Quality Assurance and Safety of Crops & Foods, 2020; 12 (3): 126–139

Geographical specificity of fatty acid and multi-element fingerprints of soybean in northern China

Dong Sheng Cui[1], Hong Yu Jia[1], Han Qing Lai[1], Zhao Hui Wang[1,*], Xue Fei Mao[2]

[1]School of Food Science and Engineering, Jilin Agricultural University, Changchun 130118, China P.R.; [2]Institute of Quality Standard and Testing Technology for Agro-products, Chinese Academy of Agricultural Sciences, and Key Laboratory of Agro-food Safety and Quality, Ministry of Agriculture and Rural Affairs, Beijing 100081, China P.R.

***Corresponding Author**: Zhao Hui Wang, School of Food Science and Engineering, Jilin Agricultural University, Changchun 130118, China P.R. Email: wzhjlndsp@aliyun.com

Received: 12 June 2020 / Accepted: 26 August 2020 / Published: 21 September 2020
© 2020 Codon Publications

OPEN ACCESS

ORIGINAL ARTICLE

Abstract

Soybean is an important food crop in China. Recently, crops cultivated in specific geographical locations have started attracting high prices. Therefore, developing a technique to identify the geographical origin of a crop is crucial to prevent fraud. In this work, we measured the contents of five fatty acids and 17 elements in soybean samples produced in Heilongjiang, the Inner Mongolia Autonomous Region, Jilin and Liaoning using gas chromatography and inductively coupled plasma mass spectrometry. Correlation analysis, principal component analysis and cluster analysis were used to identify the relationship between the metabolic fingerprint and the geographical location. Our results showed a significant correlation between the contents of fatty acids and geographical origin. Principal component analysis provided a preliminary classification of all variables. Hierarchical clustering, based on heat maps, showed that all samples could be classified based on their geographical origins. The model established by partial least squares discriminant analysis showed 89.9% predictive ability, further proving that the 14 classification indexes, comprising fatty acids and elements, could be used as molecular fingerprints to identify and distinguish soybean samples from four different production areas. Besides, pairs of soybean sample fingerprints from the four provinces were compared, and the differences in fatty acid and element contents between the provinces were explained based on the climatic environment and soil distribution. In conclusion, our method of classifying and confirming soybean production areas through fatty acid and multi-element fingerprints can potentially be used for identifying soybean of similar origins.

Keywords: soya bean; fingerprint characteristics; cluster heat map; geographical origin

Introduction

Soybean, which is rich in various nutrients, such as minerals, oils and proteins, is one of the most important food crops in China (Liu, 2014), and it is widely cultivated all over China. The northern production areas, mainly the Inner Mongolia Autonomous Region and northeast provinces, provide the highest soybean yield (Zhao *et al.*, 2018). In recent years, because of the complex production environments of agricultural products, products with specific geographical indications have started attracting higher prices. Driven by economic benefits, some illegal vendors fraudulently sell low-quality products for economic advantage, which directly damages the legitimate rights and interests of consumers (Aung and Chang, 2014). Therefore, confirming the origin of soybean and other agricultural products not only protects the geographical indications of the products but also ensures safety monitoring of soybean food "from the field to the dining table" to maintain a fair market order.

Food fingerprinting is an analytical method that provides information regarding food ingredients through non-selective methods, such as instrument fingerprinting. The fingerprint characteristics are then analysed according to the chemical composition of the food, which helps

 ISSN 1757-837X online, DOI 10.15586/qas.v12i3.767

in identifying its origin (Pérez-Castaño *et al.*, 2019). Currently, the technique underlying the confirmation of the geographical landmark of products is to analyse the differences between the variable components of products from different geographical sources, identify effective indicators for origin confirmation, establish a discriminant model, and predict the sample classification. Generally, the methods used for such analysis include mineral element fingerprinting (Jiang, 2018), stable isotope identification (Jin *et al.*, 2018), near-infrared spectroscopy (Zhang *et al.*, 2018), metabonomics (Chen *et al.*, 2016) and electro-sensing using an electronic nose (Tian *et al.*, 2018).

Currently, the mineral element fingerprint technology is one of the most effective technologies in the field of food confirmation, as extensive research on pepper, pear, green tea, cowpea and rice has been conducted using this technology (Coelho *et al.*, 2018; Gonzálvez *et al.*, 2011; Mehari *et al.*, 2019; Michael *et al.*, 2019; Zhang *et al.*, 2019a). Metabonomics is a method of detecting low-molecular-weight metabolites, such as organic acids, fatty acids, amino acids and sugars, in biological samples through high-throughput screening, data processing, information integration and biomarker recognition (Xiao *et al.*, 2018). Among the metabolites mentioned above, fatty acids are critical in confirming the geographical origin of landmark products, such as coffee beans (Mehari *et al.*, 2019), sea cucumber (Zhang *et al.*, 2017), tea (Hao *et al.*, 2016), honey (Alessandro *et al.*, 2018) and scallop (Zhang *et al.*, 2019b).

Soybean samples studied in this work were obtained from distantly located regions. Differences in geographical location, climatic environment and soil may significantly affect the elemental composition and the fatty acid content of soybean. Besides, during the process of confirming the origin of agricultural products, the specific indicators that can represent the origin information are affected by many external factors. Thus, only one traceability technology does not sufficiently support the stable selection of specific indicators, and it is challenging to comprehensively detect and confirm the complex origin of samples with a single technology (Ma *et al.*, 2014). Therefore, in this study, inductively coupled plasma mass spectrometry (ICP-MS) was used to determine various elements present in soybean obtained from the main production areas in northern China. Gas chromatography (GC) was used to determine the changes in the composition of fatty acids. By combining principal component analysis and cluster analysis, effective and stable traceability indexes were identified. The production area information of all the four provinces was compared in pairs to explore the regional specificity and commonality between fatty acid and element contents in soybean samples from each province. Furthermore, based on the distribution of the climatic environment and the soil types, this study explored the reasons for the differences in various indexes between different production areas. It also classified the fingerprint characteristics of fatty acids and elements for each production area using thermography visualisation.

Heat map clustering, a visualisation tool, can rapidly and easily classify a large amount of data. To the best of our knowledge, this is the first study in which heat map clustering has been used to analyse the fatty acid and multi-element fingerprint features of soybean samples for confirming the origin of soybean production.

Therefore, ICP-MS and GC were used to study the contents of various elements and fatty acids in soybean samples from different production areas in northern China, with the aim of providing a method for identifying metabolic fingerprints from different production areas using cluster thermography. This method was employed for the identification of soybean production areas in different regions, providing a reference for the establishment of a soybean origin traceability system. This research may help protect the stability of the agricultural product market and the legitimate rights and interests of consumers.

Materials and Methods

Sampling and pre-treatment

In total, 91 soybean samples were obtained from different production areas across northern China. Among these, 30 (H1–H30) were from Heilongjiang, 20 (N1–N20) from the Inner Mongolia Autonomous Region, 21 (L1–L21) from Liaoning and 20 (J1–J20) from Jilin. The collected samples were put in a self-sealing bag and stored in a dark, ventilated place at room temperature. All samples used in this study were provided by the local academy of agricultural sciences and research institutes. Figure 1 shows the locations of the production areas and provides the number of samples obtained from each production area, while Table 1 summarises the climatic characteristics of each production area.

The harvested soybeans were washed thrice with ultrapure water to remove dust from the epidermis. Subsequently, the soybean epidermis was dried at 30° for about 10 h in an electrothermal constant-temperature drying oven. The estimated water content of soybeans after being dried was about 6%. After drying, the soybean skin was broken by a hammer cyclone mill (IXFM110 hammer cyclone mill, Hangzhou Dacheng Photoelectric Instrument Co., Ltd., Hangzhou, China) to obtain a uniform soybean powder sample, which was stored in a self-sealing bag at room temperature.

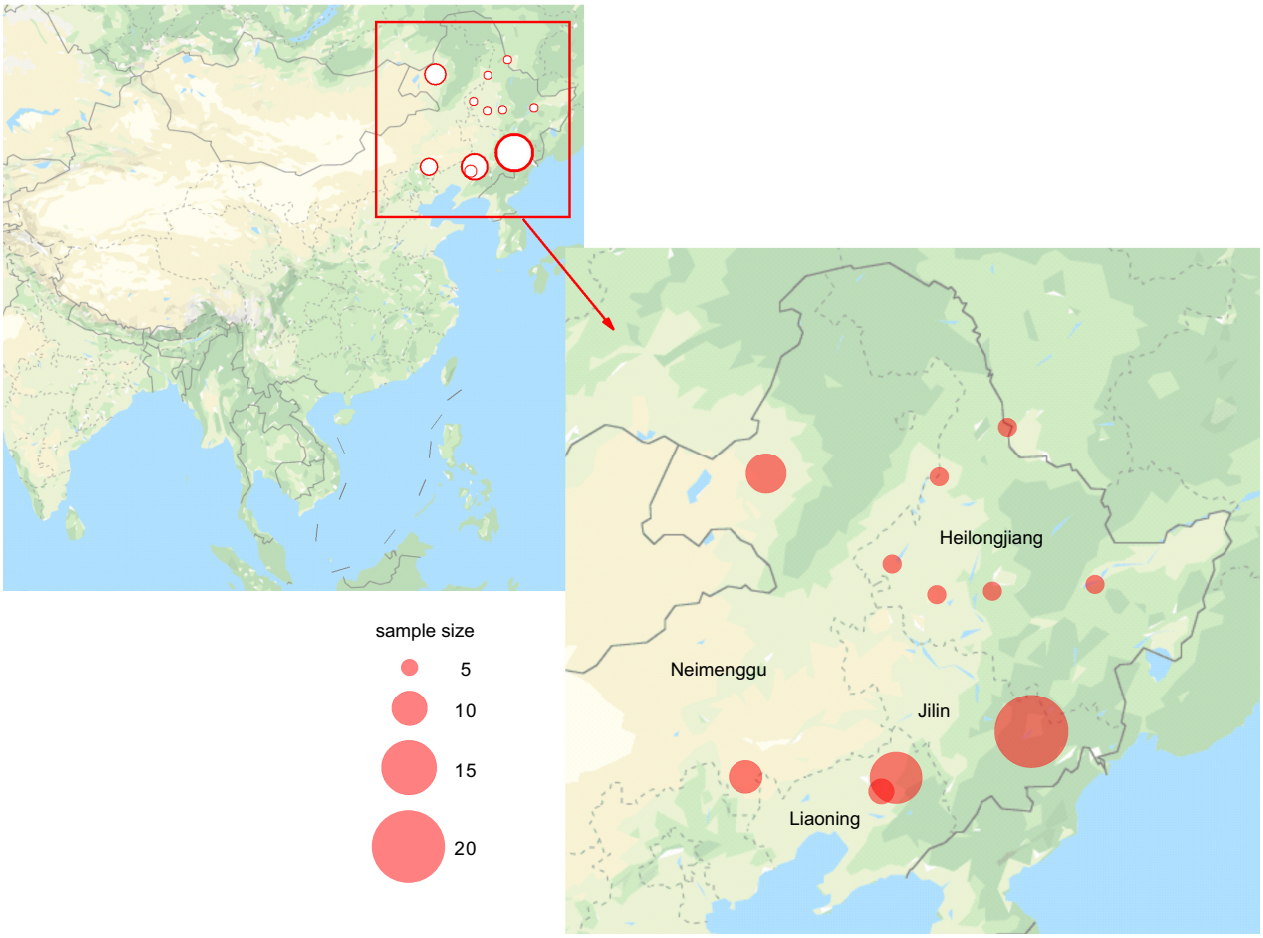

Figure 1. Locations of production areas from where soybean samples were obtained, and the number of samples obtained from each area.

Table 1. Climatic characteristics of production areas where soybean samples were obtained.

Provinces	Production area		Average annual temperature, °C	Average annual precipitation, mm	Annual sunshine, h
Heilongjiang province (n = 30)	Suihua	(H1–H5)	2.5	483	2600
	Qiqihaer	(H6–H10)	3.2	415	2600
	Jiamusi	(H11–H15)	3.0	527	2590
	Heihe	(H16–H20)	−1.3	550	2360
	Nenjiang	(H21–H25)	2.4	600	2300
	Daqing	(H26–H30)	4.2	428	2726
Inner Mongolia Autonomous Region (n = 20)	Chifeng	(N1–N9)	8.2	371	3100
	Hulunbuir	(N10–N20)	2.4	75	2700
Liaoning province (n = 21)	Tieling	(L1–L14)	6.3	700	2700
	Shenyang	(L15–L21)	6.2	800	2700
Jilin province (n = 20)	Dunhua	(J1–J20)	2.9	630	2500

Determination of element contents

In total, 0.250 g of evenly ground soybean powder was placed into an acid-washed microwave digestion inner tank. To this, 3 mL of concentrated nitric acid (70% mass fraction; Beijing Chemical Reagent Factory, Beijing, China) and 2 mL of hydrogen peroxide (30% mass fraction; Beijing Chemical Reagent Factory, Beijing, China) were added. Then, the following heating programme was run: 80° for 3 min, 100° for 3 min, 130° for 3 min, 160° for 3 min and finally, 190° for 25 min. After digestion, the mixture was cooled to room temperature. Subsequently, the digestion tank was placed on a temperature-controlled electric heating plate (G-400 intelligent temperature control electric heater, Shanghai Yiyao Instrument Technology Development Co., Ltd., Shanghai, China), which was heated to 140° for 2–3 h to remove the residual acid. Afterward, the mixture was cooled using 25 mL of ultrapure water (Mill-Q ultra-pure water system, Milibo (Shanghai) Trading Co., Ltd., Shanghai, China) to obtain a testing-ready transparent solution. Using the same method, two groups of blank reagent control and standard substance control were prepared.

The contents of 17 elements (Mg, Al, Ca, Cr, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Rb, Sr, Mo, Pd, Cd) in soybean were determined using the Thermo XSeries2 instrument (Shenzhen Ruisheng Technology Co., Ltd, Shenzhen, China) for ICP-MS. The main working parameters of ICP-MS in this experiment were as follows: radio frequency (RF) power, 1325 W; plasma gas flow rate (cooling gas flow rate), 15 L/min; carrier gas flow (atomiser flow), 0.8 L/min; auxiliary air flow rate, 0.40 L/min; atomisation chamber temperature, 2°. The GBW10013 soybean component analysis standard substance (GSB-4) was used as a standard to verify the accuracy of the analysis method. In, Ge, Rh and Re were used as internal standards; if the relative standard deviation (RSD) of these standards was ≤5%, the instrument was considered stable. Each sample was analysed thrice.

Determination of fatty acid contents

In total, 0.500 g of evenly ground soybean powder was put into a dry glass tube with a screw. Next, 5 mL of toluene (AR; Shanghai Wokai Biotechnology Co., Ltd., Shanghai, China) and 6 mL of 10% acetyl chloromethanol (prepared with methanol, AR; Merck AG, Darmstadt, Germany; and acetyl chloride, 98%, AR; Shanghai Aladdin Biochemical Technology Co., Ltd. Shanghai, China) were added sequentially, the mixture was shaken and the tube was filled with nitrogen for 25 s. Furthermore, the tube was placed in a water bath at 80° for 2 h and shaken every 30 min. The mixture was then cooled and transferred to a 50 mL centrifuge tube. The glass tube was washed thrice with 3 mL of 6% sodium carbonate solution (AR; Shanghai Guoyao Group Chemical Reagents Co., Ltd. Shanghai, China) to dissolve the residual mixture. The sodium carbonate solution was then mixed with the mixture in the 50 mL centrifuge tube. The mixture was then centrifuged at 5000 rpm. Finally, 1.50 mL of the supernatant was filtered through a Nylon 66 Jinteng organic filter membrane (0.22 μm, 13 mm) using a syringe; it was then added to a sample injection vial and stored at 4°.

The content of five fatty acids—namely oleic acid (C18:1), linoleic acid (C18:2), linolenic acid (C18:3), palmitic acid (C16:0) and stearic acid (C18:0)—in soybean was determined using the Agilent 7890A GC system (containing flame ionisation detector; Agilent Technologies Co., Ltd., Palo Alto, California, USA). The required purity of the nitrogen was more than 99.9999%. The chromatographic conditions in the test were as follows: capillary column, CP-Sil 88 for FAMES (100 m × 0.25 mm × 0.25 m, CP 7420; Agilent Technologies (China) Co., Ltd., Beijing, China); inlet temperature, 270°; injection volume, 1 L; diversion ratio, 25:1; flame ionization detector (FID) temperature, 280°; carrier gas, nitrogen. The details regarding the programmed heating method were as follows: 1 mL/min constant current mode; initial temperature, 140° for 5 min; 140° to 240° at a heating rate of 4°/min; and maintenance at 240° for 15 min. The detection time for each sample was 45 min. Peak area normalisation was done to calculate the relative contents of fatty acids, and the standard external method was used to calculate the absolute content of each fatty acid. Each sample was analysed thrice.

Data processing and statistical analysis

The SPSS24.0 software was used to pre-process and statistically analyse the continuous variables obtained through ICP-MS and GC. While measuring, the accuracy of data was checked by adding the standard to the sample and measuring the spiked recoveries and RSDs of elements and fatty acids. When the spiked recoveries of fatty acids and elements exceeded 80–120%, the RSDs of the internal standard elements were greater than 5%, the RSDs of fatty acids were greater than 10% and the samples were re-measured. The average value of three determinations for each sample was used for data analysis. First, correlation analysis was conducted to investigate whether there is a correlation between the variables, and then the correlation degree and covariant trends of variables between the samples from each production area were measured. Furthermore, a principal component analysis was conducted to classify the soybean samples and to verify the differences in element and fatty acid contents between the four provinces. The data were subjected to hierarchical clustering to explore the regional

specificity and commonality of data concentration distribution between each production area, and all test data were standardised (Z-Score method) for thermographic visualisation analysis. Finally, according to the selected characteristic indexes, a discriminant model was established using partial least squares discriminant analysis (PLS-DA) to observe its prediction ability and to test the clustering of elements, fatty acids and other indexes.

Results

Correlation between element and fatty acid contents in soybean samples

The correlations between the contents of 17 elements and five fatty acids in soybean samples procured from different production areas in the four provinces were analysed (Table 2). We noted that the contents of most elements and fatty acids in soybean samples from each production area were significantly correlated. The contents of some of these elements, such as Mg/Zn, Al/Mo, C16:0/C18:0 and C18:2/C18:3, positively correlated, indicating that they have a specific bidirectional absorption assisting effect and are interdependent and mutually promoting. Some substances, such as Zn/Cd, Fe/Cu, C18:0/C18:3 and C18:1/C18:2, showed negative correlations, indicating that they have unidirectional or bidirectional absorption inhibition. Furthermore, according to the results, there was a certain correlation between element content and fatty acid content in soybean. For example, Cu/C16:0 and Rb/C18:0 were significantly positively correlated, while Ca/C16:0 and Cd/C18:3 were negatively correlated. It could be seen that during the growth of soybean, some elements and fatty acids exhibit synergistic and antagonistic effects, and these effects can change according to the environment of the place of origin.

Principal component analysis of element and fatty acid contents in soybean samples

In order to reduce as much as possible the interference of overlapping information in data analysis, the principal component analysis was performed on the data pertaining to the contents of 17 elements and five fatty acids. This analysis involved a preliminary clustering of samples before further hierarchical clustering. The resulting rotated component matrix is shown in Table 3.

The first six principal components were obtained by extracting components with eigenvalues >1. The cumulative variance contribution rate accounted for 89.446% of the total variance, which was sufficient to represent most of the information of the original sample. Therefore, we initially thought that the variables included in the first six principal components—Mn, Ni, Se, Sr, Pd, Cd, Ca, Fe, Zn, Rb, C18:0, Co, C18:2, Al, Mo, Mg and C18:1—can be used as powerful traceability indicators, while the variables Cr, Cu, As, C16:0 and C18:3 were not reflected in the extracted principal component.

The cumulative variance contribution rate of the first three principal components accounted for 65% of the total variance. Thus, we selected the first three principal components, PC1, PC2 and PC3, to generate score charts and scatter charts. We obtained the score of each observed value and the load of each variable, as shown in Figure 2.

From the scores of each production area shown in Figure 2A, we noted that the soybean samples from Jilin and Liaoning could be effectively distinguished, and the soybean samples from Heilongjiang could be well separated from the Jilin and Liaoning samples. When PC3 was considered, samples from the production areas in Jilin and the Inner Mongolia Autonomous Region could be distinguished, as shown in Figure 2B. From the load of each variable, it was observed that the variables Ni, Sr, Se, Mn and C18:1 point to Liaoning origin, indicating that these variables may be related to the characteristics associated with Liaoning origin. The variables Mo and Rb point to the origin of the Inner Mongolia Autonomous Region; Cr points to the origin of Jilin; and Mg, Fe, C18:2 and C18:3 point to the origin of Heilongjiang. Therefore, Mo and Rb may be related to the traceability characteristics associated with the Inner Mongolia Autonomous Region, and Cr may be related to the Jilin origin, whereas Mg, Fe, C18:2 and C18:3 seem to be indicators of the Heilongjiang origin.

However, this analysis alone was not sufficient to accurately classify all the variables, because it only involves the natural clustering of all samples, and the load map only provides a preliminary assumption of the distribution of all variables. Therefore, hierarchical clustering analysis was conducted on all samples to obtain more accurate feature classification of each production area.

Hierarchical clustering analysis of fatty acid and element contents in soybean samples

Using the Ward clustering algorithm, potential markers were constructed to realise clustering visualisation based on Euclidean distance measurement (Zhao *et al.*, 2014). Firstly, 22 variables in all samples of Heilongjiang, the Inner Mongolia Autonomous Region, Liaoning and Jilin were clustered hierarchically, and the classification of different production areas and variables was presented in combination with heat map visualisation (Figure 3).

Table 2. Pearson correlation matrix of fatty acid and element contents in soybean.

	Mg	Al	Ca	Cr	Mn	Fe	Co	Ni	Cu	Zn	As	Se	Rb	Sr	Mo	Pd	Cd	C16:0	C18:0	C18:1	C18:2	C18:3
Mg	1.000																					
Al	0.004	1.000																				
Ca	0.373**	−0.183	1.000																			
Cr	−0.212*	0.459**	−0.154	1.000																		
Mn	0.080	−0.260*	0.095	−0.353**	1.000																	
Fe	0.637**	−0.180	0.548**	−0.057	0.243*	1.000																
Co	−0.135	−0.049	−0.316**	−0.170	0.214*	−0.081	1.000															
Ni	−0.261*	−0.329**	0.024	−0.389**	0.723**	−0.074	0.018	1.000														
Cu	−0.593**	0.408**	−0.568**	0.405**	−0.062	−0.684**	−0.052	0.161	1.000													
Zn	0.378**	−0.579**	0.447**	−0.406**	−0.021	0.473**	−0.266*	0.019	−0.767**	1.000												
As	−0.239*	0.269**	0.145	−0.048	−0.305**	−0.482**	−0.411**	0.037	0.242*	0.058	1.000											
Se	−0.039	−0.428**	0.245*	−0.375**	0.871**	0.254*	0.097	0.782**	−0.137	0.167	−0.194	1.000										
Rb	−0.331**	0.347**	−0.079	−0.150	−0.229*	−0.390**	0.449**	−0.076	0.054	−0.198	0.441**	−0.187	1.000									
Sr	−0.280**	−0.561**	0.009	−0.455**	0.649**	−0.043	0.297**	0.881**	0.014	0.124	−0.106	0.752**	−0.021	1.000								
Mo	0.384**	0.622**	0.139	0.064	0.063	0.176	0.114	−0.311**	−0.122	−0.288**	−0.020	−0.157	0.272**	−0.500**	1.000							
Pd	−0.037	−0.267*	0.297**	−0.216*	0.498**	0.488**	0.182	0.445**	−0.361**	0.289**	−0.328**	0.596**	0.037	0.393**	0.088	1.000						
Cd	−0.312**	0.114	−0.198	−0.176	0.577**	−0.305**	−0.004	0.818**	0.470**	−0.359**	0.155	0.536**	0.084	0.559**	0.044	0.270**	1.000					
C16:0	−0.527**	0.170	−0.627**	0.072	0.056	−0.690**	−0.053	0.448**	0.796**	−0.507**	0.275**	0.035	0.057	0.327**	−0.376**	−0.337**	0.601**	1.000				
C18:0	−0.330**	0.275**	−0.161	−0.267*	0.170	−0.364**	0.487**	0.390**	0.213*	−0.371**	0.265*	0.177	0.805**	0.368**	0.152	0.161	0.503**	0.375**	1.000			
C18:1	0.104	−0.193	0.231*	−0.669**	0.421**	0.020	0.015	0.621**	−0.328**	0.409**	0.277**	0.484**	0.272**	0.508**	0.020	0.418**	0.445**	0.046	0.476**	1.000		
C18:2	0.337**	−0.001	0.176	0.258*	−0.275**	0.462**	0.309**	−0.727**	−0.413**	0.082	−0.481**	−0.356**	−0.091	−0.506**	0.313**	0.020	−0.728**	−0.719**	−0.422**	−0.576**	1.000	
C18:3	0.221*	0.006	0.042	0.429**	−0.360**	0.310**	0.092	−0.771**	−0.182	0.015	−0.415**	−0.430**	−0.314**	−0.575**	0.138	−0.196	−0.738**	−0.541**	−0.657**	−0.774**	0.884**	1.000

Note: * Indicates significant correlation with $P < 0.05$ (double-tailed).
** Indicates significant correlation with $P < 0.01$ (double-tailed).

Table 3. The rotated component matrix.

	Components					
	1	2	3	4	5	6
Mg	−0.145	0.420	−0.443	0.119	0.392	0.579
Al	−0.212	−0.246	0.180	−0.184	0.815	−0.248
Ca	0.090	0.815	−0.071	−0.297	0.035	−0.009
Cr	−0.313	−0.131	−0.209	0.001	0.217	−0.779
Mn	0.865	0.078	−0.134	0.273	0.076	0.132
Fe	0.109	0.764	−0.373	0.275	0.129	0.077
Co	0.071	−0.114	0.574	0.761	−0.022	0.108
Ni	0.932	−0.137	0.051	−0.156	−0.226	0.087
Cu	0.104	−0.812	0.018	−0.153	0.120	−0.468
Zn	−0.061	0.644	−0.161	−0.208	−0.470	0.422
As	−0.131	−0.117	0.324	−0.851	0.027	0.024
Se	0.882	0.218	−0.063	0.106	−0.173	0.042
Rb	−0.116	−0.037	0.956	−0.113	0.167	0.042
Sr	0.777	−0.088	0.160	0.101	−0.501	0.142
Mo	−0.075	0.221	0.146	0.118	0.899	0.066
Pd	0.600	0.578	0.189	0.208	−0.054	−0.154
Cd	0.816	−0.386	0.115	−0.206	0.190	−0.041
C16:0	0.288	−0.878	0.050	−0.256	−0.108	−0.078
C18:0	0.361	−0.242	0.823	−0.051	0.191	0.117
C18:1	0.562	0.205	0.331	−0.344	−0.039	0.563
C18:2	−0.573	0.450	−0.085	0.610	0.117	−0.158
C18:3	−0.655	0.243	−0.332	0.495	0.016	−0.295
Percentage of variance (%)	30.127	23.588	11.903	10.791	8.257	4.779

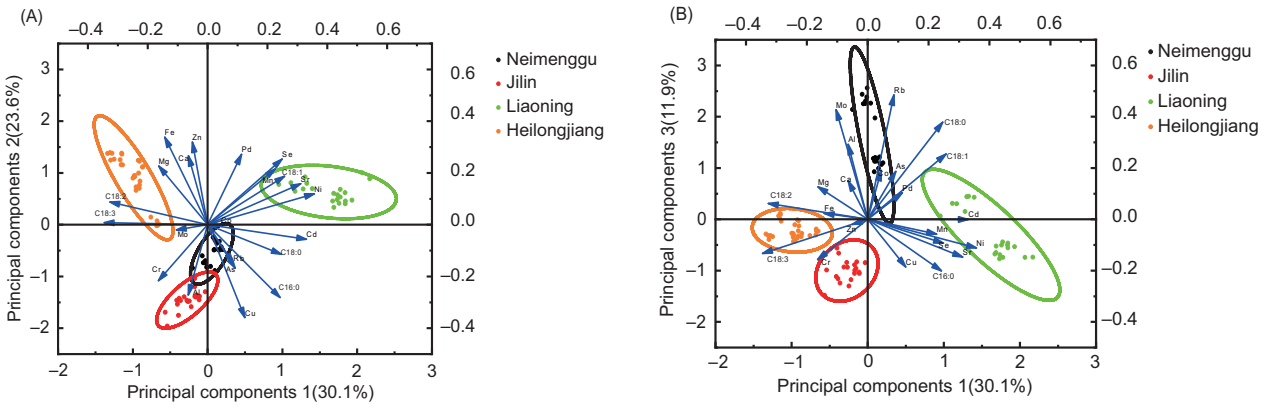

Figure 2. The first three common factor scores and loads of fatty acid and element contents in soybeans from different production areas. (A) PC1 and PC2, (B) PC1 and PC3.

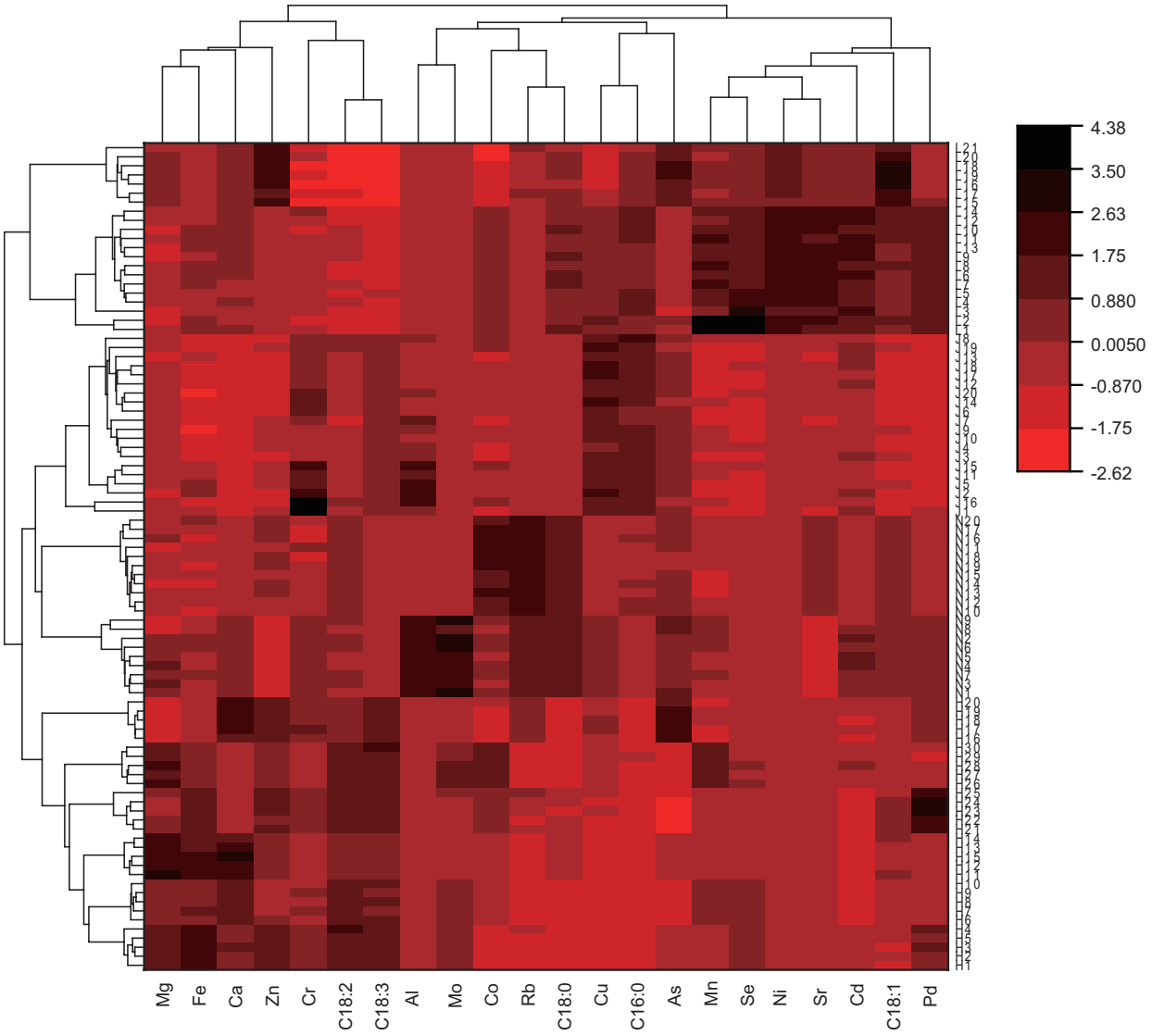

Figure 3. Cluster heat map of elements and fatty acid contents of soybean samples.

The first layer of the hierarchical tree presented on the left-hand side of Figure 3 divides the production areas into two categories according to different classification distances. One category is Liaoning, and the other constitutes Jilin, the Inner Mongolia Autonomous Region and Heilongjiang. The second layer of classification divides all samples into five categories according to their place of origin: Heilongjiang, the Inner Mongolia Autonomous Region, Jilin, Liaoning Tieling and Liaoning Shenyang. Although Liaoning is divided into two categories, it could still be generally separated from the other three provinces.

In the upper hierarchical tree, the first layer divides 22 variables into two categories according to different classification distances. The second layer divides 22 variables into three categories. The third layer is divided into seven categories. The first category comprises Mg, Fe, Ca and Zn; the second one comprises Cr, C18:2 and C18:3; the third comprises Al and Mo; the fourth comprises Co, Rb and C18:0; the fifth one comprises Cu and C16:0; the sixth one has only As; and the seventh one comprises Mn, Se, Ni, Sr, Cd, C18:1 and Pd. Among these, the classification variables of the categories 7, 1 and 4 are consistent with the variables extracted by the first, second, and third principal components, respectively. Therefore, upon general comparison, the variables Mn, Se, Ni, Sr, Cd, C18:1, Pd, Mg, Fe, Ca, Zn, Co, Rb and C18:0 may represent the origin information of soybeans collected from the four provinces.

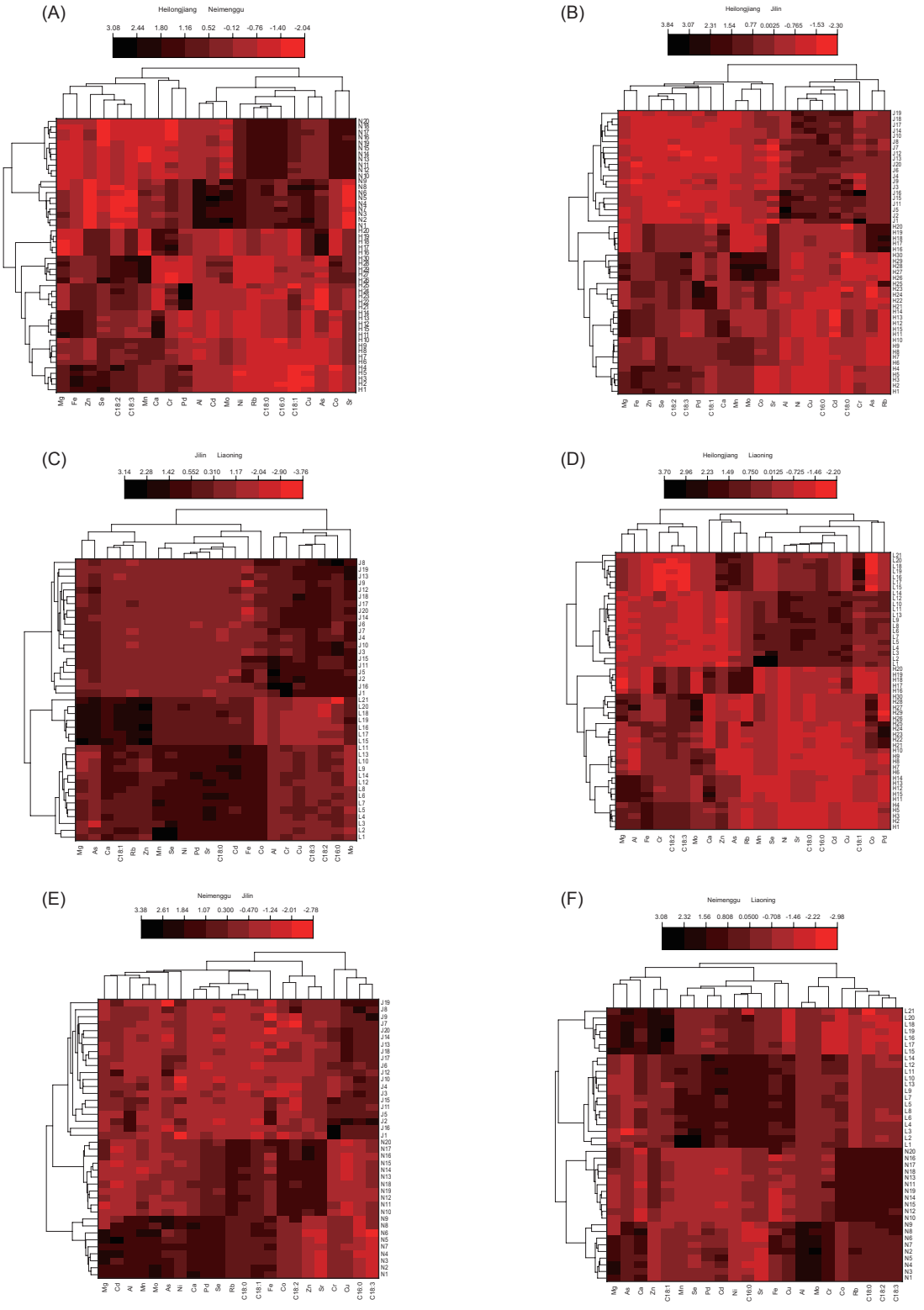

Figure 4. Cluster heat maps of pairwise comparison of fatty acid and element contents between Heilongjiang and Neimenggu(A), Heilongjiang and Jilin(B), Jilin and Liaoning(C), Heilongjiang and Liaoning(D), Neimenggu and Jilin(E) and Neimenggu and Liaoning(F).

We attempted to extract the key characteristic indexes of each production area from the heat map. In the soybean samples from Heilongjiang, the contents of Mg, Ca and Pd were higher than those from the other three provinces, whereas the contents of Mo in the soybean samples from the Inner Mongolia Autonomous Region were higher than those in the other three provinces. Similarly, the contents of Cr in the soybean samples from Jilin were higher than those in the other three provinces, and the contents of Zn, Mn, Se and C18:1 in the samples from Liaoning were higher than those from Heilongjiang, the Inner Mongolia Autonomous Region and Jilin.

To further explore the significant information and specific indicators of each production area in northern China, we compared variables of each pair of the four provinces, and results were visualised with heat maps (Figure 4).

Comparison of different production areas within the four regions

As shown in Figure 4A, Heilongjiang and the Inner Mongolia Autonomous Region were effectively distinguished. In Heilongjiang, Mg and Ca contents in the samples from Jiamusi (H11–H15); Fe, Se and C18:2 contents in the samples from Suihua (H1–H5); Mn and C18:3 contents in the samples from Daqing (H26–H30); Pd in the samples from Nenjiang (H21–H25); and As in the samples from Heihe (H16–H20) were significantly higher than those in the Inner Mongolia Autonomous Region. In this region Al, Cd, Mo, Cu and C18:1 were observed to be higher in the samples from the Chifeng production area (N1–N9), whereas Co and Sr contents in the samples from the Hulunbuir production area (N10–N20) were higher than those in Heilongjiang.

As shown in Figure 4B, the two production areas could be effectively distinguished. We noted that Mg in the samples from Jiamusi (H11–H15); Pd in the samples from Nenjiang (H21–H25); C18:3, Mn, Mo and Co in the samples from Daqing (H26–H30); and As and Rb in the samples from Heihe (H16–H20) were significantly different from those in the samples from Jilin. The Al and Cr contents in the samples from Jilin Dunhua (J1–J20) were higher than those in the samples from Heilongjiang.

As shown in Figure 4C, each production area in the two provinces could be distinguished. The contents of Mg, As, Rb, Zn and C18:1 in the samples from Liaoning Shenyang (L15–L21) and Mn and Se in the samples from Liaoning Tieling (L1–L14) were higher than those in the samples from Jilin Dunhua. The contents of Al, Cr and C16:0 in the samples from Jilin Dunhua (J1–J20) were significantly different from those in the samples from Liaoning.

As shown in Figure 4D, the production areas of the two provinces could be distinguished. Among these, Al and Mo in the samples from Daqing (H26–H30) and Pd in the samples from Nenjiang (H21–H25) in Heilongjiang were higher than those in the samples from Liaoning. On the contrary, Mn and Se in the samples from Tieling (L1–L14) and C18:1 in the samples from Shenyang (L15–L21) were higher than those in the samples from Heilongjiang.

As shown in Figure 4E, the two provinces could be effectively distinguished from each other. The contents of Mg, Cd, Mo, As and Ca in the samples from Chifeng (N1–N9) and Co, Zn, Sr and C18:2 in the samples from Hulunbuir (N10–N20) were significantly different from those in the samples from Dunhua in the Jilin province. However, the contents of Cr, C16:0 and C18:3 in the samples from the Jilin Dunhua production area (J1–J20) were significantly different from those in the samples from the Inner Mongolia Autonomous Region.

As shown in Figure 4F, the two production areas could be distinguished. Among these, the contents of Al and Mo in the samples from Chifeng (N1–N9) and Co, Rb, C18:0, C18:2 and C18:3 in the samples from Hulunbuir (N10–N20) were higher than those in the samples from Liaoning. The contents of Zn and C18:1 in the samples from Liaoning Shenyang (L15–L21) and Mn and Se in the samples from Tieling (L1–L14) were higher than those in the samples from the Inner Mongolia Autonomous Region.

Partial least squares discriminant analysis of fatty acid and elements contents in soybean samples

PLS-DA is a multivariate statistical analysis method used for discriminant analysis. It uses partial least squares regression to establish the relationship model between each expression and sample category, and reduces the dimensions of the data. This monitoring mode can generally better establish relationships among samples. The prediction model can also predict the sample category and identify more samples.

In order to verify whether Mn, Se, Ni, Sr, Cd, Pd, Mg, Fe, Ca, Zn, Co, Rb, C18:1 and C18:0 can be used as classification indexes to identify and distinguish different production areas, we conducted PLS-DA on the above 14 classification indexes, and the results are shown in Figure 5.

In the partial least squares discriminant model, R^2X and R^2Y represent the percentages of X and Y matrix information, respectively. In this model, $R^2X = 0.709$, indicating that the three prediction principal components of the model could explain 70.9% of X variable information.

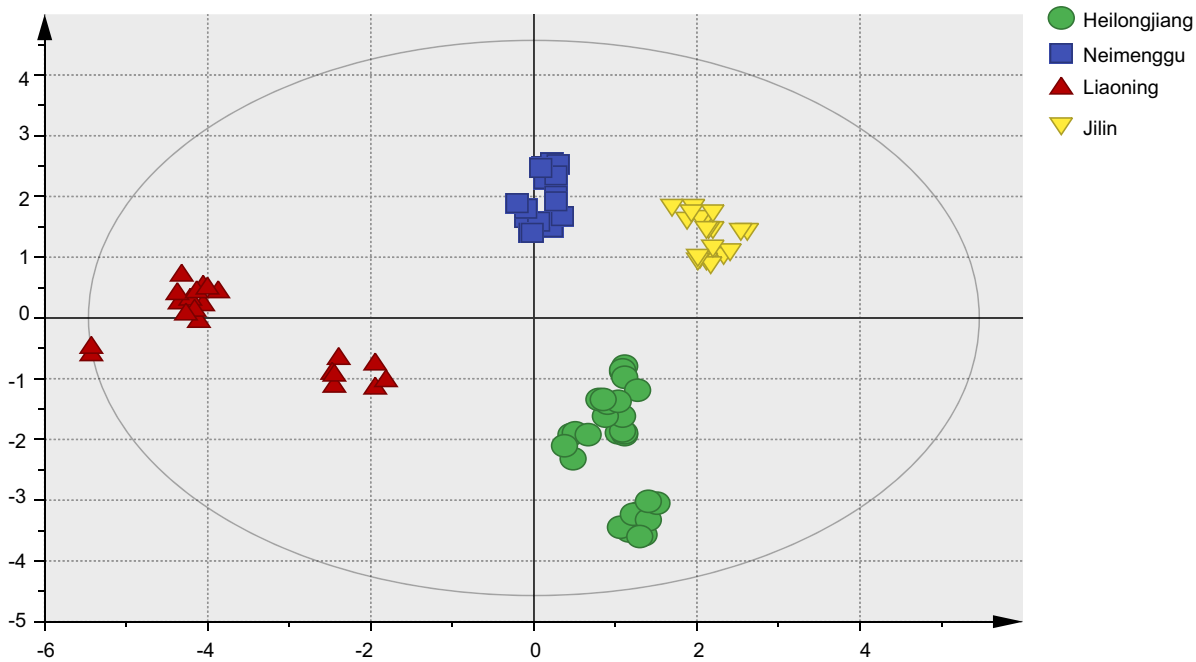

Figure 5. Scatter plot of partial least squares discriminant analysis model scores for soybean samples from different origins.

R^2Y = 0.902, indicating that the three predictive principal components of the model had 90.2% explanatory power to Y variables. Q^2 indicates that the prediction ability of the evaluation model is obtained through cross-validation calculation. Here, Q^2 = 0.899, which indicates that the PLS-DA model had 89.9% prediction ability for soybean samples from Heilongjiang, the Inner Mongolia Autonomous Region, Jilin and Liaoning. This eventually confirmed that the model is reliable for discriminating the soybean production areas of the four provinces. The score chart of Figure 5 shows that all soybean samples were divided into different regions, and the soybean samples from each production area had obvious aggregation. This result indicated that the 14 classification indexes contain sufficient information on soybean production areas and that they can be used as specific indexes to accurately identify and distinguish the soybean samples from the four production areas.

Discussion

The content and distribution of mineral elements and fatty acids in agricultural products are closely related to the natural environment in which they are cultivated. There are marked differences in the distribution of mineral elements and fatty acids in different production areas, and this difference can be reflected in agricultural products, thus serving as specific fingerprints in some areas. Among them, some elements are mainly affected by the material exchange from rock to soil and from soil to plant body in the growing environment, and therefore different soil types affect the different element contents in agricultural products (Chung *et al.*, 2015). For example, chernozem is rich in inorganic elements such as Mg, Ca and K, while black soil is rich in elements such as Fe, Zn, Mn, Cu and Mo. The composition of fatty acids is closely related to climate (temperature and precipitation) and geography (altitude, latitude and longitude). Therefore, we analysed the climatic environment and geographical characteristics of each production area in the four provinces, including the annual average temperature, annual average precipitation, annual sunshine and soil distribution.

Geographical specificity of fatty acids

The content of fatty acids in agricultural products is not only determined by the heredity of their varieties but is also related to environmental factors such as phenological period, temperature, moisture and light (Sun *et al.*, 2014). Temperature and precipitation can regulate the content of saturated and unsaturated fatty acids by affecting the activity of fatty acid desaturase in soybean plants. Different parts of plants have different responses to temperature. Under low temperature, more unsaturated fatty acids are synthesised in roots, but less in leaves. This is due to the increase of fatty acid desaturase activity in plants under a low temperature environment. Metabolism produces more unsaturated fatty acids. High

temperature increases the content of saturated fatty acids in plants, which, in turn, enhances heat resistance (Sun *et al.*, 2014). The content of C18:3 decreases and the content of C18:0 gradually increases when the water content of soybean decreases (Dornbos and Mullen, 1992).

In this study, the annual average temperature of each production area in Heilongjiang was lower than that of most production areas in other provinces, and the contents of unsaturated fatty acids, C18:2 and C18:3, were higher than those in soybean samples from other provinces. The annual average temperature of the two production areas in Liaoning was the highest, and the C16:0 content was obviously higher than that in soybean samples from Heilongjiang and the Inner Mongolia Autonomous Region. For oleic acid (C18:1), some research results showed that a high average temperature during the growth period of soybean is beneficial for the increase in oleic acid content (Cao *et al.*, 2015). While the annual average temperature in the Liaoning province was higher than that in other production areas, the content of C18:1 was obviously higher than that in soybean samples from other production areas. The relatively low temperature in Hulunbuir makes the metabolism of rich organic matter in soil stable. In this study, the average annual precipitation in the Inner Mongolia Autonomous Region was lower than that in the other provinces. Also, the content of C18:3 fatty acid in soybean samples from this province was slightly lower than that in soybean samples from Heilongjiang and Jilin, while the content of C18:0 was relatively higher than that in the other three provinces.

Geographical specificity of elements

Soil organic matter plays an important role in the process of "soil–plant" transfer of mineral elements through its influence on the physicochemical properties, moisture and structure of the soil. Change in temperature or precipitation changes the composition of organic matter in soil, thus changing its adsorption on soil structure, cation substitution, nutrient components and metal ions. The results of related studies confirmed that an increase in temperature affects the activity of microorganisms and leads to the decrease of soil organic matter content, while an increase in precipitation increases soil humidity and decreases surface temperature, thus increasing soil organic matter content (Li *et al.*, 2014). Therefore, in different natural environments, the differences of temperature and precipitation affect the content of organic matter in soil, which fundamentally changes the physical and chemical properties of soil and the form of mineral elements, eventually affecting the content of elements in soybean.

The soil types in Jiamusi, Daqing and Suihua production areas in the Heilongjiang province are mainly meadow soil, black soil and chernozem, respectively. The humus layer is relatively thick and neutral to slightly alkaline. Chernozem is rich in inorganic elements such as Mg, Ca and K, whereas black soil is relatively rich in Fe, Zn, Mn, Cu and Mo. Hailun in Suihua city is located in the core area of black soil and the Se-rich soil belt in the cold region of Songnei plain. This region is called the "selenium capital of black soil in China," and thus the contents of Se in the Suihua production area are relatively high. The soil types in the Heihe River and Nenjiang River are mainly dark brown soil and black soil, respectively, and the black soil area in the Songnei plain is rich in elements such as B, Se, Rb, Sn, Cr, Cu and Ni (Cui *et al.*, 2008). The soil in Jilin is mostly black and dark brown, and it is rich in Cr; its main sources are chemical weathering and ores (Wang *et al.*, 2020). Black soil, dark brown soil, chernozem soil and meadow soil are the main types of soils in Hulunbuir and the Inner Mongolia Autonomous Region (Yun *et al.*, 2013). In these soils, organic matter accumulation is high, and humus content is rich. Humus is the most important organic compound capable of chelation in soil because it contains many chelating groups, such as hydroxyl, amino and carboxyl groups. It has a strong fixation ability for many elements such as Fe, Mn, Zn, Co and Sr (Liao, 2004). The types of soil in the Chifeng production area are relatively complex; these are mainly brown soil, cinnamon soil, chernozem soil and meadow soil. The contents of Al, Cd, Cu and As in this production area are relatively high. Research shows that the background values of Cd, Cu and As in Chifeng soil are relatively high (Gu *et al.*, 1995). This may also be related to the pollution caused by metal elements in farmland soil (Dai *et al.*, 2014). Compared with the other three provinces, the contents of Mn, Se and Zn in the soils of Shenyang and Tieling in the Liaoning province are higher (Wang, 2012; Wu, 1986)

In this study, as the soil type and the climatic environment of each production area were different, the accumulation of the organic matter and the content of humus were also different. It is the regional specificity of this soybean production area that affects the content of elements and fatty acids in soybean from each production area. By comparing the heat maps, it was found that in the four production areas, the contents of fatty acids and elements in soybean samples in Heilongjiang were higher, followed by Liaoning. Heilongjiang has a low temperature, abundant precipitation and black soil aggregation. The favourable ecological environment has greatly maintained the material exchange of various organic matter in soil in the "soil–soybean plant" system and promoted the production of high-quality soybeans. Therefore, Heilongjiang has the largest planting area among the four provinces and the largest soybean planting area in China. Particularly, compared with other production areas in Heilongjiang, soybean from Suihua, Heihe and Daqing had a higher content of fatty acids and elements because of the ideal

geographical features in these areas. The annual average precipitation and the number of days with sunshine in Liaoning were higher than those in the other three production areas. Rich precipitation greatly increases the transformation and accumulation of organic matter in soil, while sufficient sunshine in the soybean growth period promotes the increase of some fatty acids. Among all environmental factors, the most important one is geographical location, followed by average annual temperature, average annual precipitation and finally soil organic matter (Wang *et al.*, 2019). Therefore, when planting soybeans, it is necessary to plant them densely and rationally and improve the overall ventilation and light transmission conditions. Generally, large seeds need more water and are suitable for planting in areas with sufficient precipitation, while small seeds need less water and are mostly planted in arid areas. In order to ensure the synthesis and metabolism of elements and fatty acids, it is necessary to master irrigation and drainage according to meteorological and soil moisture conditions during the pod-setting and seed-filling period of soybean growth.

Differences in fatty acid and element content are related to not only soil characteristics and climatic conditions but also to soybean varieties and agricultural practices. In this study, we screened the combination of element and fatty acid variables directly related to the geographical origin; combined with a multivariate analysis method, the interference of genotype, availability of organisms in the soil and interaction between each variable on traceability was reduced. However, it is possible that the differences in the contents of some fatty acids and elements in soybean samples may be due to the different varieties in these four provinces. Therefore, in future studies, factors such as soybean variety and agricultural practice should be considered in the identification of geographical origins of soybean.

Conclusion

In this study, metabonomics and multi-element fingerprint technology were used to analyse the regional differences in fatty acid and element contents in soybean from different habitats. The fingerprint information of soybean production areas in the four northern provinces of China was determined by principal component analysis and heat map, and the prediction model based on PLS-DA effectively distinguished the soybean samples from the four production areas. Furthermore, this study provided a theoretical basis for the establishment of a soybean fingerprint information database and confirmation platform for the selected production areas.

In conclusion, fatty acid and multi-element fingerprint technology combined with cluster heat map visualisation analysis, as a more intuitive classification method to identify soybean sources, is expected to become a powerful analysis tool for geographical specificity research.

Acknowledgements

This work was supported by the earmarked fund for Modern Agro-industry Technology Research System (CARS-04). We would like to thank Editage (www.editage.cn) for English language editing.

References

Alessandro, Z., Dora, M., Sonia, S., Antonia, Z. and Gian, L.M., 2018. Botanical traceability of unifloral honeys by chemometrics based on head-space gas chromatography. European Food Research and Technology 244: 2149–2157. https://doi.org/10.1007/s00217-018-3123-3

Aung, M.M. and Chang, Y.S., 2014. Traceability in a food supply chain: safety and quality perspectives. Food Control 39: 172–184. https://doi.org/10.1016/j.foodcont.2013.11.007

Cao, Y.Q., Xie, F.T., Dong, L.J., Wang, Y.Z., Song, S.H. and Wang, W.B., 2015. Research progress of oleic acid content in soybean seeds. Soybean Science 34: 329–334.

Chen, Y.H., Zhang, D.J., Zhang, G.F., Wang, Y. and Wang, C.Y., 2016. Construction of DNA fingerprint of the Japonica rice in Jiansanjiang area of Heilongjiang Province. Cereal & Feed Industry 07: 16–19.

Chung, I.-M., Kim, J.-K., Lee, J.-K. and Kim, S.-H., 2015. Discrimination of geographical origin of rice (Oryza sativa L.) by multielement analysis using inductively coupled plasma atomic emission spectroscopy and multivariate analysis. Journal of Cereal Science 65: 252–259. https://doi.org/10.1016/j.jcs.2015.08.001

Coelho, I., Matos, A.S., Teixeira, R., Nascimento, A., Bordado, J., Donard, O. and Castanheira, I., 2018. Combining multielement analysis and chemometrics to trace the geographical origin of Rocha Pear. Journal of Food Composition and Analysis 77: 1–8. https://doi.org/10.1016/j.jfca.2018.12.005

Cui, Y.J., Shi, Y.M., Liu, G.D. and Yang, X., 2008. Element content characteristics of black soil in Southern Songnen plain of Heilongjiang Province. Geoscience 22: 929–933.

Dai, Y.X., Liang, Y.T., Zhang, Y.H. and Zhang, J.H., 2014. Pollution status of lead and cadmium in some farmland soils in Chifeng city. Journal of Diseases Monitor & Control 8: 391–392.

Dornbos, D.L. and Mullen, R.E., 1992. Soybean seed protein and oil contents and fatty acid composition adjustments by drought and temperature. Journal of the American Oil Chemists Society 69: 228–231. https://doi.org/10.1007/BF02635891

Gonzálvez, A., Armenta, S. and Guardia, M.D.L., 2011. Geographical traceability of "Arròs de Valencia" rice grain based on mineral element composition. Food Chemistry 126: 1254–1260. https://doi.org/10.1016/j.foodchem.2010.11.032

Gu, Y.R., Zhang, T. and Bai, H.Y., 1995. Attribute classification of soil environmental background values in inner mongolia. Inner Mongolia Environmental Protection 01: 6–9.

Hao, M., Dai, X.D., Jiang, B. and Jin, L.M., 2016. Measurement of fatty acid in three kinds of tea by GC-MS. Tianjin Agricultural Sciences 22: 15–17.

Jiang, Z.Q., 2018. Research progress on traceability of grain origin produced by mineral element fingerprint analysis technology. Farm Products Processing 05: 70–71.

Jin, X.X., Pan, L.G. and Li, A., 2018. Progress in application of stable isotope technology in agricultural product safety. Vegetables 09: 29–34.

Li, L.Q., Wu, Z.Y., Zhang, Q. and Li, Y., 2014. State of the art review of the impact of climatic change on bioavailability of mineral elements in crops. Acta Ecologica Sinica 34: 1053–1060. https://doi.org/10.5846/stxb201305141057

Liao, J.F., 2004. Effect of soil environment on trace elements in crops. In: The 6th National Symposium on Research and Progress of Trace Elements in chinese chemical society, Fujian, China, p. Chinese Chemical Cociety 2004–12:3.

Liu, L., 2014. Nutritional components of soybean and its comprehensive utilization prospect. Journal of Inner Mongolia University for Nationalities (Natural Science Edition) 29: 175–178.

Ma, Y.Y., Guo, B.L., Wei, Y.M. and Zhao, H.Y., 2014. Research progress on traceability technology of origin of plant-derived food. Food Science 35: 246–250.

Mehari, B., Redi-Abshiro, M., Chandravanshi, B.S., Combrinck, S., McCrindle, R. and Atlabachew, M., 2019. GC-MS profiling of fatty acids in green coffee (Coffea arabica L.) beans and chemometric modeling for tracing geographical origins from Ethiopia. Journal of the Science of Food and Agriculture 99: 3811–3823. https://doi.org/10.1002/jsfa.9603

Michael, P.-R., Gaiad, J.E., Hidalgo, M.J., Avanza, M.V. and Pellerano, R.G., 2019. Classification of cowpea beans using multielemental fingerprinting combined with supervised learning. Food Control 95: 232–241. https://doi.org/10.1016/j.foodcont.2018.08.001

Pérez-Castaño, E., Medina-Rodríguez, S. and Bagur-González, M., 2019. Discrimination and classification of extra virgin olive oil using a chemometric approach based on TMS-4,4'-desmetylsterols GC(FID) fingerprints of edible vegetable oils. Food Chemistry 274: 518–525. https://doi.org/10.1016/j.foodchem.2018.08.128

Sun, X.Q., Mao, Z.X., Fu, H., Huang, D.J. and Li, Q., 2014. Fatty acid characteristics of forage and its influence factors. Pratacultural Science 31: 1774–1780.

Tian, X.J., Long, M., Wang, J., Ma, Z.R., Wei, Z.B., Chen, S.E., Gao, D.D. and Ding, B., 2018. Tracing the origin of wolfberry fruit based on odor information of electronic nose and multivariate statistical analysis. Acta Agriculturae Zhejiangensis 30: 1604–1611.

Wang, F., Zhao, H., Yu, C., Tang, J., Wu, W. and Yang, Q., 2020. Determination of the geographical origin of maize (Zea mays L.) using mineral element fingerprints. Journal of the Science of Food and Agriculture 100: 1294–1300. https://doi.org/10.1002/jsfa.10144

Wang, L.Y., 2012. Analysis of soil pollution in Tieling City. Journal of Environmental Management College of China 22: 60–63.

Wang, Z.H., Zheng, H., Zhao, Q. and Zhang, D.L., 2019. Canonical correspondence analysis on the distribution environment and mineral elements of Liuhe River Rice. Food Science 40: 318–324.

Wu, Y.Y., 1986. Background value of soil environment in Shenyang City. Environmental Protection Science 04: 24–28.

Xiao, R., Ma, Y., Zhang, D. and Qian, L., 2018. Discrimination of conventional and organic rice using untargeted LC-MS-based metabolomics. Journal of Cereal Science 82: 73–81. https://doi.org/10.1016/j.jcs.2018.05.012

Yun, W.L., Hou, Q. and Li, Y.W., 2013. Spatial distribution of soil hydrological characteristics in inner mongolia. Journal of Arid Land Resources and Environment 27: 193–197.

Zhang, J., Yang, R., Chen, R., Li, Y.C., Peng, Y. and Wen, X., 2019a. Geographical origin discrimination of pepper (Capsicum annuum L.) based on multi-elemental concentrations combined with chemometrics. Food Science and Biotechnology 28: 1627–1635. https://doi.org/10.1007/s10068-019-00619-3

Zhang, X., Han, D., Chen, X., Zhao, X., Cheng, J. and Liu, Y., 2019b. Combined use of fatty acid profile and fatty acid δC fingerprinting for origin traceability of scallops (Patinopecten yessoensis, Chlamys farreri, and Argopecten irradians). Food Chemistry 298: 124966. https://doi.org/10.1016/j.foodchem.2019.124966

Zhang, X., Liu, Y., Li, Y. and Zhao, X., 2017. Identification of the geographical origins of sea cucumber (Apostichopus japonicus) in northern China by using stable isotope ratios and fatty acid profiles. Food Chemistry 218: 269–276. https://doi.org/10.1016/j.foodchem.2016.08.083

Zhang, Y., Wang, D. and Li, X., 2018. Research progress on origin tracing of agricultural products based on near infrared spectroscopy. Journal of Food Safety & Quality 9: 6161–6166.

Zhao, Y., Si, W., Tian, G.Q. and He, X.R., 2018. Monitoring report on soybean production and market dynamics (August 2018). Soybean Science & Technology 04: 12–21.

Zhao, Y., Zhao, C., Li, Y., Chang, Y., Zhang, J., Zeng, Z., Lu, X. and Xu, G., 2014. Study of metabolite differences of flue-cured tobacco from different regions using a pseudotargeted gas chromatography with mass spectrometry selected-ion monitoring method. Journal of Separation Science 37: 2177–2184. https://doi.org/10.1002/jssc.201400097