Geographical origin discrimination of pepper (*Capsicum annuum* L.) based on multi-elemental concentrations combined with chemometrics

Jian Zhang¹ · Ruidong Yang¹ · Rong Chen² · Yuncong C. Li³ · Yishu Peng¹ · Xuefeng Wen⁴

Received: 28 September 2018 / Revised: 17 April 2019 / Accepted: 21 April 2019 / Published online: 20 May 2019
© The Korean Society of Food Science and Technology 2019

Abstract The concentrations of twenty-four elements in twenty-five peppers from three cultivated regions in Guizhou Province (China) were analyzed. The chemometric data processing, including one-way analysis of variance, principal component analysis, linear discriminant analysis (LDA), and orthogonal partial least squares discriminant analysis (OPLS-DA) were executed to differentiate the peppers. Consequently, the contents of 16 elements (Arsenic, Ba, Cu, Co, Cr, Ni, Pb, Sn, Sb, Mo, Sr, Y, Zn, Ca, P, and Fe) were significantly different among the three regions ($p < 0.05$). The correct discrimination

rates of 25 peppers were 92.0% for LDA and 96.0% for OPLS-DA. The variable importance in the projection (VIP) values were ranged from 1.063 to 1.262 for seven elements (Tin, Fe, Zn, Y, Cr, Sr, and Mo) indicating that they played an important role for the geographical origin traceability of pepper. To sum up, multi-element concentrations together with chemometric data-processing can be promising for the geographical origin differentiation of pepper.

Keywords Pepper · Multi-element · ICP-MS · Chemometric · Geographical origin traceability

✉ Ruidong Yang
rdyang@gzu.edu.cn
Jian Zhang
jzhanggzdxhjx@163.com
Rong Chen
rchengzu@163.com
Yuncong C. Li
yunli@ufl.edu
Yishu Peng
pengys520@126.com
Xuefeng Wen
wenxuefeng@aliyun.com

- ¹ College of Resource and Environmental Engineering, Guizhou University, Huaxi District, Guiyang 550025, China
- ² College of Mining, Guizhou University, Guiyang 550025, China
- ³ Department of Soil and Water Sciences, Tropical Research and Education Center, IFAS, University of Florida, Homestead, FL 33031, USA
- ⁴ College of Agriculture, Guizhou University, Guiyang 550025, China

Introduction

Peppers (*Capsicum annuum* L.) are the fruits of plants that belong to the family *Solanaceae* and genus *Capsicum* (Naccarato et al., 2016). As coloring and flavoring agents, peppers play an important role in the food industry for preparing barbecue, sauces, processed meats, and soups (Pino et al., 2007). It is denoted that capsicums are rich in health-promoting phytonutrients and antioxidants, e.g., protein, carotene, vitamins C, vitamins E, provitamin A, and capsaicin etc. (Naccarato et al., 2016). Moreover, pepper (*Capsicum annuum* L.) is an annual or limited perennial plant, and originated from the tropics of the central Latin America. Now it has been widely cultivated all over the world. In China, it was mainly distributed in Sichuan, Guizhou, and Hunan.

Capsicum is a characteristic and advantaged agricultural product in Guizhou Province (China). At present, pepper in Guizhou has ranked first in China in terms of planting area, processing scale, and market distribution. According to the data of 2017, the planting area, yield, and production value of pepper planting in Guizhou reached 32.9×10^4 hm²,

54.3×10^8 kg, and 15.2 billion CNY, respectively. There was approximately 3.0×10^8 kg of dry pepper from Xiazi pepper market of Zunyi exported to more than 80 countries and regions with an annual turnover of 6 billion CNY. Pepper is economically important for Guizhou (China) as a result of the enormous consumption and its effective dispelling cold and sweating role. In particular, the brand of “The godmother” capsicum is well-known both at home and abroad and favored by many people in the world, which promoted the development of capsicum industry in Guizhou Province. Xiazi in Zunyi (ZYXZ), Huaxi in Guiyang (GYHX), and Hezhang in Bijie (BJHZ) were the main capsicum cultivating areas in Guizhou Province, China. In addition, the capsicums produced in ZYXZ and GYHX were protected by the geographical indications (GI) of Ministry of Agriculture and Rural Affairs of the People’s Republic of China (MARAPRC). Thereby, the GI-authorized capsicums have been sold more expensively than unauthorized capsicums. Specifically, the GI certification includes scope of protected region, variety, site condition, cultivation and management measure, sensory characteristic, physicochemical indicator, and food safety standard. Generally, the selling price of Xiazi pepper and Huaxi pepper is approximately 25 CNY/kg (dry weight, DW) and other non-certified peppers are about 7 CNY/kg (DW). However, the impel of enormous commercial benefits promoted the emergence of fake GI capsicum in the market, which damaged the image of Guizhou capsicum in international trade and the interests of producers and consumers (Chang et al., 2016; Ma et al., 2016). As a result, the authentication of GI-authorized capsicum is needed to protect the trademarks and GI of Guizhou capsicum and it was conducive for producers and consumers. Furthermore, the establishment of capsicum origin traceability models can effectively protect the awareness of Guizhou’s GI capsicums, improve their market competitiveness and avoid the shoddy and deceptive phenomenon on the domestic and international market.

There were many research cases about the geographical origin discrimination of food and agri-products such as welsh onions (Ariyama et al., 2004), sesame seeds (Choi et al., 2017), theobroma cacao beans (Bertoldi et al., 2016), tea (Ma et al., 2016; Ni et al., 2018; Zhao et al., 2017; Zhang et al., 2018a), rice (Chung et al., 2018a), potato (Opatic et al., 2018a), tomato (Opatic et al., 2018b), and wheat (Rashmi et al., 2017). In terms of traceability of pepper geographical origin, a variety of methods were applied, including electronic nose technique (testing sensory characteristics) (Yin et al., 2018), electronic tongue on the basis of graphite pencil electrodes (GPEs) (the analyses of capsaicinoid contents) (Wu et al., 2016), and stable isotope ratios (C, H, O, N, S, and Sr) techniques (Opatic et al., 2017; Rijke et al., 2016; Song et al., 2014). Among these

different analytical methods, ICP-MS and ICP-AES may be the most available and responsible means for the accurate assay of the multi-element in peppers because of low detection limit, high throughput, wide dynamic linear range, few interference, and high analysis precision. In addition, the expenditures of stable isotopes ratios and organic components determination were more expensive than the cost of multi-element measurement. Based on the above-mentioned advantages, many scholars used the concentrations of mineral elements in peppers to carry out the geographical origin discrimination of peppers. For example, the concentrations of 32 elements in chili pepper samples combined with four class-modeling techniques have been reported to authenticate peppers in Calabria (Italy) from those cultivated outside of Calabria (Naccarato et al., 2016). Palacios-Morillo et al. (2014) achieved the identification of the two regions (Murcia and La Vera) of paprika from Protected Designations of Origin (PDO) recognized in Spain by fourteen trace metal elements. Similarly, authentic and purchased paprika from different known, stated, and unascertained geographical origins were classified faultlessly based on thirty-four elements and strontium isotope ratio ($^{87}\text{Sr}/^{86}\text{Sr}$) and an unrepeatable fingerprint of authentic Szegedi paprika (PDO) was established successfully (Brunner et al., 2010). Furthermore, the stable isotopic compositions ($\delta^{13}\text{C}$, $\delta^{15}\text{N}$, $\delta^{18}\text{O}$, and $\delta^{34}\text{S}$) of light elements combined with multi-elemental fingerprinting (Phosphorus, S, Cl, Br, Zn, Sr, Rb, K, Mn, Ca, and Fe) were used to provide instantaneous, powerful, and low-cost screening methods for discriminating sweet pepper samples from six countries (Spain, Italy, Greece, Slovenia, Morocco, and Austria), and the correct prediction rate of the suggested discriminant analysis model was 71.1% (Opatic et al., 2017).

Nowadays, to our knowledge, there are no published papers on the subject of the intra-regional authentication of chili pepper produced in Guizhou Province (China). Besides, it can be still unclear whether the multi-element assay is suitable to distinguish the geographical origin of chili pepper samples from three regions (ZYXZ, GYHX, and BJHZ) in Guizhou Province (China). Thus, this work is different from others reporting the classification of pepper came from regions separated by large intervals, where significant differences in soil composition might be anticipated and had an extraordinary effect on chemical composition of peppers. In addition, it was worthy and significant to discriminate the geographical origin of Guizhou pepper and protect the regional brand of Guizhou pepper because Guizhou is one of the most important and largest capsicum cultivating areas in China. Trustworthy differentiation of the geographical origin of Guizhou capsicum, especially very acclaimed peppers, is pivotal for developing pepper market and protecting consumer rights.

The study results can be advantageous for the regulation and governance of the unethical and swindling labeling of the geographical origin of chili peppers.

Hence, the objective of this study was to assess the potentiality of chemometrics tools (ANOVA, PCA, LDA, and OPLS-DA) associated with the multi-element concentrations achieved by ICP-MS and ICP-AES determination of chili peppers as an original provenance for the separation of peppers in line with their geographical origin. Finally, this is the first work publishing the intra-regional differentiation of peppers cultivated in Guizhou Province (China) in accordance with their geographical origin.

Materials and methods

Pepper collection

In total, 25 pepper samples were collected in three districts in Guizhou (China), including Xiazi in Zunyi (ZYXZ, $n = 16$), Huaxi in Guiyang (GYHX, $n = 5$), and Hezhang in Bijie (BJHZ, $n = 4$) in late August of 2015. The cultivated varieties are Zunyichaotianjiao, Huaxipinbanjiao, and Bijixianjiao, respectively in these three regions. The three planting areas were located in the north, central, and northwest of Guizhou Province (China), respectively. The Xiazi peppers in Zunyi were collected from Shenxi town, Xiazi town, Xinzhou town, Zhengchang town, and Yangchuan town. The Huaxi peppers in Guiyang were collected in Linka and Jialin villages of Machang town. The Hezhang peppers in Bijie were collected from Songlinpo town of Hezhang county. Moreover, the lithologies are argillaceous limestone, mudstone, shale, and dolomite of the middle and lower Triassic in ZYXZ pepper cultivated area. It is thin limestone of the Daye Formation of the lower Triassic in GYHX pepper planting area. The exposed stratum is the Permian Maokou Formation in BJHZ pepper planting area, and the lithology is limestone.

Pepper pre-treatment

Pepper samples were laved with running water and then rinsed with deionized water to remove the adhered dust on the peppers. Samples were dried in a drying oven at 60 °C until maintaining an unchanging weight. Afterwards, the pepper samples were ground using a fast-speed rotary mill (manufacturer and instrument type were seen in our previous papers (Zhang et al., 2018a; 2018b)) and sieved utilizing a 200-mesh nylon mesh (0.075 mm) and then placed in polyethylene plastic bags for further chemical analysis.

Multi-element determination

The ground and sieved pepper samples were analyzed in a certified laboratory (ALS Minerals-ALS Chemex Co. Ltd., Guangzhou, China) by making use of ICP-MS (7700x, Agilent, Santa Clara, CA, USA) and ICP-AES (Vista-MPX, Agilent, USA). In the process of analysis, the procedure criterion of ICP-MS and ICP-AES in the present work were set as proposed by Zhang et al. (2018b). First of all, a 0.2 g ground and sieved pepper sample was correctly weighted and added with 5 mL concentrated HNO₃ into a Teflon digestion vessel and the digestion was continued slowly at room temperature for approximately 8 h. Afterwards a heating plate was set up according to the following heat treatment, namely 60 min at 50 °C, 90 min at 100 °C, and 90 min at 150 °C. The dissolved samples were left to cool until it reached the room temperature. Subsequently, the samples were dissolved and transferred to the volumetric flask, and the digestion solutions were adjusted to a constant volume (25 mL) with 2% hydrochloric acid (HCl). Furthermore, all materials, including the Teflon digestion vessels, were soaked for 24 h with a 2% (v/v) nitric solution and then rinsed with deionized water prior to use for the sake of unexpected contaminations during the sample preparation.

With regard to the quality control of this measure process, the blanks, duplicates, and standard reference materials (SRMs) were analyzed. Especially, the SRMs including GL03 and SpL02 were measured one time, respectively. Two duplicate samples were scheduled in this assay, and accounted for 8.3% of the sum of all the pepper samples ($n = 25$). Besides, one blank sample was arranged using the same method as described for pepper sample digestion. In addition, the determined values of all elements in this blank were 1.5 folds equal or less than the limits of detection (LOD) of corresponding elements. The LOD, recoveries of SRMs, and the relative deviation (RD) of duplicates of this analytical procedure were presented in Table 1. The concentrations of 24 elements (Arsenic, Ba, Ca, Cd, Co, Cr, Cs, Cu, Fe, K, Mn, Mg, P, Mo, Ni, Na, Pb, Rb, Sb, Sn, Tl, Y, Sr, and Zn) in capsicum samples were determined in this study.

Statistical analysis

The IBM Statistics SPSS Statistics 20.0 and SIMCA-P 13.0 were used to perform the data processing. One-way ANOVA was first tested to find out significantly different elements among the three regions ($p < 0.05$). Principal component analysis (PCA), linear discriminant analysis (LDA), and orthogonal projection to latent structure-discriminant analysis (OPLS-DA) were conducted to establish the discriminative models based on the significantly

Table 1 Limits of detection (LOD), recoveries of standard reference materials (SRMs), relative deviation (RD) of duplicates, Concentration range and Mean concentrations and standard deviations of 24 elements in capsicum samples from the three regions (dry weight, mg/kg)

Element	LOD	Recoveries (%)	RD (%)	(ZYXZ, n = 16)		(GYHX, n = 5)		(BJHZ, n = 4)		The regulation limit ^c
				Min-Max	Mean $\pm$ SD	Min-Max	Mean $\pm$ SD	Min-Max	Mean $\pm$ SD	
As ^a	0.005	94.8–99.3	5.38–6.67	0.019–0.1	0.039 $\pm$ 0.025	0.051–0.136	0.079 $\pm$ 0.037	0.048–0.151	0.085 $\pm$ 0.045	0.5
Ba ^a	0.1	91.7–96.2	0.00–0.00	0.2–1	0.569 $\pm$ 0.221	0.5–0.8	0.620 $\pm$ 0.130	0.5–2.6	1.175 $\pm$ 0.974	–
Ca ^a	10	94.9–106.7	0.00–0.00	500–1300	962.5 $\pm$ 189.3	1300–1900	1480.0 $\pm$ 249.0	800–2100	1425.0 $\pm$ 623.8	–
Cd ^b	0.001	100.0–107.5	0.33–0.70	0.176–0.7	0.382 $\pm$ 0.181	0.125–0.242	0.175 $\pm$ 0.052	0.183–0.788	0.406 $\pm$ 0.270	0.05
Co ^a	0.002	92.1–102.9	0.88–3.07	0.022–0.346	0.137 $\pm$ 0.098	0.13–0.376	0.272 $\pm$ 0.099	0.176–0.844	0.346 $\pm$ 0.332	–
Cr ^a	0.05	94.6–100.0	0.26–9.57	0.2–0.63	0.338 $\pm$ 0.108	0.92–3.79	2.212 $\pm$ 1.460	0.84–1.35	1.020 $\pm$ 0.226	0.5
Cs ^b	0.001	96.2–117.4	1.96–11.11	0.002–0.052	0.0154 $\pm$ 0.013	0.015–0.024	0.019 $\pm$ 0.004	0.014–0.026	0.020 $\pm$ 0.005	–
Cu ^a	0.01	91.7–104.5	0.06–0.97	7.21–12.9	9.932 $\pm$ 1.467	10.25–13.55	11.900 $\pm$ 1.183	12.65–20.2	15.063 $\pm$ 3.556	–
Fe ^a	1	96.9–102.0	0.00–0.00	40–86	57.375 $\pm$ 11.099	67–102	84.200 $\pm$ 15.515	154–190	166.000 $\pm$ 16.573	–
K ^b	10	100.7–102.5	0.40–1.02	19,800–33,700	25,850 $\pm$ 3795	25,000–28,400	26,920 $\pm$ 1240	21,200–36,000	26,825 $\pm$ 6475	–
Mg ^b	10	91.9–106.1	0.00–0.33	1430–2130	1727 $\pm$ 181	1830–2030	1954 $\pm$ 81	1410–2230	1720 $\pm$ 372	–
Mn ^b	0.1	96.3–104.5	0.47–1.91	10.9–33.3	16.219 $\pm$ 5.186	18.2–28.3	21.800 $\pm$ 3.832	11.4–21.7	14.925 $\pm$ 4.605	–
Mo ^a	0.01	97.5–116.7	0.00–0.83	0.06–0.64	0.296 $\pm$ 0.188	0.24–3.02	1.326 $\pm$ 1.030	0.07–0.19	0.130 $\pm$ 0.064	–
Na ^b	10	94.6–100.0	0.00–0.00	20–110	38 $\pm$ 23	20–50	34 $\pm$ 13	20–60	40 $\pm$ 16	–
Ni ^a	0.02	93.5–98.0	0.38–3.65	0.4–1.43	0.729 $\pm$ 0.311	0.5–2.83	1.658 $\pm$ 1.049	1.68–4.81	2.733 $\pm$ 1.410	–
P ^a	5	94.4–102.5	0.00–0.54	2740–4860	3815 $\pm$ 525	4040–5560	4710 $\pm$ 581	3040–4990	3932 $\pm$ 879	–
Pb ^a	0.005	95.3–105.3	5.59–9.24	0.022–0.109	0.057 $\pm$ 0.026	0.046–0.076	0.056 $\pm$ 0.012	0.114–0.407	0.210 $\pm$ 0.133	0.1
Rb ^b	0.01	103.8–109.3	0.25–1.20	4.02–37.6	17.319 $\pm$ 9.246	19.95–27.3	22.290 $\pm$ 2.953	11.85–36.7	21.213 $\pm$ 11.066	–
Sb ^a	0.002	91.0–94.7	9.89–14.29	0.005–0.059	0.014 $\pm$ 0.013	0.018–0.093	0.052 $\pm$ 0.035	0.033–0.057	0.043 $\pm$ 0.010	–
Sn ^a	0.01	95.0–105.0	3.23–4.76	0.05–0.28	0.146 $\pm$ 0.072	0.18–0.72	0.488 $\pm$ 0.213	0.02–0.03	0.028 $\pm$ 0.005	–
Si ^a	0.02	102.1–107.7	0.24–4.29	0.73–10.25	3.416 $\pm$ 2.310	5.49–20.5	11.368 $\pm$ 6.329	2.06–5.21	3.108 $\pm$ 1.439	–
Tl ^b	0.001	92.3–107.5	0.00–1.41	0.006–0.105	0.032 $\pm$ 0.028	0.015–0.072	0.037 $\pm$ 0.024	0.014–0.111	0.044 $\pm$ 0.045	–
Y ^a	0.001	96.0–106.5	0.00–33.33	0.001–0.006	0.002 $\pm$ 0.001	0.001–0.006	0.003 $\pm$ 0.002	0.006–0.017	0.013 $\pm$ 0.005	–
Zn ^a	0.1	94.3–108.2	0.55–1.06	13.8–26.5	17.863 $\pm$ 3.358	21.4–28.5	24.980 $\pm$ 2.605	34.6–40.2	37.950 $\pm$ 2.461	–

^aThe element concentrations of peppers were significantly different among the three regions ($p < 0.05$)^bThe element concentrations of peppers were not significantly different among the three regions ($p > 0.05$)^cThe regulation limits (fresh weight) quoted the food safety standard limit in China (GB 2762-2017)

different elements ($p < 0.05$). Predictive abilities of LDA and OPLS-DA models were evaluated by the leave-one-out cross-validation.

Results and discussion

Differences of the concentrations of elements in peppers

The statistical results of 24 mineral elements (Arsenic, Ba, Cd, Co, Cr, Sb, Pb, Sn, Rb, Sr, Tl, Cs, Ca, Fe, K, Mg, Mn, P, Na, Mo, Y, Ni, Zn, and Cu) in 25 capsicum samples from the three regions (ZYXZ, GYHX, and BJHZ) were shown in Table 1. Based on the Kolmogorov–Smirnov test result, the 24 mineral elements contents in 25 pepper samples fulfilled a normal school ($p > 0.05$). The mean contents of Cadmium, Cs, Tl, Rb, Na, Mn, Mg, and K were not significantly different among these three regions ($p > 0.05$) according to the one-way ANOVA result. Therefore, these eight elements were eliminated in the further chemometrics data processing of geographical origin discrimination. On the contrary, there were significant difference for 16 elements (Arsenic, Ba, Co, Cr, Pb, Sb, Sn, Y, Ni, Cu, Sr, Mo, Zn, Fe, Ca, and P) concentrations in capsicum samples among the three regions ($p < 0.05$), which can be used for the further data process of discrimination research. Based on the food safety limits in China (GB 2762-2017) after the conversion of average dry matter (DW) rate (23.16%), only Cd and Cr concentrations exceeded the standard limit values in these three areas, and the overall exceeding rate was 60.0% and 8.0%, respectively ($n = 25$). In addition, the mean concentrations of Arsenic, Cd, Pb, Zn, and Fe in peppers from BJHZ were higher than the other two regions (GYHX and ZYXZ). This may be related to the large-scale indigenous zinc smelting in Hezhang county, Bijie city (Guizhou, China). The mean concentration of Sr in pepper samples from GYHX was the highest in these three regions, which was caused by the geological background of the thin limestone of the Triassic Daye Formation.

As indicated by the Duncan's multiple comparison, there was their own element concentration characteristic for each region. Particularly, for the pepper samples from BJHZ, the concentrations of Arsenic, Ba, Co, Cu, Fe, Pb, Ni, Y, and Zn were the highest among the three regions, but the concentrations of molybdenum, Sn, and Sr were the lowest. The peppers from GYHX were generally characterized by the highest concentrations of Calcium, Cr, Mo, P, Sb, Sn, and Sr and the lowest of Pb. Besides, the samples from ZYXZ had the lowest contents of Arsenic, Ba, Co, Cr, Cu, Ni, Sb, Y, Zn, Ca, P, and Fe. Moreover, the standard deviations of some elements were relatively large,

reflecting the great variability among peppers from the same zone, and this was analogous to the work of Ma et al. (2016). The reason was probably that the collected capsicum cultivars and the elemental concentrations in soils were different, influencing elemental levels in peppers (Chung et al., 2015; Zhao et al., 2017). However, the significant differences of element concentrations in capsicum samples provided reliable variables for further statistical analyses and made it possible to discriminate capsicums from different regions.

Principal component analysis (PCA)

Some agricultural products samples (rice, tea, and sesame seed) from different provenances were frequently classified correctly based on multi-element concentrations with significant differences and PCA (Choi et al., 2017; Chung et al., 2015; Ma et al., 2016). For the sake of estimating the difference of peppers from ZYXZ, GYHX, and BJHZ, 16 significantly different elements ($p < 0.05$) were conducted by principal component analysis (PCA). As a result, a four-factor model (the first four PCs with eigenvalues > 1) was established, which explained up to 85.238% of the total variance in the original data. The dominant variables for the first PC (PC1) were Arsenic, Ba, Co, Cr, Cu, Y, Zn, Ni, Sb, Pb, Ca, and Fe, representing 42.992% of the total variance. The second PC (PC2) represented 23.558% of the total variance, and it was strongly associated with the values of Chromium, Mo, P, Pb, Sn, and Sr. The results of the data matrix indicated that Arsenic, Co, and Pb were the dominant variables on the third PC (PC3), accounting for 11.621% of the total variance. The concentration of P had the highest weight on the fourth PC (PC4), which accounted for 7.067% of the total variability. Therefore, the elements Arsenic, Ba, Co, Pb, Sn, Sb, Y, Cr, Ni, Zn, Cu, Sr, Mo, P, Ca, and Fe can be used as the characteristic elements in the peppers. These sixteen elements may be treated as the most powerful indicators of pepper samples.

The scatter plots of the scores of PC1-PC2 and PC1-PC3 were shown in Fig. 1A, B. As can be seen from Fig. 1A, B, 66.550% and 54.613% of the total variability were explained for PC1-PC2 and PC1-PC3, respectively. Moreover, it was shown that all capsicum samples from three regions clustered into three groups (capsicums from ZYXZ, capsicums from GYHX, and capsicums from BJHZ) in the scatter plot of PC1 versus PC2 (Fig. 1A). Capsicums from ZYXZ were separated from GYHX and BJHZ in two-dimensional scatter plot of PC1 versus PC3, while the distribution of pepper samples from GYHX was a little overlapped with BJHZ samples (Fig. 1B), and it may be because Arsenic, Co, and Pb have higher weights in PC1 and PC3 and the explaining information cannot be comprehensive and enough. The classified results on the basis

Fig. 1 Score scatter plots of the first three principal components (A: PC1 vs PC2; B: PC1 vs PC3), scores scatter plot of the constructed standardized two discriminant functions (function 1 and function 2) of the LDA model based on the significantly different element concentrations (C), scores scatter plot of the two principal components of the OPLS-DA model (D), Variable importance in the projection (VIP) obtained from the OPLS-DA model for geographical origin discrimination of pepper samples (E), black indicated that the VIP values were more than one, and grey indicated that the VIP values were less than one

of PCA gave a well-performed result to distinguish capsicum samples from three regions in the scatter plot of PC1 versus PC2 (Fig. 1A). Nevertheless, there were overlaps for samples from GYHX and BJHZ in the scores scatter plot of PC1 versus PC3 (Fig. 1B). Similarly, the phenomenon of strong overlap was achieved in the studies of tea (Ni et al., 2018) and *Agaricus bisporus* mushroom (Chung et al., 2018b) traceability based on PCA. Likewise, the PCA could not trace rice samples collected from China and Korea apart from the rice samples in the Philippines and other zones (Chung et al., 2015). Therefore, it was suggested that based on these findings further complex chemometric algorithms should be used to realize the geographical origin discrimination of the pepper samples better.

Linear discriminant analysis (LDA)

The linear discriminant analysis (LDA) technique is a data-processing method with a characteristic of the supervised pattern recognition. In a supervised pattern recognition, samples were classified into groups clusters with pre-determined models for the class. This approach differed from some unsupervised data-processing tools, e.g., cluster analysis (CA) and PCA in which the prior classes are not set up (Kara, 2009). Some investigators already published some satisfactory discrimination results based on LDA (Bertoldi et al., 2016; Kara, 2009). For obtaining better identification results of capsicums from the three different regions, linear discriminant analysis (LDA) was performed on the basis of the concentrations of 16 significantly different elements ($p < 0.05$). In consequence, three metal elements (Iron, Cr, and Sn) were picked out and two

discriminant functions were established on the basis of the Wilks' lambda values. These two discriminant functions gave an explanation of 100% of the variance (86.6% for function 1 and 13.4% for function 2), and the Wilks' lambda values of function 1 and function 2 were 0.015 and 0.273, respectively ($p_1 = 0.000$ and $p_2 = 0.000$). The canonical correlations were 0.972 for function 1 and 0.853 for function 2. It was indicated that there were significant variabilities among the peppers from these three different original places. The constructed discriminant functions were listed below.

Function one = $1.220\text{Fe} - 0.513\text{Cr} - 0.578\text{Sn}$

Function two = $0.155\text{Fe} + 0.568\text{Cr} + 0.604\text{Sn}$

The separation of capsicum samples from ZYXZ, GYHX, and BJHZ was verified by plotting the two discriminant functions scores scatter plot (Fig. 1C). Obviously, capsicum samples from the three regions were effectively differentiated from each other. Consequently, the technology of multi-element concentrations assay was potential to authority the origin of capsicum and was a robust method to distinguish the geographic origin of capsicum.

The constructive discriminant functions gained the recognition ability (% of the objects belonging to the training set correctly classified) of 92.0% in accordance to 16 significantly different elements ($p < 0.05$). In particular, the capsicum samples from ZYXZ and BJHZ were 100% correctly classified, whereas two capsicum samples from GYHX were incorrectly assigned as samples from ZYXZ (Table 2). Similarly, a good discrimination results of the tomatoes with 99.9% of correctly classified cases was obtained based on LDA in the research of Fragni et al. (2015). To assess the predictive ability, the created model was then validated through the leave-one-out cross-validation method. As a result, the predictive capacity of the constructed model (% of the objects belonging to the

testing set correctly classified) was 92.0% as well, which indicated that the LDA model performed satisfactorily for the separation of capsicums from three origins in Guizhou Province, China. The established Fisher's linear discriminant functions for each region were listed below:

Group one (ZYXZ) = $0.460\text{Fe} - 3.820\text{Cr} - 4.725\text{Sn} - 12.655$

Group two (GYHX) = $0.432\text{Fe} + 0.320\text{Cr} + 21.241\text{Sn} - 25.330$

Group three (BJHZ) = $1.461\text{Fe} - 11.084\text{Cr} - 53.034\text{Sn} - 116.704$

A high percentage of objects correctly classified represented the stability and the strong connection between the derivative profiles and the origins. It was shown that a high predictive percentage of the validated set signified the high potential to separate the geographical origin of unknown samples on the basis of the relative relationship of these three elements (Iron, Cr, and Sn). Accordingly, it was illustrated that capsicums from the three zones in Guizhou Province can be classified efficiently in accordance with the LDA. In spite of a high classification rate of LDA (92.0%), another multivariate statistical analysis method should be attempted to performed with the purpose of reducing the number of erroneous classified samples.

Orthogonal partial least squares discriminant analysis (OPLS-DA)

The OPLS-DA model was developed on the basis of partial least squares discriminant analysis (PLS-DA). Moreover, in comparison to PLS-DA, in the OPLS-DA model system variation from the X variate was caused decomposition into two portions, including the portion of orthogonal to Y axial and the portion of linearly related to Y axial. The OPLS-DA will provide the more explanation and reduce the error of result with the enhancement of orthogonal variation components (Trygg and Wold, 2002). The OPLS-DA was performed to differentiate these pepper samples from three regions in line with the concentrations of 16 significantly

Table 2 Traceability results and percentage correctly classified of pepper samples from three regions by LDA and OPLS-DA

Geographical origin	Model	Examination samples	Predicted group membership			Correctly classified (%)
			ZYXZ	GYHX	BJHZ	
ZYXZ	LDA	16	16	0	0	100
GYHX		5	2	3	0	60.0
BJHZ		4	0	0	4	100
Total		25				92.0
ZYXZ	OPLS-DA	16	16	0	0	100
GYHX		5	1	4	0	80.0
BJHZ		4	0	0	4	100
Total		25				96.0

different elements ($p < 0.05$). As a result, for the model parameters, the R^2X was equal to 0.664, indicating that the two principal components (PC) have 66.40% elucidation ability for the variation of X variables. Besides, R^2Y was equal to 0.801, which revealed that the two predictive PCs of this model had the explaining power of 80.10% for the variation of Y variables. Additionally, Q^2 was equal to 0.744, which denoted that the two predictive PCs of this model had the forecast power of 74.40% for the peppers from the three regions (GYHX, ZYXZ, and BJHZ). To sum up, the results of OPLS-DA model were reliable in this work. The first, second principal component scores of pepper samples from three cultivated regions in OPLS-DA model constructed by 16 chemical indicators were shown in Fig. 1D ($p < 0.01$). It can be seen that the capsicums from each region were obviously aggregated, and that the peppers from the three regions (GYHX, ZYXZ, and BJHZ) can be well separated from each other.

The validation set back test of the OPLS-DA model was conducted according to 25 pepper samples of the training set to identify the correct rate. In the OPLS-DA model, the leave-one-out was automatically used as cross-validation to obtain the misclassification result. The validation results were summarized in Table 2. For the pepper samples, the rate of correct discrimination of this OPLS-DA model from the three regions was 96.0%, and one capsicum sample from GYHX was mistakenly assigned as sample from ZYXZ, denoting a very satisfactory result. The variable importance in the projection (VIP) values of OPLS-DA model were visible in Fig. 1E. The variable importance in the projection (VIP) was a weighted sum of squares of the PLS weight (Chung et al., 2015). A VIP value was more than 1, which revealed that the corresponding variable was important to the discrimination (Chung et al., 2015). Therefore, In the OPLS-DA model, seven elements (Tin, Fe, Zn, Y, Cr, Sr, and Mo) were regarded as significant indicators for discriminating the geographical origin of capsicums in accordance to their VIP values of > 1 (Fig. 1E). Furthermore, it was also found in the present study that the rate of correct discriminant of the OPLS-DA model (existing one misclassified sample, 96.0%) was higher than the LDA model with two misclassified samples (92.0%), showing that the OPLS-DA model was more reliable and ascendant than the LDA model for discriminating pepper samples from these three regions.

It was demonstrated that the cultivars of agricultural products had influences on the correct identification rates of their geographical origin traceability (Zhao et al., 2012). In addition, it was confirmed that stable isotope ratios ($\delta^{13}C$, $\delta^{15}N$, $\delta^{18}O$, δ^2H , $\delta^{34}S$, and $^{87}Sr/^{86}Sr$) could be efficacious and befitting chemical markers for the regional discrimination of pepper samples based on multivariate statistical approaches (Opatic et al., 2017; Rijke et al.,

2016; Song et al., 2014). Thereby, in subsequent research, the same pepper cultivar should be taken into consideration for geographical origin discrimination of peppers. Likewise, stable isotopes compositions of pepper samples, e.g., $\delta^{13}C$, $\delta^{15}N$, $\delta^{18}O$, δ^2H , $\delta^{34}S$, and $^{87}Sr/^{86}Sr$ may be determined to seek some new geographical origin differentiation indicators for regional discrimination of pepper in Guizhou Province. Simultaneously, although a successful classification was obtained, a supplementary study covering lots of pepper samples should be allowed to establish better discrimination models of pepper samples from different regions. Moreover, it was indicated that other chemometrics tools, e.g., support vector machine (SVM), random forest (RF), and k-nearest neighbors (k-NN), were applied to classify grapes seeds indicating the correct classification rates of k-NN, RF, and SVM were 85.7%, 98.3%, and 93.3%, respectively (Canizo et al., 2018). Furthermore, twenty-three elements, twelve metal isotope compositions, and four stable isotope ratios ($\delta^{13}C$, $\delta^{15}N$, δ^2H , and $\delta^{18}O$) combined with PCA, LDA, PLS-DA, and Decision-making tree (DT) were used to distinguish the geographical origin of green tea from China and an accuracy rate of 90.0% (blind dataset, $n = 107$) was obtained by means of the DT tool (Ni et al., 2018). As a result, these chemometrics means (k-NN, SVM, RF, and DT) may be also used to discriminate Guizhou pepper coupled with some significantly different indicators of pepper in our future studies.

In conclusion, the significant difference existed for the concentrations of 16 mineral elements (Arsenic, Ba, Co, Cr, Ni, Pb, Sb, Sn, Y, Mo, Sr, Cu, Zn, Fe, Ca, and P) from the three regions (GYHX, ZYXZ, and BJHZ) in Guizhou province, China ($p < 0.05$). Based on the 16 significantly different ($p < 0.05$) elements associated with the LDA and OPLS-DA models the pepper samples from these three regions (GYHX, ZYXZ, and BJHZ) can be well discriminated from each other. However, the result of PCA was not very satisfactory. In addition, the correct classification rate of OPLS-DA model (96.0%) was higher than the LDA model (92.0%). All things considered, multi-elemental concentrations combined with chemometric data-processing may be a satisfactory and feasible tool for the geographical origin traceability of peppers from different regions in Guizhou.

Acknowledgements This research was financially supported by the National Natural Science Foundation of China (Grant No. 41463009), the Innovation Group Major Research Project of Guizhou Province Education Department (Grant No. KY[2016]024), the Construction Project of the First-Class Subjects (Ecology) in Guizhou Province (Grant No. GNYL[2017]007), the Important and Special Project (Tea) of Guizhou Province Science and Technology Department and the Graduate Innovation Foundation Project of Guizhou Province Education Department (QJH-YJSCXJH-2018-049). The authors express heartfelt thanks to Dr. Jun Chen who participated in collecting the

samples. Thanks to anonymous reviewers and their constructive comments.

Compliance with ethical standards

Conflict of interest There was no conflict of interest among the authors.

References

- Ariyama K, Horita H, Yasui A. Application of inorganic element ratios to chemometrics for determination of the geographic origin of Welsh onions. *J. Agric. Food Chem.* 52(19): 5803–5809 (2004)
- Bertoldi D, Barbero A, Camin F, Caligiani A, Larcher R. Multielemental fingerprinting and geographic traceability of *Theobroma cacao* beans and cocoa products. *Food Control* 65: 46–53 (2016)
- Brunner M, Katona R, Stefanka Z, Prohaska T. Determination of the geographical origin of processed spice using multielement and isotopic pattern on the example of Szegedi paprika. *Eur. Food Res. Technol.* 231(4): 623–634 (2010)
- Canizo BV, Escudero LB, Perez MB, Pellerano RG, Wuilloud RG. Intra-regional classification of grape seeds produced in Mendoza province (Argentina) by multi-elemental analysis and chemometrics tools. *Food Chem.* 242: 272–278 (2018)
- Chang CT, You CF, Aggarwal SK, Chung CH, Chao HC, Liu HC. Boron and strontium isotope ratios and major/trace elements concentrations in tea leaves at four major tea growing gardens in Taiwan. *Environ. Geochem. Health* 38(3): 737–748 (2016)
- Choi YH, Hong CK, Kim M, Jung SO, Park J, Oh YH, Kwon JH. Multivariate analysis to discriminate the origin of sesame seeds by multi-element analysis inductively coupled plasma-mass spectrometry. *Food Sci. Biotechnol.* 26(2): 375–379 (2017)
- Chung IM, Kim JK, Lee JK, Kim SH. Discrimination of geographical origin of rice (*Oryza sativa* L.) by multielement analysis using inductively coupled plasma atomic emission spectroscopy and multivariate analysis. *J. Cereal Sci.* 65: 252–259 (2015)
- Chung IM, Kim JK, Lee KJ, Lee KJ, Park SK, Lee JH, Son NY, Jin YI, Kim SH. Geographic authentication of Asian rice multielemental and stable isotopic data combined with multivariate analysis. *Food Chem.* 240: 840–849 (2018a)
- Chung IM, Han JG, Kong WS, Kim JK, An MJ, Lee JH, An YJ, Jung MY, Kim SH. Regional discrimination of *Agaricus bisporus* mushroom using the natural stable isotope ratios. *Food Chem.* 264: 92–100 (2018b)
- Fragini R, Trifiro A, Nucci A. Towards the development of a multi-element analysis by ICP-oe-TOF-MS for tracing the geographical origin of processed tomato products. *Food Control* 48: 96–101 (2015)
- Kara D. Evaluation of trace metal concentrations in some herbs and herbal teas by principal component analysis. *Food Chem.* 114(1): 347–354 (2009)
- Ma GC, Zhang YB, Zhang JY, Wang GQ, Chen LY, Zhang ML, Liu T, Liu X, Lu CY. 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 (2016)
- Naccarato A, Furia E, Sindona G, Tagarelli A. Multivariate class modeling techniques applied to multielement analysis for the verification of the geographical origin of chili pepper. *Food Chem.* 206: 217–222 (2016)
- Ni K, Wang J, Zhang QF, Yi XY, Ma LF, Shi YZ, Ruan JY. Multi-element composition and isotopic signatures for the geographical origin discrimination of green tea in China: A case study of *Xihu Longjing*. *J. Food Compos. Anal.* 67: 104–109 (2018)
- Opatic AM, Necemer M, Lojen S, Vidrih R. Stable isotope ratio and elemental composition parameters in combination with discriminant analysis classification model to assign country of origin to commercial vegetables-A preliminary study. *Food Control* 80: 252–258 (2017)
- Opatic AM, Necemer M, Budic B, Lojen S. Stable isotope analysis of major bioelements, multi-element profiling, and discriminant analysis for geographical origins of organically grown potato. *J. Food Compos. Anal.* 71: 17–24 (2018a)
- Opatic AM, Necemer M, Lojen S, Masten J, Zlatic E, Sircelj H, Stopar D, Vidrih R. Determination of geographical origin of commercial tomato through analysis of stable isotopes, elemental composition and chemical markers. *Food Control* 89: 133–141 (2018b)
- Palacios-Morillo A, Jurado JM, Alcazar A, de Pablos F. Geographical characterization of Spanish PDO paprika by multivariate analysis of multielemental content. *Talanta* 128: 15–22 (2014)
- Pino J, Gonzalez M, Ceballos L, Centurion-Yah AR, Trujillo-Aguirre J, Latournerie-Moreno L, Sauri-Duch E. Characterization of total capsaicinoids, colour cultivars grown in Yucatan. *Food Chem.* 104(4): 1682–1686 (2007)
- Rashmi D, Shree P, Singh DK. Stable isotope ratio analysis in determining the geographical traceability of Indian wheat. *Food Control* 79: 169–176 (2017)
- Rijke ED, Schoorl JC, Cerli C, Vonhof HB, Verdegaaal SJA, Vivo-Truyols G, Lopatka M, Dekter R, Bakker D, Sjerps MJ, Ebskamp M, de Koster CG. The use of $\delta^2\text{H}$ and $\delta^{18}\text{O}$ isotopic analyses combined with chemometrics as a traceability tool for the geographical origin of bell peppers. *Food Chem.* 204: 122–128 (2016)
- Song BY, Ryu JS, Shin HS, Lee KS. Determination of the source of bioavailable Sr using $^{87}\text{Sr}/^{86}\text{Sr}$ tracers: a case study of hot pepper and rice. *J. Agric. Food Chem.* 62(38): 9232–9238 (2014)
- Trygg J, Wold S. Orthogonal projections to latent structures (OPLS). *J. Chemometrics* 16(3): 119–128 (2002)
- Wu SX, Zeng JM, Xie H, Ng SH. Capsaicin determination and chili sauce discrimination using low cost and portable electrochemical sensors based on all graphite pencil electrodes. *Anal. Method* 8(39): 7025–7029 (2016)
- Yin XQ, Xu XX, Zhang Q, Xu JG. Rapid determination of the geographical origin of Chinese red peppers (*Zanthoxylum Bungeanum* Maxim.) based on sensory characteristics and chemometric techniques. *Molecules* 23: 1001 (2018)
- Zhang J, Yang RD, Chen R, Li YC, Peng YS, Liu CL. Multielemental analysis associated with chemometric techniques for geographical origin discrimination of tea leaves (*Camelia sinensis*) in Guizhou province, SW China. *Molecules* 23(11): 3013 (2018a)
- Zhang J, Yang RD, Chen R, Peng YS, Wen XF, Gao L. Accumulation of heavy metals in tea leaves and potential health risk assessment: a case study from Puan county, Guizhou province, China. *International J. Environ. Res. Public Health* 15(1): 133 (2018b)
- Zhao HY, Guo BL, Wei YM, Zhang B. Effects of wheat origin, genotype, and their interaction on multielement fingerprints for geographical traceability. *J. Agric. Food Chem.* 60(44): 10957–10962 (2012)
- Zhao HY, Zhang SL, Zhang ZW. Relationship between multi-element composition in tea leaves and in provenance soils for geographical traceability. *Food Control* 76: 82–87 (2017)

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