICP-MS, ICP-OES	PCA, SLDA	100%	(Zhang et al., 2020)
	Yongchuan Xiuya tea	72	^{86}Sr , ^{112}Cd	ICP-MS, ICP-OES	PCA, LDA, PLS-DA	>90 %	(Liu et al., 2020)
	Biluochun tea	56	37 metal elements	ICP-MS	PCA, LDA, CA	95.2 %	(Ma et al., 2016)
Coffee	–	76	41 elements, Ca, Ba, Cd, Rb, Sr, S, Tm	ICP-MS, ICP-OES, ICP-MS/MS	PCA, LDA	>93.3 %	(Vezzulli, Fontanella, Lambri, & Beone, 2023)
	green arabica coffee	103	$\delta^{13}\text{C}$, $\delta^{15}\text{N}$, $\delta^{18}\text{O}$	ICP-OES, ICP-MS, WD-XRF,	LDA	>80 %	(Worku et al., 2019)
	Ethiopia Coffee	129	45 elements, Mn, Cu, Sr, Zn, Rb, Ni, B	ICP-OES, ICP-MS, DMA	CA, LDA, PCA	100%	(Habte et al., 2016)
	–	39	59 elements, Co, Se, Rb, Sr, Mo, Te, Cs, Ba	ICP-AES, ICP-MS	PCA, LDA	96.9 %	(Valentin & Watling, 2013)
Wine	–	104	39 mineral elements, $\delta^{18}\text{O}$, Sr/Rb ratios	ICP-MS, IRMS	PCA, HCA, LDA, ANN	>81.7 %	(Su et al., 2022)
	Diqing wines	36	45 elements, $\delta^{13}\text{C}$, $\delta^{18}\text{O}$	ICP-MS, IRMS	PCA, DA, ANN-MLP	>88.9 %	(Su et al., 2024)
Cocoa	–	61	56 elements (29 valid elements)	ICP-MS	PCA, DA	>85 %	(Bertoldi, Barbero, Camin, Caligiani, & Larcher, 2016)

Declaration of competing interest

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

Acknowledgments

Thanks to the financial support provided by the Special Support Program of Guangdong Province (2021TX06N107), Hainan Provincial Natural Science Foundation (321QN0927), the Special Fund for Basic Scientific Research Business of Central Level Public Welfare Research Institutes (NO. 1630122024005), the National Natural Science Foundation of China (NO. 31901750), and the Guangzhou Science and Technology Plan Project (NO. SL2022A04J01641).

Data availability

Data are contained within the article and will be made available on request.

References

- Barrea, L., Pugliese, G., Frias-Toral, E., Ghoche, M. E., Castellucci, B., Chapela, S. P., & de Los Angeles Carignano, M., Laudisio, D., Savastano, S., Colao, A., & Muscogiuri, G. (2023). Coffee consumption, health benefits and side effects: A narrative review and update for dietitians and nutritionists. *Critical Reviews in Food Science and Nutrition*, 63(9), 1238–1261. <https://doi.org/10.1080/10408398.2021.1963207>
- Bertoldi, D., Barbero, A., Camin, F., Caligiani, A., & Larcher, R. (2016). Multielemental fingerprinting and geographic traceability of Theobroma Cacao beans and cocoa products. *Food Control*, 65, 46–53. <https://doi.org/10.1016/j.foodcont.2016.01.013>
- Chen, L., Liu, Y., Hu, Z., Gao, S., Zong, K., & Chen, H. (2011). Accurate determinations of fifty-four major and trace elements in carbonate by La-Icp-MS using normalization strategy of bulk components as 100%. *Chemical Geology*, 284(3–4), 283–295. <https://doi.org/10.1016/j.chemgeo.2011.03.007>
- Dong, W., Tan, L., Zhao, J., Hu, R., & Lu, M. (2015). Characterization of fatty acid, amino acid and volatile compound compositions and bioactive components of seven coffee (Coffea Robusta) cultivars grown in Hainan Province. *China. Molecules*, 20(9), 16687–16708. <https://doi.org/10.3390/molecules200916687>
- Dos, S. G. W., Pereira, L. L., Rodrigues, D. L. J., Soares, D. S. M., Reis, V. T., & Partelli, F. L. (2024). Exploring the microbiome of coffee plants: Implications for coffee quality and production. *Food Research International*, 179, Article 113972. <https://doi.org/10.1016/j.foodres.2024.113972>
- Habte, G., Hwang, I. M., Kim, J. S., Hong, J. H., Hong, Y. S., Choi, J. Y., ... Kim, K. S. (2016). Elemental profiling and geographical differentiation of Ethiopian coffee

- samples through inductively coupled plasma-optical emission spectroscopy (Icp-Oes), Icp-mass spectrometry (Icp-MS) and direct mercury analyzer (Dma). *Food Chemistry*, 212, 512–520. <https://doi.org/10.1016/j.foodchem.2016.05.178>
- Haile, M., & Kang, W. H. (2019). The role of microbes in coffee fermentation and their impact on coffee quality. *Journal of Food Quality*, 2019(12). <https://doi.org/10.1155/2019/4836709>
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A-Mathematical Physical and Engineering Sciences*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- Khan, N., Choi, J. Y., Nho, E. Y., Jamila, N., Habte, G., Hong, J. H., ... Kim, K. S. (2014). Determination of minor and trace elements in aromatic spices by Micro-wave assisted digestion and inductively coupled plasma-mass spectrometry. *Food Chemistry*, 158(1), 200–206. <https://doi.org/10.1016/j.foodchem.2014.02.103>
- Khan, N., Jeong, I. S., Hwang, I. M., Kim, J. S., Choi, S. H., Nho, E. Y., ... Kim, K. S. (2013). Method validation for simultaneous determination of chromium, molybdenum and selenium in infant formulas by Icp-Oes and Icp-MS. *Food Chemistry*, 141(4), 3566–3570. <https://doi.org/10.1016/j.foodchem.2013.06.034>
- Liu, D., Huang, Q., Song, W., Li, Z., & Mao, C. (2023). Association of Sugar-Sweetened, artificially sweetened, and unsweetened coffee consumption with all-cause and cause-specific mortality. *Annals of Internal Medicine*, 176(1). <https://doi.org/10.7326/L22-0369>
- Liu, H., Zeng, Y., Zhao, X., & Tong, H. (2020). Improved geographical origin discrimination for tea using Icp-MS and Icp-Oes techniques in combination with Chemometric approach. *Journal of the Science of Food and Agriculture*, 100(8), 3507–3516. <https://doi.org/10.1002/jsfa.10392>
- Liu, Z., Yuan, Y., Zhang, Y., Shi, Y., Hu, G., Zhu, J., & Rogers, K. M. (2019). Geographical traceability of Chinese green tea using stable isotope and multi-element Chemometrics. *Rapid Communications in Mass Spectrometry*, 33(8), 778–788. <https://doi.org/10.1002/rcm.8405>
- Ma, G., Zhang, Y., Zhang, J., Wang, G., Chen, L., Zhang, M., Liu, T., Liu, X., & Lu, C. (2016). Determining the geographical origin of Chinese green tea by linear discriminant analysis of trace metals and rare earth elements: Taking Dongting Biluochun as an example. *Food Control*, 59, 714–720. <https://doi.org/10.1016/j.foodcont.2015.06.037>
- Ma, J., Li, J., He, H., Jin, X., Cesarino, I., Zeng, W., & Li, Z. (2022). Characterization of sensory properties of Yunnan coffee. *Current Research in Food Science*, 5, 1205–1215. <https://doi.org/10.1016/j.crf.2022.07.010>
- Moor, C., Devos, W., Guecheva, M., & Kobler, J. (2000). Inductively coupled plasma mass spectrometry: A versatile tool for a variety of different tasks. *Fresenius Journal of Analytical Chemistry*, 366(2), 159–164. <https://doi.org/10.1007/s002160050030>
- Nguyen, Q. T., Nguyen, T. T., Le, V. N., Nguyen, N. T., Truong, N. M., Hoang, M. T., ... Bui, Q. M. (2023). Towards a standardized approach for the geographical traceability of plant foods using inductively coupled plasma mass spectrometry (Icp-MS) and principal component analysis (Pca). *Foods*, 12(9). <https://doi.org/10.3390/foods12091848>
- Schaneberg, B. T. (2020). Green and roasted coffee authentication: Species, origin, and diluent methods of analysis. *Journal of AOAC International*, 103(2), 281–282. <https://doi.org/10.1093/jaoacint/qs2022>
- Song, M., Xu, H., Xin, G., Liu, C., Sun, X., Zhi, Y., Li, B., & Shen, Y. (2022). Comprehensive evaluation of Actinidia Arguta fruit based on the nutrition and taste:

- 67 germplasm native to Northeast China. *Food Science and Human Wellness*, 11(2), 393–404. <https://doi.org/10.1016/j.fshw.2021.11.020>
- Souza, A. S., Bezerra, M. A., Cerqueira, U. M. F. M., Rodrigues, C. J. O., Santos, B. C., Novaes, C. G., & Almeida, E. R. V. (2024). An introductory review on the application of principal component analysis in the data exploration of the chemical analysis of food samples. *Food Science and Biotechnology*, 33(6), 1323–1336. <https://doi.org/10.1007/s10068-023-01509-5>
- Su, Y., Zhang, J., Wang, L., Dias Araujo, L., Tan, D., Yuan, C., & Zhang, A. (2024). Assessing geographical origin of Dìqīng wines based on their elemental and isotopic profiles. *Journal of Food Composition and Analysis*, 125, Article 105671. <https://doi.org/10.1016/j.jfca.2023.105671>
- Su, Y., Zhao, Y., Cui, K., Wang, F., Zhang, J., & Zhang, A. (2022). Wine characterisation according to geographical origin using analysis of mineral elements and rainfall correlation of oxygen isotope values. *International Journal of Food Science & Technology*, 57(1), 552–565. <https://doi.org/10.1111/ijfs.15236>
- Valentin, J. L., & Watling, R. J. (2013). Provenance establishment of coffee using solution Icp-Ms and Icp-Aes. *Food Chemistry*, 141(1), 98–104. <https://doi.org/10.1016/j.foodchem.2013.02.101>
- Van Dam, R. M., Hu, F. B., & Willett, W. C. (2020). Coffee, caffeine, and health. *The New England Journal of Medicine*, 383(4), 369–378. <https://doi.org/10.1056/NEJMr1816604>
- Velasquez, S., & Banchon, C. (2023). Influence of pre-and post-harvest factors on the organoleptic and physicochemical quality of coffee: A short review. *Journal of Food Science and Technology*, 60(10), 2526–2538. <https://doi.org/10.1007/s13197-022-05569-z>
- Vezzulli, F., Fontanella, M. C., Lambri, M., & Beone, G. M. (2023). Specialty and high-quality coffee: Discrimination through elemental characterization Viaicp-Oes, Icp-Ms, Andicp-Ms/Ms of origin, species, and variety. *Journal of the Science of Food and Agriculture*, 103(9), 4303–4316. <https://doi.org/10.1002/jsfa.12490>
- Worku, M., Upadhyay, H. R., Latruwe, K., Taylor, A., Blake, W., Vanhaecke, F., ... Boeckx, P. (2019). Differentiating the geographical origin of Ethiopian coffee using Xrf- and Icp-based multi-element and stable isotope profiling. *Food Chemistry*, 290, 295–307. <https://doi.org/10.1016/j.foodchem.2019.03.135>
- Zhang, J., Yang, R., Li, Y. C., Wen, X., Peng, Y., & Ni, X. (2020). Use of mineral multi-elemental analysis to authenticate geographical origin of different cultivars of tea in Guizhou, China. *Journal of the Science of Food and Agriculture*, 100(7), 3046–3055. <https://doi.org/10.1002/jsfa.10335>
- Zhao, H., Guo, B., Wei, Y., Zhang, B., Sun, S., Zhang, L., & Yan, J. (2011). Determining the geographic origin of wheat using multielement analysis and multivariate statistics. *Journal of Agricultural and Food Chemistry*, 59(9), 4397–4402. <https://doi.org/10.1021/jf200108d>