DETERMINATION OF GEOGRAPHIC ORIGIN OF SENGCU RICE USING MULTIELEMENT AND MULTIVARIATE ANALYSIS

Pham Quoc Trung, Le Truong Giang
Institute of Chemistry, Vietnam Academy of Science and Technology
Dao Hai Yen*
Graduate University of Science and Technology, Vietnam Academy of Science and Technology

TÓM TẮT

XÁC ĐỊNH NGUỒN GỐC ĐỊA LÝ GẠO SÉNG CÙ SỬ DỤNG PHƯƠNG PHÁP PHÂN TÍCH ĐA NGUYÊN TỐ VÀ THUẬT TOÁN ĐA BIẾN

Phân tích đa nguyên tố kết hợp với thuật toán đa biến được sử dụng để đánh giá khả năng truy xuất nguồn gốc địa lý gạo Séng Cù ở một quy mô nhỏ (bán kính <30 km) của tỉnh Lào Cai. Hàm lượng 16 nguyên tố kim loại trong gạo Séng Cù được xác định bằng phương pháp ICP-MS. Kết quả phân tích thành phần chính (PCA) thu được đã chỉ ra mức độ phân tách sơ bộ của các mẫu gạo. As, Ba Sr, Cr, Zn và Cd là các nguyên tố chính trong truy xuất nguồn gốc địa lý mẫu gạo Séng Cù được canh tác ở ba cánh đồng của Lào Cai bao gồm Mường Vi (SC-MV), Bản Qua (SC-BQ) và Bản Xen (SC-BX). Trong mô hình phân tích phân biệt tuyến tính (LDA) và phân tích phân biệt bình phương nhỏ nhất (PLS-DA), tỷ lệ nhận dạng chính xác tổng thể về nguồn gốc địa lý của các mẫu gạo Séng Cù là 100% trong cả tập dữ liệu huấn luyện và thẩm định. Kết quả trên cho thấy nghiên cứu này chứng tỏ khả năng phân biệt nguồn gốc địa lý của gạo Séng Cù trong phạm vi nhỏ của tỉnh Lào Cai. Đây là điều kiện tiên quyết để xác định nguồn gốc xuất xứ gạo Séng Cù và các thương hiệu gạo cao cấp như ST24, ST25, Hương Lai,... trên phạm vi địa lý lớn hơn.
Từ khóa: *gạo Séng Cù, nguồn gốc địa lý, phân tích đa biến, đa nguyên tố, phân tích thành phần chính.*

1. INTRODUCTION

In recent years, with the development of international markets and the demand for high-quality agricultural products, the safety and authenticity of foodstuffs have become major concerns for consumers. Rice with geographical origin labels, for example, Jasmine rice from Thailand, Koshihikari from Japan, and Basmatice from Pakistan is considered high-quality and is usually high-priced[1]. Besides, it helps prevent fake and inferior products to protect consumers' rights and improve the credibility of producers and traders in light of the increase in global trade and free markets[2].

Sengcu rice was known as a specialty of Lao Cai province, Vietnam. This rice with high nutritional contents such as vitamin E, vitamin B1, B3, B6, fiber, etc. is 4-6 times higher than other types of rice[1, 3]. Because of the high price, many traders often defraud Sengcu rice by mixing it with other rice types and adding flavor[4]. Therefore, brand-name protection and distinction of Sengcu rice is essential.

In recent years, the most common techniques used for food authenticity and traceability include isotope ratio; liquid and gas chromatography; elemental analysis, spectroscopic techniques, DNA-based techniques, and sensor techniques[5, 6].In terms of multi-elemental method, inductively coupled plasma mass spectrometry (ICP-MS) is a powerful analytical technique useful for geographical origin information by the analysis of

several inorganic elements that serve as the fingerprint for geographical traceability. For example, geographical identifications of Chianti red wine[7], Wuchang rice [8], Asian rice[9] have been carried out with correct rate stood at above 95% by multi-element analysis combined with multivariate analysis. It can be seen that the accuracy of the multi-elemental methods is very high. Moreover, compared with some other methods such as isotope ratio mass spectrometry, fingerprint chromatography, the multi-element analysis method has many advantages such as low cost, ease of use and data processing[6, 10, 11].

In food authenticity research, chemometrics plays an important role in classifying products based on their origin, variety, or other properties[2]. The most common multivariate analysis method for verification purposes includes two main gender algorithms: unsupervised and supervised methods [12]. For the former, principal component analysis (PCA) is known as a useful tool for reducing database and screening preliminary[2, 6]. For the later, linear discriminant analysis (LDA), and partial least square discriminant analysis (PLS-DA) are major classification algorithms [6]. These methods have been used in many previous publications with high sensitivity and accuracy[11, 13-15].

Currently, much research only carried out traceability rice, wine, honey, ginseng,.. among distant locations. Hence, in this study, we aimed to: (i) measuring the multi-elements in Sengcu rice, and assessing the contamination level of heavy metals in Sengcu rice; (ii) constructing a model using multivariate analysis to discriminate geographical origins of Sengcu rice among three cultivating fields in Lao Cai province.

2. MATERIALS AND METHODS

2.1. Reagents and samples

Nitric acid (HNO_3) 65% and hydrogen peroxide (H_2O_2) 30% solution were purchased from Sigma-Aldrich (Merck, Darmstadt, Germany). Four-teen element calibration standard solutions including Al, As, Ba, Cd, Cr, Cu, Fe, Mn, Ni, Pb, Rb, Se, Sr, Zn were provided from CPA Chem (CPAchem Ltd, Stara Zagora, Bulgaria). Ultrapure water ($18.2 M\Omega.cm^{-1}$) was produced by the Millipore-Q system (Merck Millipore, Burlington, Massachusetts, United States). Certified reference material NMIJ CRM 7502-a (rice) was provided from the National Metrology Institute of Japan (NMIJ) (National Institute of Advanced Industrial Science and Technology, Tokyo, Japan) and was used to evaluating the accuracy of the analysis method.

2.2. Sample collection and digestion

A total of 30 Sengcu rice samples were collected from three fields (Muong Vi, Ban Qua, and Ban Xen) of Lao Cai province from 2019 to 2020.

All rice samples were lyophilized at −45°C for 3 days and then ground with a pulverizer to obtain find powder (<400 µm particle size) before storing in clean plastic bags. Rice samples (0.25 g) were weighed into a polytetrafluoroethylene (PTFE) vessel and then 4mL H_2O, 2mL HNO_3, and 2mL H_2O_2 were added. After 30 min, the sample was digested in a Mars X-press plus microwave digestion system (CEM, NC, USA). The digestion program was as follows: the sample was heated to 120ºC for 15min, held for 10min, before heating to 160ºC for 10min, held for 10min, and finally heated to 180ºC within 10min and held for 30min. After cooling, the solution was diluted to 25mL in a volumetric flask with ultrapure water.

2.3. Instrumental parameters

The concentration of 16 elements was determined by a quadrupole inductively coupled plasma mass spectrometer (ICP-MS) (iCap TQ, Thermo Scientific, USA) working under the following operating conditions: RF power, 1.550 kW; plasma argon flow rate, 14 L/min; auxiliary Ar flow rate, 0.8L/min; carrier Ar flow rate, 1.0L/min; sample depth, 9.0mm; spray interface temp, 2ºC; sample flow rate, 400µL/min. Additionally, to obtain the best sensitivity (^{115}In 25.000 cps/ppb), minimal formation of oxides ($^{156}CeO^+/^{140}Ce^+< 1\%$) and doubly charged ions ($^{138}Ba^{2+}/^{138}Ba^+< 3\%$), the source parameters (gas flows and ion lens voltages) were optimised as recommended by Thermo Scientific . An external calibration curve method was used for the quantitation of all elements in rice. The instrument was run under a linear multipoint calibration 0.05-50 µg/L (trace elements), and 20-200 µg/L (micro elements). Dilution method with deionized water is used to determine the content of Mg,Ca.

2.4. Method validation (QA/QC)

In this study, instrument detection limits (IDL) were calculated using the raw intensity data from the standard and the blank sample (using ultrapure 2% nitric acid matrix) as per the

following equation: $IDL = 3SD_{blank} \times C_x/(S_x - S_{blank})$ where SD_{blank} is the standard deviation of the intensities of the multiple blank measurements, C_x is the mean signal for the standard, and S_x, S_{blank} are signal for C_x and blank sample. Method detection limits (MDLs) were calculated as followed: MDL = IDL x constant volume/sample weight.

The CRM of white rice flour (NMIJ CRM7502-a) was digested according to the above method to verify the precision and accuracy procedure.

2.5. Data analysis

First, one-way analysis of variance (ANOVA) was used to evaluate the difference between each element in rice among different regions. This eliminates factors that do not differ statistically between regions. In this study, analysis of variance was performed using Minitab 18.0 software. Furthermore, the multivariate analysis of data obtained from ICP-MS was performed by using Origin 2018 pro-software. The principal component analysis (PCA) was performed for exploring the preliminary visual classification groups of the database. Partial least squares-discriminant analysis(PLS-DA) and linear discriminant analysis (LDA) were carried out to evaluate whether rice samples from different regions could be classified based on elements that had significant differences among the regions. The

data of rice samples were divided into training (24 samples) and validation sets (6 samples).

3. RESULTS AND DISCUSSION

3.1. Method validation

The MDL value has been estimated following section 2.4 for each element, with sample volume 25mL, sample mass 0.25 g. As can be observed in Table 1, the MDL values fluctuated from 0.109 to 172.61 µg/kg, which was suitable for the determination of trace elements in rice.

The repeatability of the method was calculated as the relative standard deviation (%RSD) of the 7 repeated digestion times CRM sample. The RSD values obtained ranging from 2.0-9.5% (<15%), indicating that the method has good repeatability.

The CRM of NMIJ CRM 7502-a reference material (rice flour) was digested and determined according to the above method to verify the accuracy of the method. The results are shown in Table 1, the elemental concentrations obtained from the experiment agreed well with the certified values. Furthermore, the mean recoveries of all the elements at three-level spike (0.2; 0.5; 2.0 mg/kg) ranged from 90 to 104%.

The method used to determine the content of the elements in rice has high reliability and repeatability.

Table 1. Method detection limits (MDL), Limit of Quantitation (LOQ), precision and accuracy of the method for determining the elements in rice (n=3)

Elements	MDL (µg/kg)	LOQ (µg/kg)	RSD (%)	Recovery (%)	CRM(mg/kg)	Measured values (mg/kg)
Al	5.30	17.6	7.88±5.62	102.46±4.11	-	-
Cr	0.17	0.58	2.64±1.10	102.06±4.95	0.075±0.013	0.080±0.021
Mn	0.11	0.36	5.90±3.15	96.15±3.25	11.2±0.4	10.7±0.5
Fe	0.49	1.62	9.47±4.65	90.05±6.79	4.48±0.20	4.68±0.72
Ni	1.76	5.85	6.93±7.21	98.38±4.45	0.390±0.022	0.413±0.019
Cu	0.63	2.10	4.63±4.05	95.17±8.52	3.02±0.11	3.10±0.36
Zn	0.42	1.40	6.20±4.85	91.15±4.62	26.0±0.9	24.0±0.8
As	0.48	1.61	5.08±3.32	95.63±6.66	0.109±0.005	0.118±0.007
Se	0.32	1.08	8.52±5.43	104.25±6.92	-	-
Rb	0.52	1.74	6.20±3.36	102.32±6.84	1.77±0.07	1.67±0.08
Sr	0.17	0.59	2.22±0.58	96.37±9.37	0.068±0.003	0.113±0.007
Cd	0.39	1.30	3.33±2.73	103.09±3.83	0.548±0.020	0.548±0.033
Ba	0.90	3.02	2.36±1.32	101.14±5.74	0.137±0.005	0.175±0.007
Pb	0.38	1.27	2.01±1.91	95.11±5.01	-	-

Concentration elements in Sengcu rice

Table 2 shows the profile for 14 elements in Sengcu rice from Muong Vi, Ban Qua, and Ban Xen fields. In the present study, 14 elements were classified as micro-elements (Cu, Fe, Mn, Zn), and other elements.

Figure 1 and Table 2 demonstrate the fluctuation content of micro-elements and other elements in Sengcu. Generally, the total concentration of micro-elements (Al, Mn, Fe, Cu, Zn, Rb) accounted for around 5% of 16 elements. Looking more detail, the mean concentration of Cr from SC-MV was nearly 446 µg/kg, and the highest, while the figures of Sengcu rice origin BX was the lowest (approximately 117 µg/kg). Standing at 294 µg/kg, the mean of As of SC-BX was the highest while that of rice samples from other regions ranged from 72 to 212 µg/kg. The average amount of Ni in SC-MV was 424

µg/kg, which was two and four times higher than that of SC-BQ and SC-BX, respectively. A similarity in the concentration of Cu was witnessed between SC-MV and SC-BQ, which was higher than that of SC-BX. This may be because these two fields are located close to the Sin Quyen copper mine. Remarkably, the content of Pb, Cd, As in Sengcu rice were below the maximum allowable limit in rice as required by EU Commission regulation (EC-1881/2006); China (GB 2762-2012); Codex (193-1995).

As compared by ANOVA test, 14 elements in Sengcu rice samples were significantly different among the regions. Since there was no significant difference in the mean concentration of Mg and Fe among Muong Vi, Ban Qua, and Ban Xen ($p > 0.05$), these elements were not included in any downstream analyses.

Table 2. Mineral elements content in Sengcu from three regions (µg/kg)

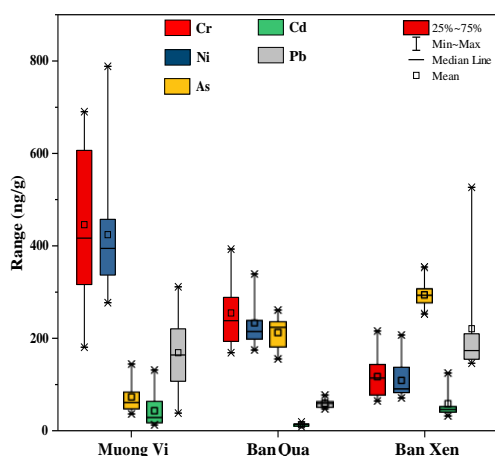| Elements	Muong Vi field			Ban Qua field			Ban Xen field		
	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
Al	**5.80**[a]	**22.90**[a]	**13.65**[a]	**5.23**[a]	**19.58**[a]	**8.55**[a]	**4.45**[a]	**9.72**[a]	**6.53**[a]
Cr	180.62	690.14	445.722	168.79	392.77	254.579	66.32	214.49	117.78
Mn	**13.11**[a]	**22.16**[a]	**17.12**[a]	**16.34**[a]	**27.87**[a]	**19.23**[a]	**14.76**[a]	**17.40**[a]	**16.01**[a]
Fe	**10.99**[a]	**48.18**[a]	**28.36a**	**13.99**[a]	**51.62**[a]	**21.01**[a]	**17.84**[a]	**25.07**[a]	**21.12**[a]
Ni	276.93	788.32	424.017	175.81	339.12	233.88	70.93	205.84	108.547
Cu	**1.76**[a]	**3.85**[a]	**2.52**[a]	**1.43a**	**2.79**[a]	**2.09**[a]	**1.53**[a]	**1.97**[a]	**1.70**[a]
Zn	**19.46**[a]	**37.99**[a]	**28.99**[a]	**15.75**[a]	**18.94**[a]	**17.36**[a]	**24.46**[a]	**36.68**[a]	**29.43**[a]
As	36.81	142.52	72.63	156.63	260.02	212.37	254.37	352.19	293.90
Se	15.17	35.58	30.65	27.32	79.01	48.25	11.72	25.23	19.54
Rb	**7.27**[a]	**21.73**[a]	**14.83**[a]	**3.40**[a]	**27.37**[a]	**13.11**[a]	**8.61**[a]	**10.54**[a]	**9.41**[a]
Sr	243.5	694.47	473.64	114.33	462.33	334.22	79.4	145.14	109.93
Cd	13.65	129.91	43.66	8.71	19.18	12.75	31.96	122.98	58.03
Ba	441.14	780.61	643.65	381.51	622.08	500.11	174.2	330.18	248.08
Pb	38.33	311.39	168.65	47.38	75.68	58.86	145.95	526.73	220.61

Note: a- measured in mg/kg

Figure 1. Comparison of element contents in Sengcu rice samples among three fields in Lao Cai province

3.2. Multivariate analysis

3.2.1. Principal component analysis

Because the content of different elements in rice will affect their weight during PCA analysis, the normalization (min-max) method was used to bring the variables to the same scale before using PCA analysis to provide a preliminary assessment of the ability to discriminate geographical origins of Sengcu rice in Lao Cai.

The results of PCA indicated that only the first four components of PCA with eigen-value more than 1, which can explain approximately 87.5% of the total variance of the database (PC_1:43.9%, PC_2:20.7%; PC_3: 12.5%, and PC_4: 5.8%). Therefore, it is evident that PC_1 was governed significantly by As, Ba, Sr, and Cr with absolute loading values stood at 0.44, 0.43, 0.40, and 0.38 respectively. Moreover, Zn (0.60); Cd (0.44); and Pb (0.38) were the main influence on PC_2. Figure 2 illustrates a score plot based on the first two PCs to classify the samples from three fields cultivating Sengcu rice in Lao Cai province. PC_1 divided the rice into two parts: BX and MV-BQ, while PC_2 play as a tool to separate rice origin between MV-BX and BQ.

The above results indicated that regression algorithms were necessary to construct discriminant models to identify the geographical origin of rice accurately. Hence,

LDA and PLS-DA were carried out in the following studies to obtain accurate classification rates of rice samples. Moreover, because the cumulative variance of the first five components stood at approximately>80%, therefore the results of PCA were used as input data for LDA and PLS-DA.
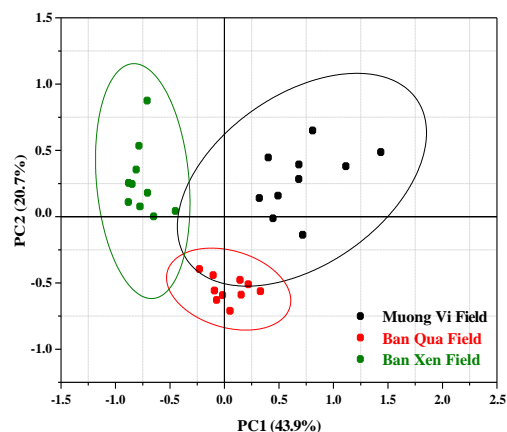

Figure 2. Loading plot of the PCA results obtained from data on 14 multi-element in rice samples from three regions culturing SengCu varieties in Lao Cai province
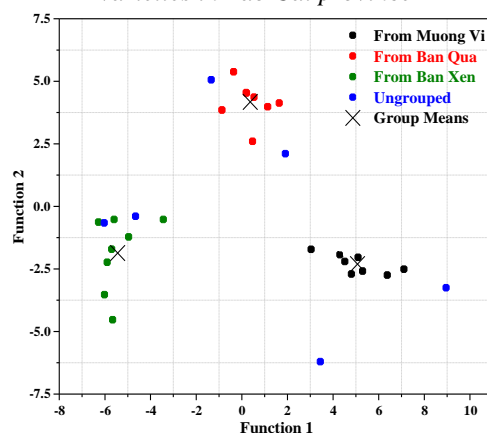

Figure 3. Scatter plot of Sengcu rice from different regions based on the two discriminant functions
Classification model

LDA is a classification technique that aims for maximizing the ratio of between-class variance to the within-class variance to achieve maximal separability. In this study, the predicted model was constructed based on 14 elements (ANOVA test in section 3.2) for separating three sites cultivating Sengcu rice. The results of LDA indicates that discriminant

function 1 (F1) accounted for67.91% of the total variance. And it was almost two times higher than the discriminant function 2 (F2) (32.09%). The eigen-value of F1 stood at 21.13 in comparison with 9.98 of F2. Wilks' Lambda factor was used as a criterion for the evaluated statistical significance of each discriminant function. The λ-value ranges from 1 (no discriminatory power) to 0 (perfect discriminatory power), this parameter of F1, F2 was very low (0.004 and 0.091)combined with a p-value of approximately 0.00 indicating good separation among the Sengcu rice cultivation regions in Lao Cai.

Figure 3 demonstrates the separation of Sengcu rice from MV, BQ, and BX fields based on the first two discriminant functions. Overall, it is clear evidence that function 1 provided the main key to separate between Sengcu rice from MV and that of BX. By contrast, function 2 provided the discrimination rice from between MV and BQ. Otherwise, the degree of separation between Sengcu-MV and Sengcu-BQ is clear, which was shown in Table 3. The original test shows that 24 out of24 rice samples SC-MV, SC-BQ, and SC-BX were discriminant accuracy. Remarkably, the correct rate reaches 100% for model validation samples.

In recent years, PLS-DA was known as an algorithm with high power to discriminate geographical origin. Instead of finding hyperplanes of maximum variance between the response and independent variables, this method finds a linear regression model by projecting the predicted variables and the observable variables to a new space. Hence, with the difference in nature, it is expected to have better classification capacity than the LDA method. In the same way as LDA, themodels were constructed based on 14 elements to geographical traceability of Sengcu rice.

The result of the first model was shown in Fig.4a,b. Seven elements (Al, Cr, Ni, Cu, As, Sr, Ba) which the variable importance in the projection (VIP) values stood at higher than 0.8, were found to be significant in creating a discrimination model for determining the geographical origin of Sengcu rice grains. Noticeably, standing at nearly 1.6 the VIP value of As, which was the highest in 14 elements, indicated that As was the most relevant to discriminate geographical origin Sengcu rice.

The cross-validation discriminant accuracy of the training set was up to 100% (8/8) for SC-MV, (8/8) for SC-BQ, and (8/8) SC-BX. For the validationset, the correct rate of SC-MV, SC-BQ, and SC-BX were also 100% (2/2). These results indicated that SC-MV could be accurately differentiated from SC-BQ and SC-BX rice using a PLS-DA model.

Table 3. Prediction result of PCA-LDA and PLS-DA model for traceability Sengcu rice

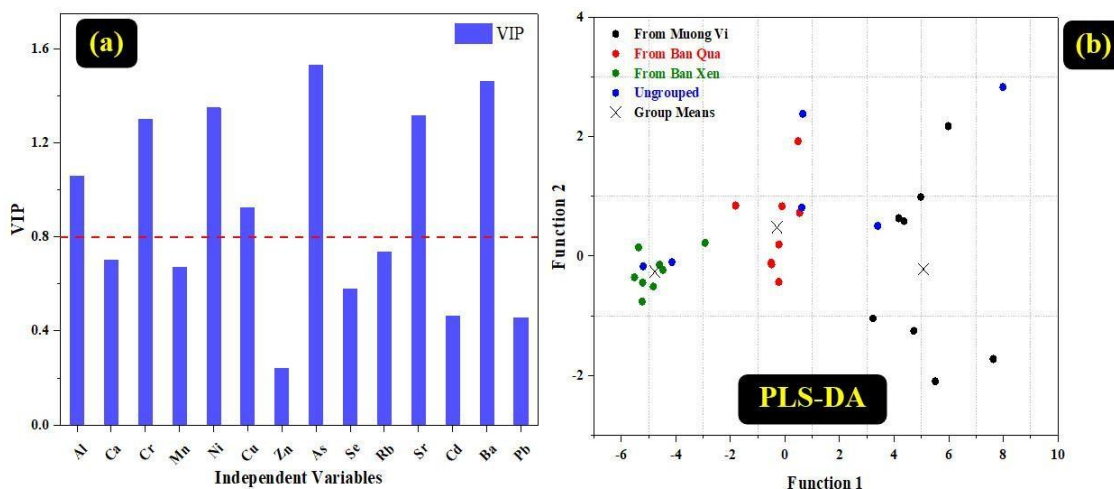| Model	Sample source	Training set			Validation set		
		Number of sample	Mistake	Correct rate	Number of sample	Mistake	Correct rate
PCA-LDA	MV	8	0	100	2	0	100
	BQ	8	0	100	2	0	100
	BX	8	0	100	2	0	100
PLS-DA	MV	8	0	100	2	0	100
	BQ	8	0	100	2	0	100
	BX	8	0	100	2	0	100

Figure 4. (a)-VIP value of each element and, (b)-PCA-PLS-DA score plot of Sengcu rice samples from Lao Cai.

4. CONCLUSION

In this work, the combination of ICP-MS with multivariate analysis has demonstrated to be useful for geographical traceability Sengcu rice. According to the suggested PCA method, As, Ba Sr, Cr, Zn, and Cd were good indicators for origin identification of Sengcu rice from Muong Vi, Ban Qua, and Ban Xen fields in Lao Cai. For the geographical traceability of Sengcu rice, LDA performed better than PLS-DA because of its advantage in maximizing the ratio of between-class and within-class distance. The above results confirmed that multi-element analysis is a promising method and suitable for the geographical traceability Sengcu rice. It helps to improve product value, protect brands, interests of consumers, and prevent commercial fraud. Our research also indicates the potential application for other agricultural products of high economic value such as premium rice brands, coffee, cashew, etc. in the future.

REFERENCE

1. Kusano, E., *Food Value Chain in ASEAN: Case Studies Focusing on Local Producers*, in *Economic Research Institute for ASEAN and East Asia*. 2019.

2. Maione, C. and R.M. Barbosa, *Recent applications of multivariate data analysis methods in the authentication of rice and the most analyzed parameters: A review*. Crit Rev Food Sci Nutr, 2019. **59**(12): p. 1868-1879.

3. Bui Thi Lam, P.L., Tran Huu Cuong, *Determintants of rice production in remote areas of Vietnam: a case study in Lao Cai province*, in *International conference on Agriculture development in the context of international integration: Opportunities and challenges*. 2016.

4. Srinuttrakul, W., V. Permnamtip, and A. Busamongkol, *Stable isotopic fingerprint of Sangyod rice*. Journal of Radioanalytical and Nuclear Chemistry, 2019. **322**(3): p. 1777-1782.

5. Qian, J., et al., *Traceability in food processing: problems, methods, and performance evaluations-a review*. Crit Rev Food Sci Nutr, 2020: p. 1-14.

6. Wadood, S.A., et al., *Recent development in the application of analytical techniques for the*

traceability and authenticity of food of plant origin. Microchemical Journal, 2020. **152**.

7. Bronzi, B., et al., *Geographical identification of Chianti red wine based on ICP-MS element composition.* Food Chem, 2020. **315**: p. 126248.

8. Qian, L., et al., *Determination of Geographical Origin of Wuchang Rice with the Geographical Indicator by Multielement Analysis.* Journal of Food Quality, 2019. **2019**: p. 1-7.

9. Chung, I.M., et al., *Geographic authentication of Asian rice (Oryza sativa L.) using multi-elemental and stable isotopic data combined with multivariate analysis.* Food Chem, 2018. **240**: p. 840-849.

10. Wu, H., et al., *Verification of imported red wine origin into China using multi isotope and elemental analyses.* Food Chem, 2019. **301**: p. 125137.

11. Choi, S.-H., et al., *Geographical origin identification of garlic cultivated in Korea using isotopic and multi-elemental analyses.* Food Control, 2020. **111**.

12. Gautam, R., et al., *Review of multidimensional data processing approaches for Raman and infrared spectroscopy.* EPJ Techniques and Instrumentation, 2015. **2**(1).

13. Park, J.H., S.-H. Choi, and Y.-S. Bong, *Geographical origin authentication of onions using stable isotope ratio and compositions of C, H, O,N, and S.* Food Control, 2019. **101**: p. 121-125.

14. Tian, F., F. Tan, and H. Li, *An rapid nondestructive testing method for distinguishing rice producing areas based on Raman spectroscopy and support vector machine.* Vibrational Spectroscopy, 2020. **107**.

15. Wang, F., et al., *Determination of the geographical origin of maize (Zea mays L.) using mineral element fingerprints.* J Sci Food Agric, 2020. **100**(3): p. 1294-1300.