Atomic Spectroscopy

Article

Discrimination of geographical origin of blueberry from three major producing areas of China using mineral element analyses

Lixue Kuang,[a] Jiyun Nie,[a,b,*] Jianyi Zhang,[a] Guofeng Xu,[a] Jing Li,[a] Yang Cheng,[a] Youming Shen,[a] and Farooq Saqi[b]

[a] Laboratory of Quality and Safety Risk Assessment for Fruit (Xingcheng), Ministry of Agriculture and Rural Affairs; Quality Inspection and Test Center for Fruit and Nursery Stocks (Xingcheng), Ministry of Agriculture and Rural Affairs; Research Institute of Pomology, Chinese Academy of Agricultural Sciences, Xingcheng, 125100, P.R. China

[b] College of Horticulture, Qingdao Agriculture University/Qingdao Key Lab of Modern Agriculture Quality and Safety Engineering, Qingdao, 266109, P.R. China

Received: *December 11, 2020;* ***Revised:*** *January 19, 2020;* ***Accepted:*** *January 19, 2020;* ***Available online:*** *January 19, 2021.*

DOI: 10.46770/AS.2021.004

ABSTRACT: Food quality and safety are closely related to the geographical origin of food. This study combined mineral element analysis and multivariate statistical analysis to discriminate the origins of 148 blueberry samples from three regions of China. The concentrations of K, Ca, Mg, Na, Fe, Cu, Mn, B, P, and Zn were determined inductively coupled plasma-atomic emission spectroscopy (ICP-AES). Variance analysis (ANOVA), Duncan's multiple-comparison test, and principal component analysis (PCA) were used to compare the element concentrations, and statistically significant differences were found among samples from different regions. Linear discriminant analysis (LDA), decision tree (DT), multilayer perceptron neural network (MLP-NN), and support vector machine (SVM) were utilized to build models for blueberry authentication. The results showed that the average concentrations of the minerals were in the order of K > P > Ca > Mg > Na > Mn > Fe > Zn > B > Cu, and the levels of K, Ca, Mg, Fe, Cu, Mn, B, P, and Zn were significantly different among regions by ANOVA and Duncan's multiple-comparison test. The study indicates that LDA, DT, MLP-NN, and SVM chemometric tools have the potential to discriminate the origin of blueberries. The results revealed that the MLP-NN and SVM models were more discriminative than the other two mathematical methods. The MLP-NN yielded an average discrimination rate of 92.7% for the training set and 94.7% for the test set, and the SVM with linear kernel function (SVM-lin) obtained an average identification rate of 91.8% for the training set and 94.7% for the test set. The order of successful identification rates was as follows: MLP-NN > SVM-lin > DT > LDA. This study can serve as a reference to identify the origin of blueberries and perform quality assurance for the fruit.

INTRODUCTION

Blueberry (*Vaccinium vitis-idaea L.*) is an important commercial fruit crop in China and is popular around the world for its appealing flavor and high nutraceutical value.[1-3] Its flesh is rich in not only vitamins, protein, minerals, and other nutrient elements but also unique precious anthocyanin, which constitutes a high proportion.[4-7] Moreover, blueberries also contain a large amount of calcium, iron, phosphorus, potassium, zinc, and other trace mineral elements.[8] Therefore, blueberry is a high-nutrition fruit recommended as one of the five healthy fruits by the Food and Agriculture Organization, and it has been named as the king of berries.[9] In addition, when fresh, blueberries can be processed into fruit juice, fruit wine, and other foods.[10,11] Currently, blueberries are one of the most commercialized fruits in the international market, and their consumption has increased substantially in recent years.[12]

In China, the total area under blueberry production increased from 681 ha in 2006 to 55,344 ha in 2018, and the corresponding total production increased from 342 t to 184,238 t.[13] Jilin Province (JL), Liaoning Province (LN), and Shandong Province (SD) are important blueberry cultivation regions in China.[14] Because of the differences in climate and soil conditions, blueberries from various regions have different qualities and characteristics.[15,16] Therefore, research on the technique of tracing the geographical origins of blueberries is beneficial for protecting the origin of product, regional brand, and regional featured products.[17-19]

In recent years, several studies have investigated the technique of tracing the geographical origin of agricultural products.[20-22] Mineral element analysis is considered an effective tool for geographical origin distribution.[19, 22-24] The sample pretreatment method of mineral element analysis is simple, fast, low-cost, and highly accurate.[25,26] Although element analysis technology has been widely used in studying the geographical origins of many fruits, including apple,[27] table grape,[28] cherry,[29] kiwifruit,[30] and lemon,[31] it has not been applied to blueberries.

Multivariate data analysis and machine learning techniques are powerful tools for food authentication.[19,32-34] Principal component analysis (PCA) and discriminant analysis are the preferred techniques for performing multivariate data analysis and exploratory and predictive analyses for origin identification.[35-40] Recently, the use of popular machine learning algorithms, such as decision tree (DT), neural network, and support vector machine (SVM), has become more common, because they have certain advantages over conventional methods.[34,41,42] Machine learning algorithms can mine information from the data themselves and better reflect the natural mechanism of data.

In this study, we analyzed and compared the contents of K, Ca, Mg, Na, Fe, Cu, Mn, B, P, and Zn in blueberries from different production regions. Models for blueberry authentication were built using the following machine learning techniques: linear discriminant analysis (LDA), DT, multilayer perceptron neural network (MLP-NN), and SVM.

EXPERIMENTAL

Apparatus and reagents. The apparatus used in sample preparation and determination comprised the following: a freeze grinder (SPEX 6875D, CEM, USA), microwave digestion system (Mars 6, CEM, USA), and inductively coupled plasma-atomic emission spectrometer (ICP-AES 9000, Shimadzu, Japan). The reagents HNO_3 (MOS grade, 70%) and H_2O_2 (GR grade, 30%) were purchased from Sinopharm Chemical Reagent Beijing Co., Ltd. (Beijing, China). Ultrapure deionized water (resistivity > 18 MΩ·cm) was obtained from a Milli-Q Plus water purification system (Millipore, Bedford, MA, USA). High-purity argon was used as an inert gas and was purchased from Huludao Weiye Gas Reagent Co., Ltd. (Liaoning, China). Standard solutions of mineral elements including K, Ca, Mg, Na, Fe, Cu, Mn, B, P, and Zn (100 mg/L), and the certified reference material (CRM) Apple (GBW10019) were obtained from the National Institute of Metrology (Beijing, China). Analytical reagent blanks and Apple (GBW10019) standards were prepared with each batch of digestion. They were then subjected to laboratory quality control analyses. The analysis results were only accepted when the measured concentrations in the reference material were within one standard deviation of the certified values.

Fig. 1 Picture showing the map of China with detailed origins from where blueberry samples were collected.

Sample collection and preparation. A total of 148 blueberry samples were collected from Shandong Province (SD) (n = 22), Liaoning Province (LN) (n = 46), and Jilin Province (JL) (n = 80) (Fig.1). For each sample, approximately 3 kg of blueberries was collected at the harvest-maturity stage from May to September 2018. The samples were transported to the laboratory within two days after collection. During the pretreatment, the samples were rinsed with water and then with ultrapure water. They were cut into pieces, frozen in liquid nitrogen, and ground using a SPEX Sample Prep system (CEM, USA). The powdered samples were aliquoted into individual tubes and stored at −20°C until further use.

Sample preparation and analysis. Each sample was weighed (5.0 g) and introduced into digestion tubes. Then 8.0 mL of 70% HNO_3 and 2.0 mL of 30% H_2O_2 were added to the samples. The samples were then vortexed and kept at room temperature overnight. The next day, the samples were digested using a CEM digestion system according to the digestion procedure in Table S1. Then, the digested samples were placed on an electric hot plate at 140°C to remove the nitric acid. The residual aqueous layer (approximately 1 mL) was cooled down to room temperature, and added to deionized water up to 50 mL. After filtration through a 0.22-μm nylon membrane, the supernatant was analyzed via inductively coupled plasma–atomic emission spectroscopy (ICP-AES). The K, Ca, Mg, Na, Fe, Cu, Mn, B, P, and Zn contents were detected via ICP-AES under the following operating conditions: forward power of 1.2 kW, plasma gas flow rate of 14 $L \cdot min^{-1}$, auxiliary gas flow rate of 1.2 $L \cdot min^{-1}$, and nebulizer gas flow rate of 0.7 $L \cdot min^{-1}$. The wavelength and calibration curves of each element are presented in Table S2. The linear correlation coefficients of the analyzed elements were all above 0.99. For accuracy evaluation, the obtained values of CRM were compared with the reference values. Analytical reagent blanks were run in parallel to check for the presence of analytes in the reagents. The detection limits of all of the elements evaluated from three times the standard deviation of 10 replicate blank measurements and the analytical results of elements in certification are shown in Table S2. Each sample was analyzed in triplicate. The element recoveries ranged from 93% to 106%, which agree with the certified concentrations in the reference materials.

Statistical analysis. The obtained data were analyzed using the SPSS 20.0 package and the SPSS Modeler 18.0 package. One-way analysis of variance (ANOVA) was first performed for each element of the blueberry samples (p-value < 0.05 was considered significant).[44,45] Duncan's multiple-comparison procedure was performed to determine the significant differences between individual regions.[34,] Principal component analysis was used for dimensionality reduction, and the most significant variables were selected through a forward stepwise analysis.[46,47] Moreover, LDA was performed to construct the classification model using the SPSS 20.0 package. Then MLP-NN, SVM, and DT models were implemented using the SPSS Modeler 18.0 package.

RESULTS AND DISCUSSION

Mineral element content in blueberries

The K, Ca, Mg, Na, Fe, Cu, Mn, B, P, and Zn concentrations were determined via ICP-AES. The ICP-AES technique has the advantages of relatively high sensitivity, simple operation, low cost, simultaneous analysis of multiple elements, and short analysis time.[48,49] It is also one of the most advanced and accurate analytical methods in the field of multi-element determination.[19,23] The average concentration mineral elements in the blueberries studied is of the following order: K > P > Ca > Mg > Na > Mn > Fe > Zn > B > Cu. The average values obtained for K, P, Ca, and Mg were 805.87, 104.41, 76.06, and 52.91 mg kg^{-1}, respectively; the concentrations of Na, Mn, Fe, Zn, B, and Cu were lower, with average values of 7.91, 7.86, 4.51, 1.23, 0.31, and 0.25 mg kg^{-1}, respectively. Table 1 presents the results of the mineral element analyses of blueberry samples from three different origins, in terms of minimum, maximum, mean, and median values. Duncan's multiple test was used to compare the element mean concentrations with statistically significant differences among samples from different regions.[34,41] The mean concentrations of P, Ca, Mg, Fe, and Cu in the SD samples were higher than those in samples from other regions, and the mean concentration of B in the JL samples was higher than those in other regional samples. The mean concentrations of P, Ca, Mg, and B in the LN samples were lower than those in other regional samples. However, there was a high level of overlap between the ranges of these elements despite their significant differences. From Table 1, we also noticed that some elements of the standard deviation are large, which may indicate that the content of some elements within different regions in the same province is different. These results suggest that each region had a typical profile which was influenced by the environment, and the traceable element choice depends on the source and extent of the sample area.[34]

Furthermore, PCA was performed for exploratory data analysis. It was applied to the auto-scaled data matrix to provide a data structure study in a reduced dimensionality, keeping the maximum amount of variability present in data.[31,46,47] Through the dimensionality reduction function of PCA, all the information can be described by a few comprehensive variables, and the differences in feature elements among regions can be more intuitive and simpler to understand.[50] Principal component analysis was performed using the nine elements with significant regional differences in blueberries. The PCA analysis showed that KMO value was 0.788, indicating that there was a high correlation between variables, which was suitable for factor analysis. The characteristic roots of the first four principal factor s were 3.816, 1.243, 1.049 and 0.793, respectively. Because the load value of Mn on factor 4 was 0.711, so the first four principal components were selected and together explained 76.7% of the total variance. The factor extraction values of each principal factor were all greater than 0.55. PC1 contained the K, P, Ca, Mg, and B contents (42.4%), while PC2 contained the Fe and Cu contents (13.8%). PC3 explained 11.7% variance and contained the Zn content. PC4 explained 8.8% variance and contained the Mn content.

Table 1. Concentrations of Mineral Elements in Blueberry Samples, mg kg^{-1}

Element	SD (n=22)			LN (n=46)			JL (n=80)		
	Range	Mean (±sd)	Median	Range	Mean(±sd)	Median	Range	Mean(±sd)	Median
K	609.5-1120	876.07 ± 150.30^{a}	827	509.5-878.5	674.83 ± 87.33^{b}	669	580.5-1120	861.91 ± 119.79^{a}	868
P	79.8-179	122.60 ± 32.20^{a}	120	56.5-128	85.89 ± 18.69^{c}	83.5	66.85-153.5	110.06 ± 18.94^{b}	108.5
Ca	45.2-168.7	88.85 ± 27.81^{a}	94.35	35.85-102.7	65.48 ± 14.99^{c}	67.03	50.05-205.85	78.64 ± 22.36^{b}	74.93
Mg	40.9-88.9	61.08 ± 10.72^{a}	60.23	30.05-56.3	44.11 ± 6.70^{c}	44.65	36.95-79.75	55.72 ± 7.75^{b}	54.93
Na	1.39-16.8	8.01 ± 4.26^{a}	7.55	1.11-39.65	7.69 ± 7.35^{a}	5.37	0.03-35.42	8.01 ± 6.35^{a}	5.70
Mn	3.29-14.3	7.69 ± 2.70^{ab}	7.14	2.95-13.95	6.19 ± 2.43^{b}	6.07	2.85-22.95	8.86 ± 3.86^{a}	8.63
Fe	2.76-12.9	5.70 ± 2.87^{a}	4.43	1.54-17.45	4.28 ± 2.97^{b}	3.77	1.84-14.85	4.31 ± 2.25^{b}	3.76
Zn	0.74-5	1.35 ± 0.85^{a}	1.17	0.55-2.14	1.06 ± 0.33^{b}	1.04	0.48-3.31	1.29 ± 0.57^{ab}	1.13
B	0-0.51	0.25 ± 0.16^{b}	0.23	0.01-0.4	0.13 ± 0.10^{c}	0.10	0.12-0.91	0.43 ± 0.15^{a}	0.42
Cu	0.22-0.67	0.42 ± 0.13^{a}	0.41	0.06-0.36	0.20 ± 0.08^{b}	0.19	0.11-0.41	0.23 ± 0.07^{b}	0.23

Note: SD, Shandong Province; LN, Liaoning Province; JL, Jilin Province. sd, standard deviations. [a, b] Numbers with different superscripts are significantly different among blueberry samples from different regions by the Duncan test. ($P < 0.05$).

Fig. 2 The PCA scatter plot of the scores of blueberries from different geographical origin on PC 1 and PC 2.

Fig. 3 The PCA 3D scatter plot of the scores of blueberries from different geographical origin on the first three principal components.

Fig. 4 Scatter plot of discriminant scores of blueberry samples by LDA.

To visualize the data trends, a score plot was obtained using the top three principal components. As shown in Fig. 2 and Fig. 3, although some of the samples of the three provinces overlapped, the trend of the discrimination of samples based on their origin is evident. The LN samples were concentrated on the negative part of the PC1 coordinate, and the JL samples were concentrated on the positive part, which shows that the content of representative elements (K, P, Ca, Mg, and B) for PC1 in JL samples was higher than those in LN samples. Furthermore, SD samples were mainly distributed in the positive part of PC2 and the skewed positive part of PC1; this shows that the content of representative elements (Fe and Cu) for PC2 in SD samples was higher than those in JL and LN samples; it also shows that the K, P, Ca, Mg, and B content for PC1 in SD samples was slightly higher than that in JL samples and significantly higher than that in LN samples. The distribution of SD samples was relatively dispersed. The is mainly because the samples were from three cities, including Qingdao City, Yantai City, and Weihai City of Shandong Province. The samples were widespread, and the number of samples was relatively small; thus, the sample size should be further increased in future research. Based on this statistical analysis, it is clear that blueberry samples from different regions have differences in mineral element content; this property can be leveraged to trace blueberry regional sources.

Machine learning for geographical origin evaluation

Machine learning is a technique of using algorithms to extract hidden patterns from large amounts of data and then using the patterns for prediction or classification.[51] Specifically, machine learning can be regarded as searching for a function whose input is the sample data and whose output is the expected result. Machine learning is widely used in several emerging fields, including food authenticity identification.[29] It has potential application in the field of food origin tracing.[52,53] In this work, four classification algorithms (LDA, MLP-NN, SVM, and DT) were evaluated to classify blueberries from different geographical origins based on the content of mineral elements.

Linear discriminant analysis is of great use in the authentication of the geographical origins of food/food products.[54-58] In our work, LDA was performed on the basis of the contents of 10 elements. A stepwise discriminant procedure was performed to extract the best discriminant variable separating blueberry samples from different origins. As a result, four elements (B, Cu, Mg, and K) were selected, and two discriminant functions were constructed on the basis of Wilks' lambda values. The two functions explained 100% of the variance (function 1 explained 52.5% of the total variance, and function 2 explained 47.5%). The discrimination functions for each province are shown as follows:

$$\text{Group 1 (SD)} = -1.896B + 2.988Cu - 0.21K + 1.475Mg - 4.338$$
$$\text{Group 2 (LN)} = -1.208B - 0.387Cu - 0.757K - 0.428Mg - 2.276$$
$$\text{Group 3 (JL)} = 1.216B - 0.599Cu + 0.441K - 0.159Mg - 1.580$$

The separation of blueberry samples from SD, LN, and JL was checked by plotting the two function scores (Fig. 4). To check the reliability of the developed classification model, a cross-validation method was implemented for the classification and probability estimation for the blueberry samples. The LDA classification and cross-validation results based on mineral elements were summarized in Table 2. From the results in Fig. 4 and Table 3, the

Table 2. Classification of Blueberry Samples from Different Regions by LDA.

		Predicted Group Membership				
		SD	LN	JL	Total	Correct, %
Original	SD	17	4	1	22	77.3
	LN	0	42	4	46	91.3
	JL	1	11	68	80	85.0
						85.8
Cross-validated	SD	17	4	1	22	77.3
	LN	1	40	5	46	87.0
	JL	1	11	68	80	85.0
						84.5

Table 3. Classification of Blueberry Samples from Different Regions by DT

	SD	LN	JL	Rate of classification (%)
Training (110)				
SD	17	1	1	89.5
LN	0	32	0	100.0
JL	0	7	52	88.1
Total				91.8
Test (38)				
SD	2	0	1	66.7
LN	0	14	0	100.0
JL	4	0	17	81.0
Total				86.8

Table 4. Classification of Blueberry Samples from Different Regions by MLP-NN

	SD	LN	JL	Rate of classification (%)
Training (110)				
SD	18	0	1	94.7
LN	0	26	6	81.3
JL	0	1	58	98.3
Total				92.7
Test (38)				
SD	2	1	0	66.7
LN	0	13	1	92.9
JL	0	0	21	100.0
Total				94.7

total empirical grouped observations classification rate was 85.8%, and the cross-validated grouped observations classification rate was 84.5%. The classification rates of SD, LN, and JL were 77.3%, 91.3%, and 85.0%, respectively. Four SD samples were incorrectly identified as LN samples, and only one sample was incorrectly identified as JL samples. This may be because the two production areas had temperate monsoon climates and similar soil types. Eleven LN samples were incorrectly identified as JL samples, and only one sample was incorrectly identified as an SD sample. This may be due to the geographical proximity of the two provinces in which the two appellations are located.

The DT uses a tree structure to build the classification models.[59] A dataset is divided into smaller subsets. A leaf node represents a decision. Each node represents a feature in an instance in a DT that is to be classified, and each branch represents a value.[60] The classification of instances begins from the root node, and the instances are sorted based on their feature values. Categorical and numerical data can be handled by decision trees.[61] The C5.0 algorithm of DT was used in this study. The C5.0 algorithm was developed on the basis of C4.5 and can provide more accurate and efficient results than C4.5.[61] The blueberry origins were defined as the target variable, and the 10 mineral element indicators were defined as input variables. The data were divided into a training set (110 samples) to build the models and a test set (38 samples) to calculate their classification performance. Table 3 presents the discrimination results of the DT model. The average identification rate was 91.8% for the training set and 86.8% for the test set. From the table, the LN samples were correctly classified in the training set. The identification rates of SD and JL were 89.5% and 88.1%, respectively. In the test set, all the samples from LN were correctly identified. Among the 21 JL samples, four samples were misidentified as originating from SD, and the identification rate was 81.0%. Three samples were from SD, of which one sample was misidentified to be from JL. Although only one sample was wrongly identified, the accuracy rate was low, only 66.7%. This is because the sample size was small. This situation can be found in MLP-NN and SVM models. Therefore, the sample size of SD should be further increased in future research.

A neural network is a mathematical model for information processing using a structure similar to the synaptic connections of the brain, and it has nonlinear dynamic properties.[62] The good performance of a neural network in the field of food origin traceability has been verified.[63] In this work, an MLP-NN with a sigmoid activation function was utilized. Artificial neural networks such as multilayer perceptrons (MLPs) trained by back-propagation neural network (BPNN) are efficient tools for classifying and discriminating food products.[64,65] The MLP consists of formal neurons and connections between them. The neurons are arranged in layers (an input layer, one or more hidden layers, and an output layer), and the connections are unidirectional from the input to the output. Adjacent layers are fully connected, and no connections exist between neurons within the same layer. In our study, the samples were divided into a training data set (110 samples) and a test data set (38 samples). We used an MLP with three layers according to the architecture $10 \times 4 \times 3$: an input layer with 10 neurons, a hidden layer with 4 neurons, and an output layer with 3 neurons, corresponding to the classes to be assigned (SD, LN, and JL).

Table 4 presents the discrimination results of the MLP-NN model. As shown in the table, the MLP-NN model performed well

Table 5. Classification of Blueberry Samples from Different Regions by SVM

	SD	LN	JL	Rate of classification (%)
Training (110)				
SD	17	1	1	89.5
LN	0	27	5	84.4
JL	0	2	57	96.6
Total				91.8
Test (37)				
SD	2	1	0	66.7
LN	0	14	0	100.0
JL	1	0	20	95.2
Total				94.7

in discriminating different source regions, with an average identification rate of 92.7% for the training set and 94.7% for the test set. For the test set, the classification rate of LN was 92.9%. All the samples from JL were correctly identified. There were only three samples from SD, and one sample was misidentified to be from LN.

In recent years, SVM has shown some advantages in the development of machine learning.[53,66] Support vector machine is a nonlinear supervised learning method, and it maps the training data into a high-dimensional space and constructs a linear classifier in that space.[67] If the categories are linearly separated, the SVM algorithm finds optimal hyper-plane boundaries, which separate the classes of the training set and the unknown sample. However, if the classes are separated by a nonlinear boundary, then the kernel function is used to find the boundary by mapping the non-separable data into a higher-dimensional space.[50]

In this study, SVM algorithms with linear kernel function (lin), polynomial kernel function (pol), and radial basis kernel function (rbf) were compared. Of the three algorithms, the SVM-lin was the best, with an average successful identification rate of 91.8% for the training set and 94.7% for the test set. The second-best was the SVM-pol, which had an average accuracy of 90.0% for the training set and 89.5% for the test set. The SVM-rbf exhibited the worst performance, with a success rate of only 83.6%. Table 5 presents the discrimination results of the SVM-lin model. The classification rates of LN and JL were 92.9% and 100% for the test set, respectively. Although most of the blueberry samples were accurately discriminated, the correct prediction of SD samples was relatively low. This may be the result of the fewer samples. Therefore, it is necessary to incorporate more samples to develop a more accurate model.

Through the above analysis, we can see that the four algorithms displayed different degrees of success. Of the four algorithms, MLP-NN had the best performance, with a success rate of 92.7% in the training set and 94.7% in the test set. The order of successful identification rates among the algorithms is as follows: MLP-NN > SVM-lin > DT > LDA. The MLP-NN and SVM-lin are ideal models for discriminating different origins of blueberry samples. With the continuous improvement of people's living standards, consumers pay more attention to the origin of fruits. However, there has hardly been any research on data mining techniques applied to the geographical classification of the blueberries. Until now, no relevant study has been reported on the geographical classification of blueberry samples. During the last few decades, the applications of some pattern recognition techniques combined with multi-element analysis for tracing the geographical source have achieved good results in some areas. The main adopted techniques include PCA, cluster analysis, and LDA. However, other promising methods such as DT, ANN, and SVM have been scarcely used for this purpose. In this study, we combined multi-element analysis and these promising methods to identify the geographical origin of blueberries and achieved satisfactory results. Therefore, our study presents a novel investigation in the field of origin identification.

CONCLUSIONS

A multi-element–based geographical verification of blueberries from three regions of China was conducted in this study. The results of MLP-NN, SVM, DT, and LDA indicated the possibility of the geographical authentication of the Chinese blueberry based on multiple elements. Among the statistical tests performed, MLP-ANN and SVM analyses proved to be the more successful for differentiating the origins of the analyzed samples than DT and LDA models. Future studies may incorporate more samples and identify more region-specific elements including rare earth elements, precious metals and ultra-trace elements to improve the identification accuracy. The multi-element–based geographical authentication of blueberries in this study can provide a basis for further geographical origin studies.

ASSOCIATED CONTENT

Please contact the corresponding author for the Supporting Information (Table S1 and S2).

AUTHOR INFORMATION

Corresponding Author

* J. Y. Nie

Email address: jiyunnie186@sina.com

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was financially supported by the National Program for Quality and Safety Risk Assessment of Agricultural Products of China (GJFP2017003, GJFP2018003), the Agricultural Science and Technology Innovation Program of Chinese Academy of Agricultural Sciences (CAAS-ASTIP).

REFERENCES

1. Gao, S. Cao, X. Fang, H. Chen and S. Xiao, *Food Chem.*, 2017, **219**, 436-442. https://doi.org/10.1016/j.foodchem.2016.09.186
2. G. Mazza, C. D. Kay, T. Cottrell and B. J. Holub, *J. Agr. Food Chem.*, 2002, **50**, 7731-7737. https://doi.org/10.1021/jf020690l
3. J. Zhang, J. Y. Nie, J. Li, H. Zhang, Y. Li, S. Farooq, S. Asim and J. Wang, *J. Integr. Agr.*, 2020, **19**, 2352-2361. https://doi.org/10.1016/S2095-3119(20)63236-1
4. Y. Wu, Q. Zhou, X. Chen, X. Li, Y. Wang and J. Zhang, *Food Res. Int.*, 2017, **100,** 312-324. https://doi.org/10.1016/j.foodres.2017.07.004
5. N. Istek and O. Gurbuz, *J. Funct. Foods*, 2017, **38,** 298-307. https://doi.org/10.1016/j.jff.2017.09.039
6. S. Skrovankova, D. Sumczynski, J. Mlcek, T. Jurikova and J. Sochor, *Int. J. Mol. Sci.*, 2015, **16**, 24673-24706. https://doi.org/10.3390/ijms161024673
7. S. Sellappan, C. C. Akoh and G. Krewer, *J. Agric. Food Chem.*, 2002, **50,** 2432-2438. https://doi.org/10.1021/jf011097r
8. S. H. Nile and S. W. Park, *Nutrition*, 2014, **30**, 134-144. https://doi.org/10.1016/j.nut.2013.04.007
9. Y. R. Ji, W. Z. Hu, A. L. Jiang, Z. L. Xiu, J. Liao, X. Z. Yang, Y. G. Guan, G. W. Saren and K. Feng, *J. Sci. Food Agr.*, 2019, **99**, 6296-6306. https://doi.org/10.1002/jsfa.9904
10. P. S. Melo, A. P. Massarioli, C. Denny, L. F. dos Santos, M. Franchin, G. E. Pereira, T. M. F. D. S. Vieira, P. L. Rosalen and S. M. D. Alencar, *Food Chem.*, 2015, **181,** 160-169. https://doi.org/10.1016/j.foodchem.2015.02.087
11. B. S. B. Bamba, J. Shi, C. C. Tranchant, S. J. Xue, C. F. Forney and L. T. Lim, *Molecules*, 2018, **23,** 1685. https://doi.org/10.3390/molecules23071685
12. T. Tefera, K. R. Tysnes, K. S. Utaaker and L. J. Robertson, *Food Waterborne Parasitol.*, 2018, **10,** 23-38. https://doi.org/10.1016/j.fawpar.2018.04.002
13. Y. D. Li, H. Y. Sun and L. Chen, *China Fruits*, 2016, **5**,1-10. https://doi.org/10.16626/j.cnki.issn1000-8047.2016.05.001
14. Y. D. Li, J. P. Pei and H. Y. Sun, *J. Jilin Agri. Univ.*, 2018, **40**, 421-432. https://doi.org/10.13327/j.jjlau.2018.4606
15. Q. Z. Liu, J. W. Wang, Q. D. Gong and L. Xin, *Deciduous Fruits*, 2018, **50**, 1-4. https://doi.org/10.13855/j.cnki.lygs.2018.06.001
16. J. Rohloff, E. Uleberg, A. Nes, T. Krogstad, R. Nestby and I. Martinussen, *J. Appl. Bot. Food Qual.*, 2015, **88**, 274-287. https://doi.org/10.5073/JABFQ.2015.088.040
17. M. M. Aung and Y. S. Chang. *Food Control*, 2014, **39**, 172-184. https://doi.org/10.1016/j.foodcont.2013.11.007
18. S. A. Drivelos and C. A. Georgiou, *TRAC-Trend. Anal. Chem.,* 2012, **40**, 38-51. https://doi.org/10.1016/j.trac.2012.08.003
19. M. de la Guardia and A. Gonzálvez, Food protected designation of origin: methodologies and applications. *Amsterdam Elsevier*, 2013.
20. Gonzalvez, S. Armenta and M. de la Guardia, *TRAC-Trend. Anal. Chem.,* 2009, **28**, 1295-1311. https://doi.org/10.1016/j.trac.2009.08.001
21. R. J. Luo, T. Jiang, X. B. Chen, C. C. Zheng, H. B. Liu and J. Yang, *Food Chem.*, 2019, **274**, 1-7. https://doi.org/10.1016/j.foodchem.2018.08.104
22. S. Kelly, K. Heaton and J. Hoogewerff, *Trends Food Sci. Technol.*, 2005, **16**,555-567. https://doi.org/10.1016/j.tifs.2005.08.008
23. N. Sharma, V. K. Singh, Y. Lee, S. Kumar, P. K. Rai, A. K. Pathak and V. K. Singh, *At. Spectrosc.*, 2020, **41,** 234-241. https://doi.org/10.46770/AS.2020.06.003
24. S.A. Drivelos and C.A. Georgiou, *TRAC-Trend. Anal. Chem.*, 2012, **40**, 38-51. https://doi.org/10.1016/j.trac.2012.08.003
25. F. Di Giacomo, A. Del Signore and M. Giaccio, *J. Agric. Food Chem.,* 2007, **55**, 860-866. https://doi.org/10.1021/jf062690h
26. S. M. Sun, B. L. Guo, Y. M. Wei and M. T. Fan, *Food Chem.*, 2010, **124**, 1151-1156. https://doi.org/10.1016/j.foodchem.2010.07.027
27. J. Y. Zhang, J. Y. Nie, L. X. Kuang, Y. M. Shen, H. D. Zheng, H. Zhang, S. Farooq and S. Asim, *J. Sci. Food Agric.*, 2019, **99**, 6182-6190. https://doi.org/10.1002/jsfa.9890
28. F. Longobardi, G. Casiello, V. Centonze, L. Catucci and A. Agostiano, *J. Sci. Food Agric.,* 2017, **97**, 3173-3180. https://doi.org/10.1002/jsfa.8161
29. F. Longobardi, G. Casiello, A. Ventrella, V. Mazzilli, A. Nardelli, D. Sacco, L. Catucci, and A. Agostiano, *Food Chem.*, 2015, **170**, 90-96. https://doi.org/10.1016/j.foodchem.2014.08.057
30. J. Guo, Y. Yuan, P. Dou and T. Yue, *Food Chem.*, 2017, **232**, 552-559. https://doi.org/10.1016/j.foodchem.2017.04.037
31. J. E. Gaiad, M. J. Hidalgo, R. N. Villafañe, E. J. Marchevsky and R. G. Pellerano, *Microchem J.,* 2016, **129**, 243-248. https://doi.org/10.1016/j.microc.2016.07.002
32. M. Jiménez-Carvelo, A. González-Casado, M. G. Bagur-González and L. Cuadros-Rodríguez, *Food Res. Int.*, 2019, **112**, 25-39. https://doi.org/10.1016/j.foodres.2019.03.063
33. H. Y. Zhao, B. L. Guo, Y. M. Wei, B. Zhang, S.M. Sun, L. Zhang and J. H. Yan, *J. Agric. Food Chem*., 2011, **59,** 4397-4402. https://doi.org/10.1021/jf200108d
34. C. Li, S. C. Yan, Q. S. Guo, K. Y. Zheng, Y. F. Shi, X. F. Xiao and G. Q. Long, *Chemometrics Intell. Lab. Syst.,* 2014, **136**, 115-120.

http://dx.doi.org/10.1016/j.chemolab.2014.05.008

35. Gonzalvez, S. Armenta and M. de la Guardia, *Food Chem.,* 2010, **121**, 878-886. https://doi.org/10.1016/j.foodchem.2009.11.091
36. Y. Yücel and P. Sultanoğlu, *Food Chem.,*2013, **140,** 231–237. https://doi.org/10.1016/j.foodchem.2013.02.046
37. C. Benincasa, J. Lewis, E. Perri, G. Sindona and A. Tagarelli, *Anal. Chim. Acta*. 2007, **585,** 366-370. https://doi.org/10.1016/j.aca.2006.12.040
38. Y. Suzuki, Y. Chikaraishi, N.O. Ogawa, N. Ohkouchi and T. Korenaga, *Food Chem.*, 2008, **109,** 470-475. https://doi.org/10.1016/j.foodchem.2007.12.063
39. G. Li, L. Nunes, Y. Wang, P.N. Williams, M. Zheng, Q.F. Zhang and Y. Zhu, *Chin. J. Environ. Sci.*, 2013, **25,** 144-154. https://doi.org/10.1016/S1001-0742(12)60007-2
40. S. M. Sun, B. L. Guo, Y. M. Wei and M. T. Fan, *Food Chem.*, 2011, **124**, 1151-1156. https://doi.org/10.1016/j.foodchem.2010.07.027
41. S. Pérez-Magariño, M. Ortega-Heras, M.L. González-San José and Z. Boger, *Talanta*, 2004, **62,** 983-990. https://doi.org/10.1016/j.talanta.2003.10.019
42. P.Y. Cheng, W.L. Fan and Y. Xu, *Food Control*, 2014, **35,** 153-158. https://doi.org/10.1016/j.foodcont.2013.07.003
43. L. X. Kuang, J. Y. Nie, Z. X. Li, Y. L. Wu, Z. Yan and Y. Cheng, *Scientia Agricultura Sinica*, 2016, **49**, 3993-4003. https://doi.org/10.3864/j.issn.0578-1752.2016.20.013
44. H. Zhao, B. Guo and Y. Wei, *J. Agr. Food Chem.,*2011, **59,** 4397-4402. https://doi.org/10.1021/jf200108d
45. Y. Zhao, B. Zhang, G. Chen, A. Chen, S. Yang and Z. Ye, J. Agr. *Food Chem.*, 2013, **61,** 7055-7060. https://doi.org/10.1021/jf400947y
46. U. Kropf, M. Korosec, J. Bertoncelj, N. Ogrinc, M. Necemer and P. Kump, *Food Chem.,* 2010, **121,** 839-846. https://doi.org/10.1016/j.foodchem.2009.12.094
47. M. Savio, M. S. Ortiz, C. A. Almeida, R. A. Olsina, L. D. Martinez and R. A. Gil, *Food Chem.,* 2014, **159,** 433-438. https://doi.org/10.1016/j.foodchem.2014.03.041
48. S. M. Cruz, L. Schmidt, F. M. D. Nora, M. F. Pedrotti, C. A. Bizzi, and J. S. Barin, *Microchem. J.*, 2015, **123**, 28-32. https://doi.org/10.1016/j.microc.2015.05.008
49. T. H. Hansen, T. C. de Bang, P. Pedas, S. Husted and J. K. Schjoerring, *Plant Mineral Nutr.,* 2013, **953,** 121-141. https://doi.org/10.1007/978-1-62703-152-3_8
50. J. Qi, Y. Y. Li, C. Zhang, C. Wang, J. Q. Wang, W. P. Guo and S. W. Wang, *Food Chem.*, 2021, **337**, 127779. https://doi.org/10.1016/j.foodchem.2020.127779
51. Y. S. Bong, W. J. Shin, M. K. Gautam, Y. J. Jeong, A. R. Lee and C. S. Jang, *Food Chem.*, 2012, **135**, 2666-2674. https://doi.org/10.1016/j.foodchem.2012.07.045
52. H. Du, B. Tang, S. Cao, C. Xi, X. Li, and L. Zhang, *Microchem. J.*, 2020, **152**, 104427. https://doi.org/10.1016/j.microc.2019.104427
53. M. J. Hidalgo, D. C. Fechner, E. J. Marchevsky and R. G. Pellerano. *Food Chem.*, 2016, **210**, 228-234. https://doi.org/10.1016/j.foodchem.2016.04.120
54. G. Ma, Y. B. Zhang, J. Y. Zhang, G. Q. Wang, L. Y. Chen, M. L. Zhang, T. Liu, X. Liu and C. Y. Lu, *Food Control*, 2016, **59**, 714-720. https://doi.org/10.1016/j.foodcont.2015.06.037
55. S. M. Sun, B. L. Guo, Y.M. Wei and M. T. Fan, *Food Chem.*, 2011, **124,** 1151-1156. https://doi.org/10.1016/j.foodchem.2010.07.027
56. H. Zhao, B. Guo, Y. Wei and B. Zhang, *J. Cereal Sci.,* 2013, **57,** 391-397. https://doi.org/10.1016/j.jcs.2013.01.008
57. X. F. Liu, C. H. Xue, Y. M. Wang, Z. Li, Y. Xue and J. Xu, *Food Control*, 2012, **23,** 522-527. https://doi.org/10.1016/j.foodcont.2011.08.025
58. A.E. Martin, R.J. Watling and G.S. Lee, *Food Chem.*, 2012, **133**, 1081-1089. https://doi.org/10.1016/j.foodchem.2012.02.013
59. K. Ni, J. Wang, Q. F. Zhang, X. Y. Yi, L. F. Ma, Y. Z. Shi and J. Y. Ruan, *J. Food Compos. Anal.*, 2018, **67**, 104-109. https://doi.org/10.1016/j.jfca.2018.01.005
60. R. Pandya, and J. Pandya, *J. Food Compos. Anal.*, 2015, **117**, 18-21. https://doi.org/10.5120/20639-3318
61. D. Jankowski and K. Jackowski, 13th IFIP International Conference on Computer Information Systems and Industrial Management (CISIM), 2014, 23-32. https://doi.org/10.1007/978-3-662-45237-0_4
62. O. Anjos, C. Iglesias, F. Peres, J. Martínez, Á. García and J. Taboada, *Food Chem.*, 2015, **175**, 128-136. https://doi.org/10.1016/j.foodchem.2014.11.121
63. R. Zbikowski, *Estuar. Coast. Shelf Sci.*, 2011, **94,** 138-143. https://doi.org/10.1016/j.ecss.2011.05.027.
64. P. Evans, K.C. Persaud, A.S. McNeish, R.W. Sneath, N. Hobson and N. Magan, *Sensors Actuators B: Chem.,* 2000, **69,** 348-358. https://doi.org/10.1016/S0925-4005(00)00485-8
65. B. Dębska and B.Guzowska-Świder, *Anal. Chim. Acta,* 2011, **705**, 283-291. https://doi.org/10.1016/j.aca.2011.06.033
66. X. Capron, D. L. Massart, and J. Smeyersverbeke, *Eur. Food Res. Technol.*, 2007, **225,** 559-568. https://doi.org/10.1007/s00217-006-0454-2
67. L. Xu and X.G. Shao, Chemometric Methods, *Science Press, Beijing*, 2006.