Food Control 98 (2019) 457–464

Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont

Multivariate statistical analysis combined with e-nose and e-tongue assays simplifies the tracing of geographical origins of *Lycium ruthenicum* Murray grown in China

Zi-Chao Wang[a,b,c,1], Yuzhen Yan[a,1], Tanzeela Nisar[c], Lijun Sun[c], Yang Zeng[a], Yurong Guo[c], Huichun Wang[a], Zhongxiang Fang[b,*]

[a] School of Life Science, Qinghai Normal University, Qinghai Province, Xining 810000, China
[b] Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Vic 3010, Australia
[c] College of Food Engineering and Nutritional Science, Shaanxi Normal University, Shaanxi Province, Xi'an 710000, China

ARTICLE INFO

Keywords:
Lycium ruthenicum murray
E-nose
E-tongue
Principal component analysis
Linear discriminant analysis

ABSTRACT

This study aims to develop a fast and simple method to trace the geographical origins, harvest years and varieties of *Lycium ruthenicum* Murray (LRM) grown in China by employing e-nose and e-tongue assays and their combination. Principal component analysis (PCA) and linear discriminant analysis (LDA) were applied for qualitative classification and quantitative prediction. The results showed that e-nose and e-tongue assays and their combination failed to recognize harvest years and varieties of LRM, but achieved reliable results for tracing LRM geographical origins with a total classification ability of 86.4%, 86.8% and 92.6% respectively. In addition, the analysis procedure required shorter time and less chemical reagents as compared to high-end instrumental analysis or traditional methods like chemical analytical methods and sensory evaluation. This study demonstrated that the multivariate statistical analysis combined with e-nose and e-tongue assays could be a reliable and simplified method of tracing the geographical origins of LRM.

1. Introduction

Lycium ruthenicum Murray (LRM) is a unique nutraceutical food that has a variety of biological functionalities like cell-mediated immunity enhancement, anti-oxidation, anti-aging, anti-fatigue and hypoglycemic activities (Lv, Wang, Cheng, Huang, & Wang, 2013). LRM has been used as a traditional herb for the treatment of heart diseases, abnormal menstruation and menopause etc. (Liu et al., 2013). In addition, LRM possess strong drought- and salt-resistance, making it an ideal plant for preventing soil desertification and alleviating soil salinity-alkalinity (Zheng et al., 2011). Consequently, LRM is considered important for the agriculture and ecosystem particularly in China's western provinces where LRM is widely distributed (Wang et al., 2018).

Qinghai Province is one of the major producers of LRM in China due to its suitable climate and geographical conditions (Lv et al., 2013). Qinghai LRM contains higher content of nutritional components and therefore has higher market value as compared to that from other provinces (Zheng et al., 2011). Mislabeling and selling of fake and inferior quality LRM has been increased in the market due to economic considerations (Wang et al., 2018). This severely infringes the reputation of famous high-quality products and damages the interests of consumers, and may also cause serious food safety issues (Zhang, Liu, Li, & Zhao, 2017). A possible solution for this issue is to establish a reliable traceability and identification system (Qiu, Wang, Tang, & Du, 2015).

Previously, high-end instrumental analysis method has been established by our team to trace the LRM geographical origins (Wang et al., 2018). The results suggested that high-performance liquid chromatography-mass spectrometry (HPLC-MS) combined with principal component analysis (PCA) and linear discriminant analysis (LDA) was a powerful analytical method for the traceability of geographical origins of LRM. However, the HPLC-MS facility is too expensive to the remote LRM producing and trading regions. The operation of HPLC-MS system and the interpretation of analytical results are time- and reagents-consuming and require well-trained personnel (Qiu, Wang, Tang, et al., 2015). Therefore, it has become increasingly important to develop a rapid, robust, and simple alternative technique for determining the authenticity of LMR.

* Corresponding author.
E-mail address: zhongxiang.fang@unimelb.edu.au (Z. Fang).
[1] Z.-C. Wang and Y. Yan contributed equally to this article.

https://doi.org/10.1016/j.foodcont.2018.12.012
Received 5 July 2018; Received in revised form 18 November 2018; Accepted 13 December 2018
Available online 14 December 2018
0956-7135/

Electronic nose (e-nose) and electronic tongue (e-tongue) mimic the olfactory and gustatory systems of the mankind and are good alternatives for traditional analysis of foods (Hong, Wang, & Hai, 2012). By monitoring the flavour and taste, they have been successfully applied for classifying the types and predicting the deterioration of foods such as fruit juices, peanuts, and honey etc. (Haddi et al., 2014; Naila, Flint, Sulaiman, Ajit, & Weeds, 2018; Qiu, Wang, & Gao, 2015). The combination of e-nose and e-tongue system have also been suggested to discriminate and classify foods (Ghasemi-Varnamkhasti et al., 2012). In general, algorithms techniques such as PCA and LDA were applied in those researches to mine useful information from the responses of e-nose and e-tongue (Qiu, Wang, Tang, et al., 2015). These approaches have good practical applications because they have the advantages of low cost, minimal sample preparation, non-destructive detection, less time consuming, and simple sampling procedure etc. (Li, Lei, Zhang, Shao, & Xie, 2015; Liu & Tu, 2012; Shen et al., 2018). Therefore, the objective of this study was to develop a fast and simple method to trace the geographical origins, harvest years and varieties of LRM based on e-nose and e-tongue assays and their combination, by using PCA and LDA assays. This would provide an alternative reliable technique for the authenticity recognition of LRM.

2. Materials and methods

2.1. Plant materials and treatments

A total of 243 LRM samples belonging to two different varieties were randomly collected from 5 different provinces in China in 2016 and 2017 (Table 1). The environmental factors of geographical locations of LRM sampling have been given in our previous work (Wang et al., 2018). The fully ripened fruits were hand-picked and transported right after collected in heat preservation box with ice bags to the laboratory, in where they were freeze-dried with a Sihuan® LGJ-18C vacuum freeze drier (Beijing, China), crushed to a fine powder using a Zhaoshen® XS-10 pulverizer (Shanghai, China), and sieved through a 40-mesh sieve. The e-nose and e-tongue measurements were following the method of Tian, Deng, and Chen (2007) by stirring 1.00 g of the powder in 100 mL of deionized water. The mixture was heated at 85 °C for 5 min and then cooled in an ice-bath to room temperature. The residues were filtered and the supernatants were analyzed immediately by the e-nose and e-tongue. Each sample was successively analyzed for three times and average values were used in score plot for further analysis.

2.2. E-nose system and sampling procedure

The headspace analysis was performed with an ISENSO® iNose e-nose (New York, USA). The e-nose system consists of a sampling apparatus, a detector that contains sensor arrays and a computer equipped with pattern-recognition software for data recording and elaboration (Fig. 1A). The sensor array system is composed of 10 metal-oxide semiconductor chemical sensors to detect volatile compounds. The 10 sensor arrays are named as follows: N-S1 (aromatics), N-S2 (nitrogen oxides), N-S3 (ammonia, aromatic molecules), N-S4 (hydrogen), N-S5 (methane, propane, and aliphatic non-polar molecules), N-S6 (broad methane), N-S7 (sulfur-containing organics), N-S8 (broad alcohols), N-S9 (aromatics, sulfur- and chlorine-containing organics) and N-S10 (methane and aliphatics). The details of the e-nose system have been introduced by Zhang, Wang, Tian, Yu, and Yu (2007).

Prior to detection, the gas path of e-nose was cleaned by cleaning gas (ambient air filtered through activated charcoal) for 20 min to normalize the sensor signals. Each sample solution (5.00 mL) was placed in a 50 mL airtight glass vial and sealed with plastic wrap for 10 min. Then, the headspace gaseous compounds were pumped into the sensor arrays through Teflon tube connected to a Luer-lock needle in the plastic wrap at a flow rate of 300 mL/min (Wang et al., 2016). The measurement phase was lasted for 80 s, which was long enough for the sensors to reach stable signal values. The signal data from the sensors were collected by the computer at an interval of 0.1 s. When the measurement process was complete, the acquired data were stored for further analysis. After each measurement, zero gas (air filtered by active carbon) was pumped into the sample gas path from the other port of the instrument for 240 s to normalize sensor signals. All the e-nose measurements were performed at room temperature.

2.3. E-tongue system and sampling procedure

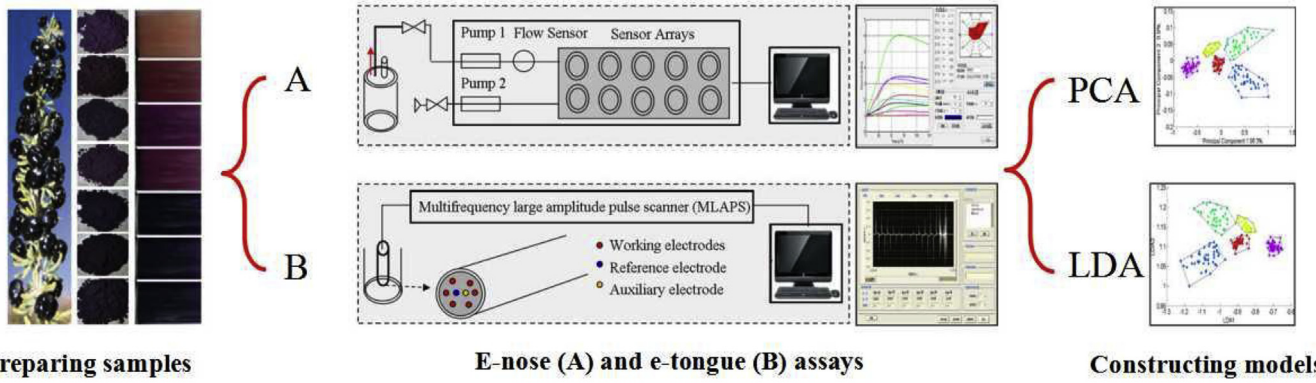
An ISENSO® SmarTongue e-tongue (New York, USA) was employed to classify and characterize LRM samples. The e-tongue system consists of an electronic tongue, a device called “Multi-frequency large amplitude pulse scanner, MLAPS” and a computer (Fig. 1B). It comprises six metallic disc electrodes (T-S1, platinum; T-S2, gold; T-S3, palladium; T-S4, tungsten; T-S5, nickel; T-S6, silver) as working electrodes, a Ag/AgCl electrode as reference electrode and a platinum counter electrode as auxiliary electrode for standard three-electrode systems. MLAPS is a potentiostat with six channels that makes the potential pulses on the working electrodes and enables them to work consecutively at three-electrode configurations. The computer contains data acquisition system and basic data analysis software and was used to set and control the potential pulses, measure and store pulse current responses. The details of the e-tongue system have been described by Tian et al. (2007).

The measurements were directly performed without any sample solution pre-treatments. About 25.00 mL of solution was used to ensure that sensors were fully immersed. Before measurement, sensors were rinsed for 10 s using deionized water to minimize and correct the drift of sensors. The measurement time was 180 s (30 s for each working electrode) for each sample. The principle of the method was to measure the changes in the voltage (mv) intensity between the working electrodes and the reference electrode. The applied potential waveform was multi-frequency large amplitude pulse voltammetry (MLAPV), consisted of three segments of 1 Hz, 10 Hz and 100 Hz, respectively. The

Table 1
The information of LRM samples.[a]

	Province	Detailed origin	Variety	Year	Code	Number of samples collected in each county or farm	Total number of samples
1	Gansu	Guazhou, Yumen, Jinta	Wild	2017	GS-W-17	15	45
2	Inner Mongolia	Ejina, Alashanyou, Alashanzuo	Wild	2017	IM-W-17	15	45
3	Ningxia	Pingluo, Helan, Yongning	Wild	2017	NX-W-17	15	45
4	Qinghai	Gomud, Delingha, Dulan	Wild	2017	QH-W-17	15	45
5	Xinjiang	Akesu, Kuerle, Hetian	Wild	2017	XJ-W-17	15	45
6	Qinghai	Gomud, Delingha, Dulan	Wild	2016	QH-W-16	3	9
7	Qinghai	Hedong farm, Delingha farm, Nomuhong farm	Cultivated	2017	QH-C-17	3	9

[a] Hedong, Delingha and Nomuhong farm located in the counties of Golmud, Delingha and Dulan, respectively.



Procedures	Instruments	Time consumption		Chemical reagents consumption
		Preparing samples	Instrument assay	
The present study	E-nose /e-tongue	5 min	80 s/3 min	Water
The previous study	HPLC-ESI-ToF-MS	8 h	100 min	Cyanidin 3-glucoside, methanol, formic acid, trifluoroacetic acid, acetonitrile etc.

Fig. 1. The procedure containing the set-up of electronic nose (A) and electronic tongue (B) for classifying LRM used in the present study and the comparison between it and the procedure proposed in the previous study (Wang et al., 2018).

collection rate is 1 point every 0.001 s. All the samples were tested at room temperature.

2.4. Data fusion and analysis

According to the abstraction level, there are three approaches for data fusion, i.e., high-abstraction level, mid-abstraction level and low-abstraction level (Callao & Ruisanchez, 2018; Haddi et al., 2014). For the high-level fusion approach, each data source is analyzed to construct a model separately. Then the results from all the models are combined (Liu & Brown, 2004). For the mid-level fusion approach, each data source is feature extracted separately. Then all the extracted features are merged. For the low-level fusion approach, all the data sources are simply concatenated. Then the combined dataset is used to construct a model. The last approach was elaborated by several researchers and satisfactory results were obtained (Haddi et al., 2014; Qiu, Wang, Tang, et al., 2015). Therefore, we adopted the low-level fusion for the combination of data from the e-nose and e-tongue assays in the present study.

2.5. Data analysis

The data obtained from e-nose and e-tongue assays and their combination were subjected to pattern recognition analysis. Rows of the dataset represent the analyzed LRM samples (243 objects) and columns represent the response values of the sensors (10 for e-nose, 6 for e-tongue and 16 for the two instruments combined). Unsupervised (PCA) and supervised (LDA) approaches were employed using chemometric techniques as described by Granato et al. (2018) and Abad-García, Berrueta, Garmón-Lobato, Urkaregi, Gallo, & Vicenteto (2012) to extract the main information in multivariate data and to construct classification models according to geographical origins, harvest years and varieties.

PCA is a most popular unsupervised exploratory technique to reduce the multidimensional data to a lower dimensional approximation and simplify the interpretation of large datasets containing many objects (samples) and variables (responses) (Callao & Ruisanchez, 2018). In this method, PCA interprets the data in a two or three dimensions by the first two or three principal components (PCs). The chosen PCs contain the maximum data variance and thus linear combined to the original response vectors.

LDA is a main discriminant technique to explicitly model the difference between the classes of data, and it provides a classification model showing the classification scores with respect to the descriptors (Callao & Ruisanchez, 2018). In this method, LDA maximizes the variance between categories and minimizes the variance within categories. To ensure the most significant variables involved in the differentiation are selected, a stepwise variable selection procedure is performed using a Wilks' λ and F statistic. The procedure will not check the previously selected variables before the inclusion of new variable (e.g. the early selected variable will be removed if it is no longer useful) until there is no other variables meet the criteria for entry or when the next included variable is the one that was just removed. A cross-validation procedure (leave-one-out test) is used to evaluate the classification performance. Compared with PCA, LDA can notice the distribution of points in the same category and the distance between them (Hong et al., 2012). The reliability of the classification models constructed in this study were evaluated based on recognition and prediction capacity. All data were processed with SPSS19.0 software (SPSS Inc., Chicago, USA).

3. Results and discussion

3.1. E-nose and e-tongue response to LRM

Typical responses of e-nose to LRM are summarized in Fig. 2A–G. The response signal was expressed as G/G_0, where G and G_0 are the conductivities of the sensors when exposed to the sample gas and zero gas, respectively. The response signals of all e-nose sensors reached at dynamic balances after 60 s. Response signals for each sample were increased at different rate. The response signals of N-S2 increased and

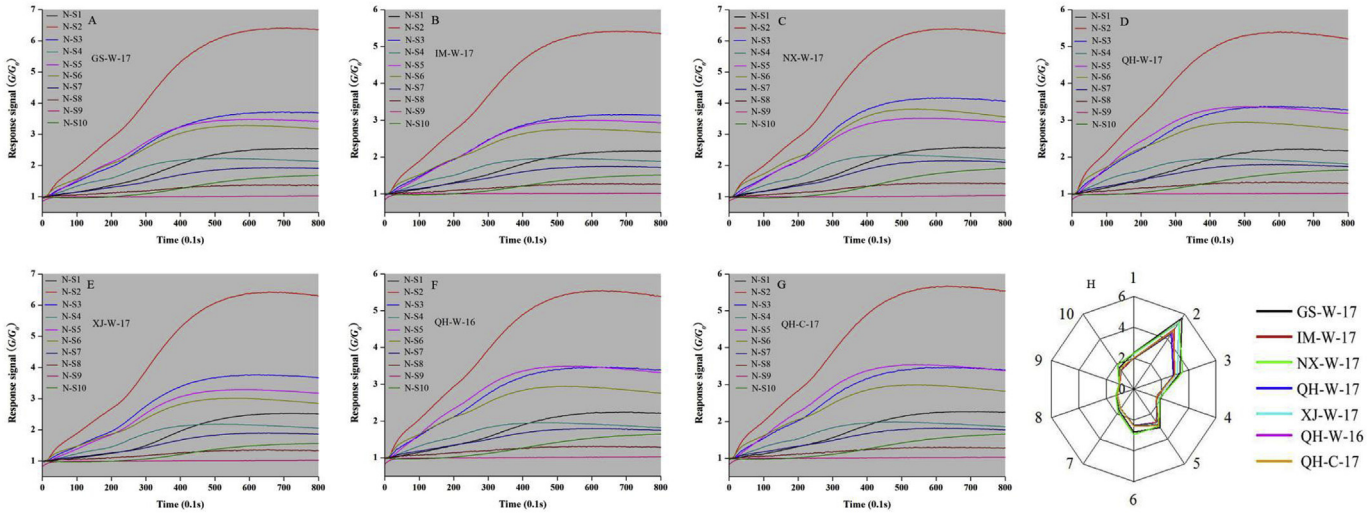


Fig. 2. E-nose responses to LRM samples of GS-W-17 (A), IM-W-17 (B), NX-W-17 (C), QH-W-17 (D), XJ-W-17 (E), QH-W-16 (F) and QH-C-17 (G), and radar chart analysis (H) of response values of ten e-nose sensor responses.

remained at a relatively high level, while the response signals of N-S3, N-S5 and N-S6 were also increased but remained at middle high level; The remaining response signals were increased slightly and remained at low level, especially the response signals of N-S9 (Fig. 2A–G). Overall, the response signals of N-S2, N-S3, N-S5 and N-S6 were remained higher as compared to other response signals (Fig. 2A–G). The concentration and threshold values of volatile compounds should be considered in order to assess their contribution to flavor perception (Tian, Gou, Niu, Sun, & Guo, 2018). The results suggested that nitrogen oxides (N-S2); ammonia, aromatic molecules (N-S3); methane, propane, and aliphatic non-polar molecules (N-S5) and broad methane (N-S6) mainly contributed to the aroma of LRM.

Since the original experimental data were very large, the pulse current profiles of e-tongue to LRM samples were formed using the maximum, minimum and two inflection points of each cycle (Tian et al., 2007) and are summarized in Fig. 3A–G. It should be noted that, to separate different types of samples, different frequency segments and their optimization are needed, because one working electrode has different separation ability in different frequency segments (Tian et al., 2007). The segments of the MLAPV were set at 1 Hz, 10 Hz and 100 Hz, and the optimal combination of T-S1 (1 Hz), T-S2 (1 Hz), T-S3 (100 Hz), T-S4 (1 Hz), T-S5 (10 Hz) and T-S6 (1 Hz) was used to form the original data of e-tongue. Although the results of e-tongue analysis were summarized as pulse currents, it suggested that different matrix may present in food samples (Bougrini et al., 2014). The main compound families that contribute to the taste of foods are minerals, sugars, acids, proteins, lipids and fats (Bougrini et al., 2014; Dias et al., 2009; Winquist et al., 2005). In this study, LRM samples were swelled in the liquid state when they were measured by e-tongue. Therefore, sugars and acids could be the main components in the liquids. It was previously identified by Zhang, Chen, Zhao, and Xi (2016) that LRM juice was the most acidic with a higher amount of titratable acid and had lesser amount of total soluble solid as compared to *Lycium barbarum* L. genotypes, another species of the genus *Lycium*. As total soluble solid generally contains sugars, acids and secondary metabolites present in fruits (Beckles, 2012), acids could be responsible for the specific taste of LRM.

It was detected that both the e-nose and e-tongue responding fingerprints of different LRM samples were same (Fig. 2A–G and 3A-G), suggesting the LRM samples may have similar genetic background (Li et al., 2017). It is not surprising that the cultivated LRMs have same genetic background as those of wild LRMs, because they are actually transplanted from the wild ones. However, significant differences were observed in the strength of response signals and pulse current values of LRM samples from different provinces. The differences could be extracted with each type of sensor by radar plot as depicted in Figs. 2H

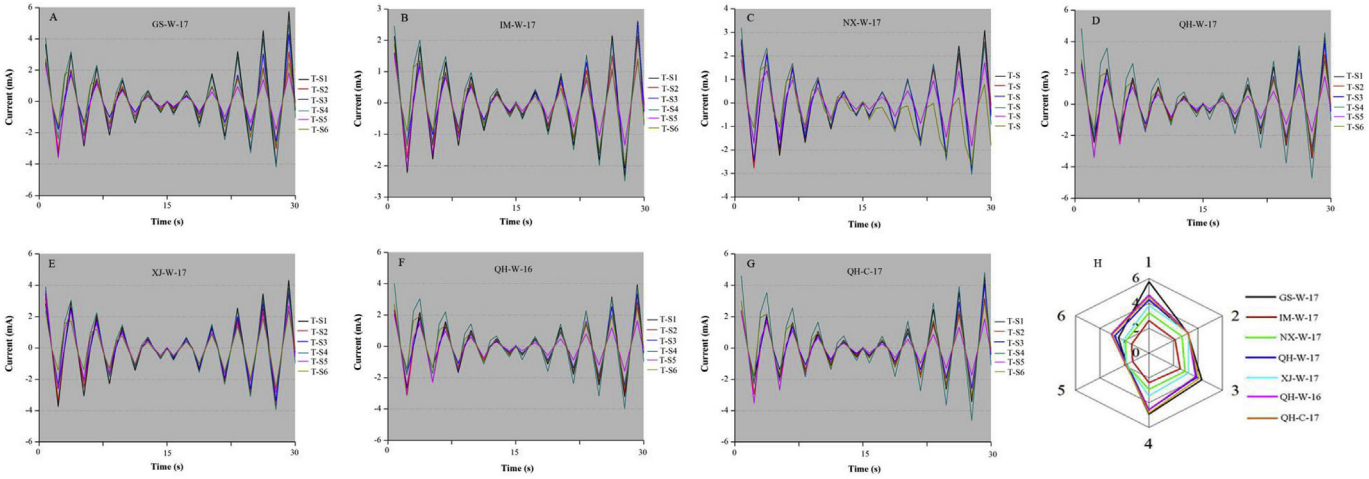


Fig. 3. E-tongue responses to LRM samples of GS-W-17 (A), IM-W-17 (B), NX-W-17 (C), QH-W-17 (D), XJ-W-17 (E), QH-W-16 (F) and QH-C-17 (G), and radar chart analysis (H) of response values of six e-tongue sensor responses.

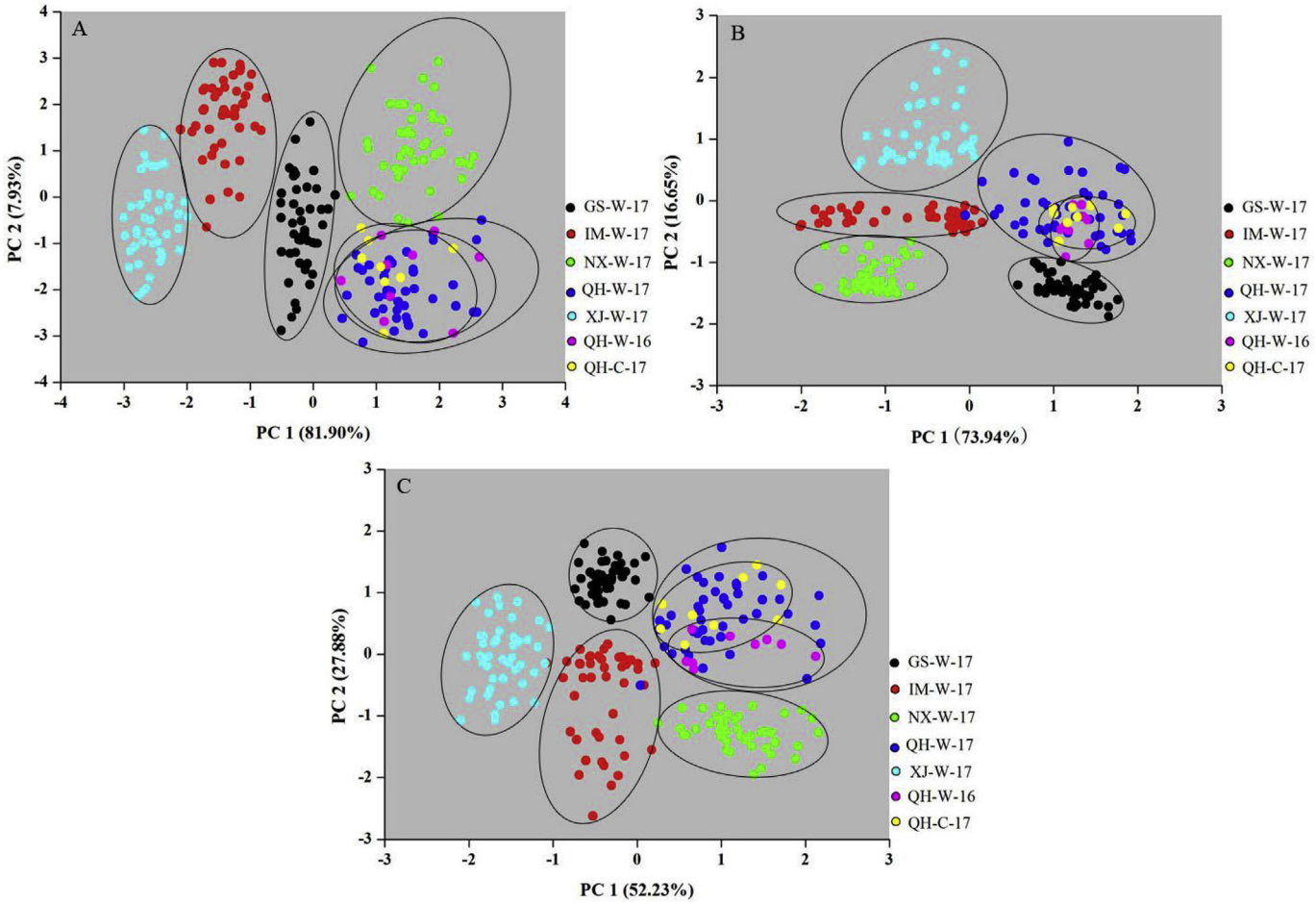


Fig. 4. Two-dimensional PCA plots performed on LRM samples with data gathered using the e-nose (A), e-tongue (B) and fusion system (C).

and 3H. They were formed with the mean values of the response signals of each e-nose sensor during 70–80 s and the maximum values of the pulse currents of each e-tongue sensor, respectively. Although the radar charts of different LRM samples showed a similar shape, the strength of response signals by e-nose and pulse current values by e-tongue were different. The response strengths of N-S2, N-S3, N-S5, N-S6 and N-S10 and the pulse current values of T-S1, T-S3, T-S4 and T-S6 were particularly different among the samples (Figs. 2H and 3H). However, the response strength and the pulse current values of all the samples of Qinghai Province were found same (Figs. 2H and 3H). The radar plots show a clear pattern variation among different LRM samples excepting Qinghai Province samples. Environmental factors like soil components, weather and climatic conditions (rainfall, sunlight, diurnal temperature) are the primary sources of variations among different provinces (Wang et al., 2018; Zheng et al., 2011; Łata, 2007). The rainfall has been reported to be a key factor directly linked to the formation of aromatic and taste components (Russo et al., 2014). Therefore, it is expected that LRM samples from Gansu and Ningxia Provinces have a higher level of aroma components and pulse currents than these from other provinces, because these two provinces have been reported to have more rainfall than other provinces (3–126 and 500-200 mm, respectively) in the sample growing years (Wang et al., 2018). The results showed that the samples of GS-W-17 and NX-W-17 had higher values of N-S2, N-S3, N-S5, N-S6 and N-S10 than that of other provinces (Fig. 2H). Similarly, GS-W-17 exhibited the highest pulse current in the sensors of TS-1, TS-3 and TS-4 (Fig. 3H). It has been reported that environmental variations (rainfall, sunlight and temperature) can trigger year-to-year variations in the concentrations of components in fruits (Łata, 2007). In addition, the component compositions in fruits could also be influenced by agronomic management practices like irrigation, fertilization, herbicide and pesticide treatments etc. (Tomaas-Barberaan & Espin, 2001). However, similar response signal strengths and pulse current values were observed in the present studied samples (Figs. 2 and 3). The possible reasons are that compounds in the fruits are genetically controlled, and no discernible difference was found in the climatic conditions including temperature, air pressure, relative humidity, water vapor pressure, wind, and precipitation etc. between the two years in Qinghai Province (National Meteorological Information Center, Available: https://data.cma.cn/data/detail/dataCode/A.0012.0001.html). The results suggested that LRM obtained from different geographical origins exhibited significant differences in aromatic components and pulse currents, indicating that geographical origin pays an important role in forming the aromatic and taste components of LRM. For this reason, the differences in aromatic components and pulse currents could be revealed by appropriate statistical analyses, and might be credible indices for LRM classification based on geographical origins.

3.2. Classification of LRM based on individual e-nose and e-tongue dataset

In general, not all the data from each working electrode of the e-nose and e-tongue is required for multivariate statistics process, only some features extracted from the response curve of each sensor are needed (Haddi et al., 2014; Tian et al., 2007). Hence, the mean values of the response signals of each e-nose sensor during 70–80 s and the maximum values of the pulse currents of each e-tongue sensor were used as the data matrix for multivariate statistical analysis (Haddi et al., 2014; Tian et al., 2018).

First, PCA was applied to the data matrix of e-nose (10 × 243) and e-tongue (6 × 243). In e-nose analysis, the total contribution variance of PC1 and PC2 is 89.82%, which is sufficient to explain the total variance of the e-nose dataset. From the loadings of the variables, the first principal component (PC1) contributed 81.90% to total variability and contains five major influential features (N-S2, N-S3, N-S5, N-S6, and N-S10). The second principal component (PC2) contributed 7.93% to total variability and contains one major influential feature (N-S9) (Table S1A). PC1 and PC2 were used to defined a two-dimensional spaces for representing the scores of the samples, in which a natural separation between samples on the basis of geographical origins was observed (Fig. 4A). It shows that LRM samples obtained from different geographical origins were clustered into five groups including those from Qinghai, Gansu, Inner Mongolia, Xinjiang and Ningxia. However, LRM samples of QH-W-16 and QH-C-17 were almost located in the same cluster of QH-W-17 (Fig. 4A). In e-tongue analysis, the total contribution variance of the first two PCs was higher than 90.0%, indicating that they were the main source of total system variance. PC1 explained 73.94% of the variance, and T-N1, T-N3, T-N4 and T-N6 predominantly contributed to it. PC2 yielded 16.65% of explainable results, with T-N5 loaded heavily (Table S1B). Discrimination of LRM samples according to geographical origins was completed by two-dimensional score plot defined by PC1 and PC2 (Fig. 4B). However, one data point from the QH-W-17 group was connected to the IM-W17 group, and one data point from the GS-W-17 group was very close to the QH-W-16 group. These data points may be misclassified into its neighboring groups in the above classification analysis. In addition, there was a certain overlapped area with no clear differentiation was observed between LRM samples from Qinghai Province (Fig. 4B). The clustering results (Fig. 4A and B) suggested that PCA could be used to classify LRM samples from each other on the basis of geographical origins. However, the samples within the same province having different varieties and harvest years were unable to be separated satisfactorily.

LDA was then applied to the same data matrix to construct discriminant model for distinguishing and identifying LRM samples. The leave-one-out test was applied to validate the discriminant functions. Six discriminant functions were obtained for e-nose dataset, which resulted in 100.0% total recognition ability (Table S2A) and 86.4% prediction ability (Table S3A). LRM samples of GS-W-17, IM-W-17, NX-W-17 and XJ-W-17 were clearly distinguished from each other, and from the samples of Qinghai Province. The prediction correct rate for LRM samples of GS-W-17, IM-W-17, NX-W-17 and XJ-W-17 were 97.8%, 100.0%, 100.0% and 100.0%, respectively (Table S3A), which can be considered satisfactorily. However, as shown in Table S3A, the prediction correct rate for samples of Qinghai Province was not satisfactory. The sum of the first two discriminant functions covered 92.5% of the total variance explained. The first function accounted for 67.9% and the second one accounted for 24.6% of the total variability (Table S2A). The scores of the first two functions were plotted as a two-dimensional scatter diagram (Fig. 5A). Meanwhile, five discriminant functions were obtained for e-tongue dataset. The model shows a highly satisfactory classification performance allowing to correctly classify 76.7% of the samples (Table S2B), and 86.8% for the cross-validation procedure (Table S3B). The first two discriminant functions explained 92.3% of the total variance of the e-tongue data (the first explaining 68.7% and the second 23.6%) (Table S2B). The graphical representation of LRM samples in the plane defined by the first two discriminant functions is presented in Fig. 5B. The samples of GS-W-17 and XJ-W-17 were clearly distinguished from each other and from other samples, with prediction abilities of 95.6% and 100.0% respectively (Table S3B). Although the samples of IM-W-17 and NX-W-17 were not completely separated, they exhibited a considerably satisfactory prediction abilities of 75.6% and 80.0%, respectively (Table S3B). The samples of QH-W-17 and QH-W-16 were completely overlapped. The data points of QH-W-17 samples were not observed, which may be misrecognized as the samples of QH-W-16 (Fig. 5B). Fig. 5A and B clearly confirmed that the LDA achieved an unambiguous classification of LRM samples based on geographical origin by showing a sufficient division of LRM samples into different clusters. However, the samples within the same province were not clearly distinguished from each other.

The results of PCA and LDA suggested that both the aroma components and pulse currents could be considered as indicators to determine LRM geographical origins, thus providing an effective method to trace geographical origins, but failed to differentiate the samples on the basis of harvest years and varieties. Moreover, geographical origin-based classifications applied on e-nose dataset revealed that the major aroma compounds such as N-S2, N-S3, N-S5 and N-S6 exhibited a great impact on geographical differentiation. Meanwhile, the minor compound such as N-S9 also significantly influenced the geographical origin-based classification. The remaining aromas in LRM samples were less important for the classification. A similar result was obtained in the analysis of the geographical origin-based differentiation applied on e-tongue dataset. The results suggested that not only the major components provided by the e-nose and e-tongue, but also the minor ones can properly characterize LRM samples on the basis of geographical origins. This is in consistent with the previous studies, which suggested that both major and minor components in fruits can be considered as representative indices for reliable differentiation of them from different geographical origins (Guo, Yuan, Dou, & Yue, 2017; Guo, Yue, Yuan, & Wang, 2013; Wang et al., 2018). The results suggested that both e-nose and e-tongue analysis were able to separate the LRM samples according to geographical origins. However, they could not be used for variety- and harvest year-based classifications of LRM.

3.3. Classification of LRM based on combined e-nose and e-tongue dataset

Data fusion has been applied in various foods and quality control processes (Callao & Ruisanchez, 2018). Particularly, low-level abstraction data fusion was proved a good approach to surpass the lack of recognition, which has been successfully applied to separate the overlapped samples (Haddi et al., 2014; Qiu, Wang, & Gao, 2015). Hence, PCA and LDA were conducted on the combined dataset of e-nose and e-tongue to improve the geographically based classification of LRM samples and to discriminate LRM samples within the same province (Qinghai). The e-nose and e-tongue data were combined and the merged matrix was composed of 16 variables. Before classification modeling, PCA was carried out for exploratory data analysis. The total contribution variance of first two PCs obtained from the data fusion was 80.1% that is generally sufficient to explain the total variance of the dataset. As shown in Table S1C, the response signal of N-S2, N-S3, N-S5 and N-S6 have the highest weight in the first PC (explaining 52.23% of the variability). The second PC that explained 27.88% of the total variance was related to T-S1, T-S3, T-S4 and T-S6. Fig. 4C shows the two-dimensional scatter plot of the scores of PC1 versus PC2. It shows that LRM samples obtained from different geographical regions were clustered into five groups. However, a total of two data points from the IM-W-17 and QH-W-17 groups was connected to the XJ-W-17 and IM-W-17 group, respectively, and one data point from the NX-W-17 group was close to the XJ-W-17 group. Besides, the LRM samples from Qinghai Province were not clearly differentiated, giving a completely overlapped distribution. LDA was also applied to the above data matrix. Six discriminant functions were constructed. A satisfactory differentiation on the basis of the five provinces was achieved with a total recognition ability of 100% (Table S2B) and a prediction ability of 92.6% (Table S2C). The first two functions explained the 82.3% of the variance (58.2% and 24.1%, respectively) (Table S2C). The separation of LRM samples was checked by plotting the two functions scores (Fig. 5C), showing that the LRM samples were well distinguished from each other based on geographical origins. The prediction correct rate for both were 100.0% (Table S3C). However, the samples within Qinghai Province were not clearly distinguished from each other (Table

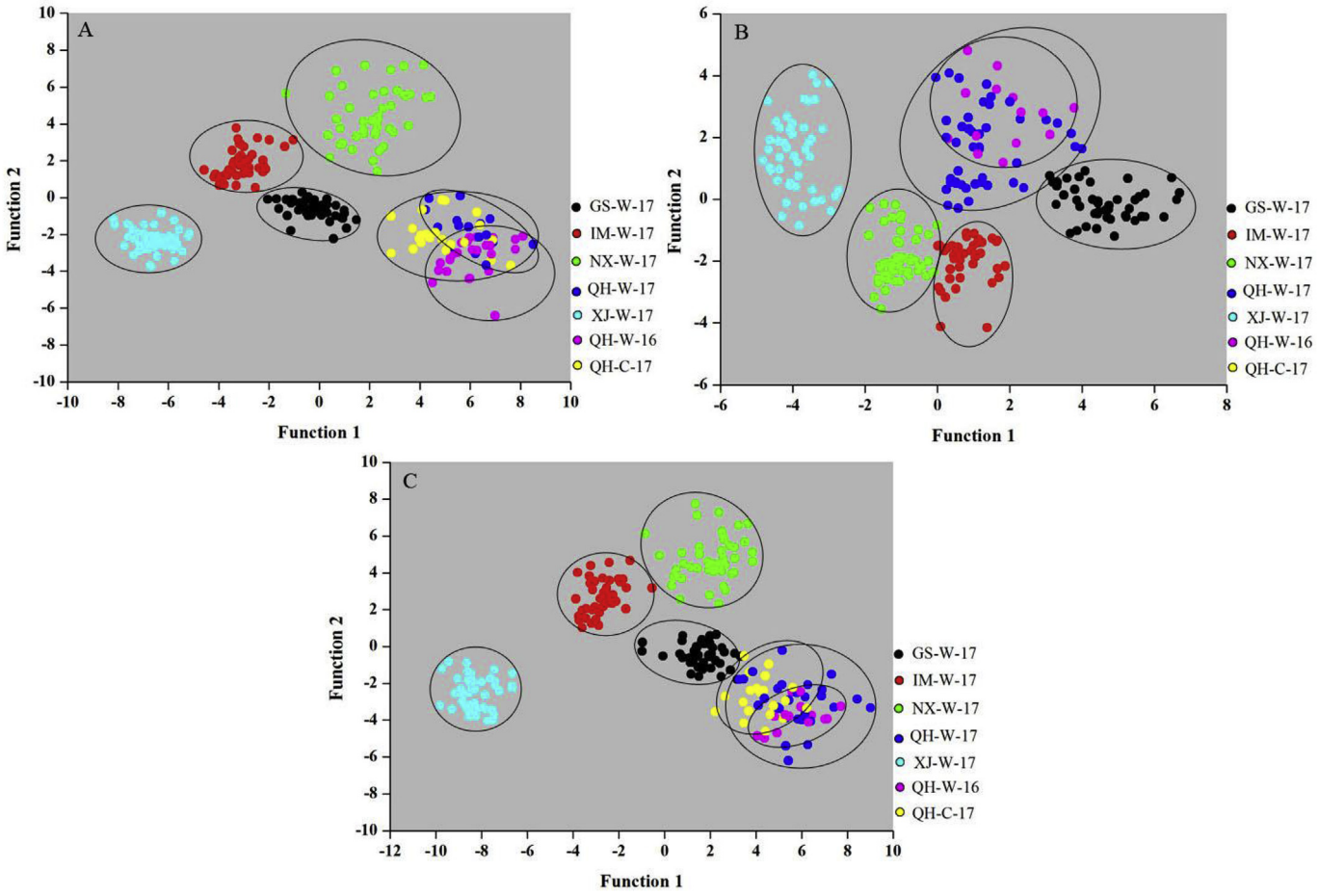


Fig. 5. Two-dimensional LDA plots performed on LRM samples with data gathered using the e-nose (A), e-tongue (B) and fusion system (C).

S3C).

After merging data, although the total contribution variance of the first two PCs (80.1%) was lower than that of individual e-nose (89.82%) and e-tongue (90.59%) datasets (Fig. 4), it was generally sufficient for explaining the total variance of the combined dataset. The potential reason is that, the merged measurement of two data sources provided more redundant information which directly influenced the ability of identification (Haddi et al., 2014). The total classification ability (92.6%) was higher than that of individual e-nose (86.4%) and e-tongue (86.8%) datasets (Table 3S). The higher total classification ability was resulted from the high cross-sensitivity of the combined system (Haddi et al., 2014). The results suggested that fewer extraction gives higher rate of anticipation. Besides, principal response signals of the individual e-nose and e-tongue assays contributed heavily to the total variance, indicating that data fusion have no negative effect on the stability of sensor responses (Sadrieh et al., 2005) and can be used to evaluate the distribution of LRM samples. In summary, the results indicated that data fusion of both instruments from perceptual knowledge could be a better way than individual utilization of e-nose or e-tongue. However, both the PCA and LDA did not show a sufficient division of LRM samples on the basis of harvest years and varieties, even though a clear separation of LRM samples on the basis of geographical origin was obtained (Figs. 4C and 5C). Specific frequency segments for working electrodes of e-tongue may be required to be develop in the discrimination of various samples (Tian et al., 2007). The research will be presented in further study.

In terms of efficiency, fast learning, and high accuracy rate, multivariate statistical analysis combined with e-nose and e-tongue systems has great advantages over traditional detection methods such as chemical analytical methods and sensory evaluation (Qiu, Wang, Tang, et al., 2015). Moreover, as a kind of instrumental analyses, it requires less time and chemical reagent consumption in comparison to the high-end instrumental analysis proposed in our previous study (Wang et al., 2018) (Fig. 1). Furthermore, with the development of technology, more advanced portable e-nose with a novel headspace diffusion sampling method is developed, making the measurement cheaper and more convenient (Li et al., 2015). Hence, the procedure proposed in this study could be used as a fast and simple method. In addition, previously we suggested to establish a classification methodology for LRM with a broader array of samples (Wang et al., 2018). This study expanded the LRM samples from 9 to 45 for each province (Table 1). LRM samples from different harvest years and varieties were also considered at the present study. Hence, this geographical origin-based classification methodology of LRM have potential practical application features.

4. Conclusion

In this study, e-nose and e-tongue systems and their combination were applied to distinguish LRM samples from different provinces, harvest years and varieties. The selected features of the two instruments were used as the input variables of the multivariate statistical analysis. Irrespective of using the two instruments separately or in combination, the results confirmed that multivariate statistical analysis combined with e-nose and e-tongue assays achieved an unambiguous classification between LRM samples from different geographical origins, but failed in distinguishing samples from different harvest years and varieties. The data of LDA were in agreement with PCA and showed satisfactory discrimination for LRM samples based on geographical

origins. Furthermore, it was demonstrated that the data fusion using low-level of abstraction approach was a good approach to improve the recognition. In conclusion, multivariate statistical analysis combined with e-nose and e-tongue assays represents a fast and reliable pattern screening method and could be an appropriate method for original traceability of LRM. Further work is currently in progress to develop an effective and reliable technique to characterize LRM with different harvest years and varieties.

Acknowledgments

This work was supported by China Agriculture Research System (CARS-28), National Natural Science Foundation of China (21557001 & 81460652), Qinghai Province Research Project (2017-ZJ-774), Research Fund of Qinghai Normal University, and Study Abroad Research Program at the University of Melbourne, Australia.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodcont.2018.12.012.

References

Abad-García, B., Berrueta, L. A., Garmón-Lobato, S., Urkaregi, A., Gallo, B., & Vicente, F. (2012). Chemometric characterization of fruit juices from Spanish cultivars according to their phenolic compound contents: I. Citrus fruits. *Journal of Agricultural and Food Chemistry, 60*, 3635–3644.

Beckles, D. M. (2012). Factors affecting the postharvest soluble solids and sugar content of tomato (*Solanum lycopersicum* L.) fruit. *Postharvest Biology and Technology, 63*, 129–140.

Bougrini, M., Tahri, K., Haddi, Z., El Bari, N., Llobet, E., Jaffrezic-Renault, N., et al. (2014). Aging time and brand determination of pasteurized milk using a multisensor e-nose combined with a voltammetric e-tongue. *Materials Science and Engineering: C, 45*, 348–358.

Callao, M. P., & Ruisanchez, I. (2018). An overview of multivariate qualitative methods for food fraud detection. *Food Control, 86*, 283–293.

Dias, L. A., Peres, A. M., Veloso, A. C. A., Reis, F. S., Vilas-Boas, M., & Machado, A. A. S. C. (2009). An electronic tongue taste evaluation: Identification of goat milk adulteration with bovine milk. *Sensors and Actuators B: Chemical, 136*, 209–217.

Ghasemi-Varnamkhasti, M., Rodríguez-Méndez, M. L., Mohtasebi, S. S., Apetrei, C., Lozano, J., Ahmadi, H., et al. (2012). Monitoring the aging of beers using a bioelectronic tongue. *Food Control, 25*, 216–224.

Granato, D., Putnik, P., Kovacevic, D. B., Santos, J. S., Calado, V., Rocha, R. S., et al. (2018). Trends in chemometrics: Food authentication, microbiology, and effects of processing. *Comprehensive Reviews in Food Science and Food Safety, 17*, 663–677.

Guo, J., Yuan, Y., Dou, P., & Yue, T. (2017). Multivariate statistical analysis of the polyphenolic constituents in kiwifruit juices to trace fruit varieties and geographical origins. *Food Chemistry, 232*, 552–559.

Guo, J., Yue, T., Yuan, Y., & Wang, Y. (2013). Chemometric classification of apple juices according to variety and geographical origin based on polyphenolic profiles. *Journal of Agricultural and Food Chemistry, 61*, 6949–6963.

Haddi, Z., Mabrouk, S., Bougrini, M., Tahri, K., Sghaier, K., Barhoumi, H., et al. (2014). E-Nose and e-Tongue combination for improved recognition of fruit juice samples. *Food Chemistry, 150*, 246–253.

Hong, X., Wang, J., & Hai, Z. (2012). Discrimination and prediction of multiple beef freshness indexes based on electronic nose. *Sensors and Actuators B: Chemical, 161*, 381–389.

Li, D., Lei, T., Zhang, S., Shao, X., & Xie, C. (2015). A novel headspace integrated E-nose and its application in discrimination of Chinese medical herbs. *Sensors and Actuators B: Chemical, 221*, 556–563.

Li, D., Li, B., Ma, Y., Sun, X., Lin, Y., & Meng, X. (2017). Polyphenols, anthocyanins, and flavonoids contents and the antioxidant capacity of various cultivars of highbush and half-high blueberries. *Journal of Food Composition and Analysis, 62*, 84–93.

Liu, Y., & Brown, S. D. (2004). Wavelet multiscale regression from the perspective of data fusion: New conceptual approaches. *Analytical and Bioanalytical Chemistry, 380*, 445–452.

Liu, Z., Dang, J., Wang, Q., Yu, M., Jiang, L., Mei, L., et al. (2013). Optimization of polysaccharides from *Lycium ruthenicum* fruit using RSM and its anti-oxidant activity. *International Journal of Biological Macromolecules, 61*, 127–134.

Liu, P., & Tu, K. (2012). Prediction of TVB-N content in eggs based on electronic nose. *Food Control, 23*, 177–183.

Lv, X., Wang, C., Cheng, Y., Huang, L., & Wang, Z. (2013). Isolation and structural characterization of a polysaccharide LRP4-A from *Lycium ruthenicum* Murr. *Carbohydrate Research, 365*, 20–25.

Naila, A., Flint, S. H., Sulaiman, A. Z., Ajit, A., & Weeds, Z. (2018). Classical and novel approaches to the analysis of honey and detection of adulterants. *Food Control, 90*, 152–165.

Qiu, S., Wang, J., & Gao, L. (2015a). Qualification and quantisation of processed strawberry juice based on electronic nose and tongue. *Lebensmittel-Wissenschaft und -Technologie- Food Science and Technology, 60*, 115–123.

Qiu, S., Wang, J., Tang, C., & Du, D. (2015b). Comparison of ELM, RF, and SVM on E-nose and E-tongue to trace the quality status of Mandarin (*Citrus unshiu* Marc.). *Journal of Food Engineering, 166*, 193–203.

Russo, M., Serra, D., Suraci, F., Di Sanzo, R., Fuda, S., & Postorino, S. (2014). The potential of e-nose aroma profiling for identifying the geographical origin of licorice (*Glycyrrhiza glabra* L.) roots. *Food Chemistry, 165*, 467–474.

Sadrieh, N., Brower, J., Yu, L., Dou, W., Straughn, A., Machado, S., et al. (2005). Stability, dose uniformity, and palatability of three counterterrorism drugsdhuman subject and electronic tongue studies. *Pharmaceutical Research, 22*, 1747–1756.

Shen, F., Wu, Q., Liu, P., Jiang, X., Fang, Y., & Cao, C. (2018). Detection of *Aspergillus* spp. contamination levels in peanuts by near infrared spectroscopy and electronic nose. *Food Control, 93*, 1–8.

Tian, S., Deng, S., & Chen, Z. (2007). Multifrequency large amplitude pulse voltammetry: A novel electrochemical method for electronic tongue. *Sensors and Actuators B: Chemical, 123*, 1049–1056.

Tian, Y., Gou, X., Niu, P., Sun, L., & Guo, Y. (2018). Multivariate data analysis of the physicochemical and phenolic properties of not from concentrate apple juices to explore the alternative cultivars in juice production. *Food Analytical Methods, 11*, 1735–1747.

Tomaas-Barberaan, F. A., & Espin, J. C. (2001). Phenolic compounds and related enzymes as determinants of quality in fruits and vegetables. *Journal of the Science of Food and Agriculture, 81*, 853–876.

Wang, H., Hu, Z., Long, F., Guo, C., Yuan, Y., & Yue, T. (2016). Early detection of Zygosaccharomyces rouxii–spawned spoilage in apple juice by electronic nose combined with chemometrics. *International Journal of Food Microbiology, 217*, 68–78.

Wang, Z., Yan, Y., Nisar, T., Zou, L., Yang, X., Niu, P., et al. (2018). Comparison and multivariate statistical analysis of anthocyanin composition in *Lycium ruthenicum* Murray from different regions to trace geographical origins: The case of China. *Food Chemistry, 246*, 233–241.

Winquist, F., Bjorklund, R., Krantz-Rülcker, C., Lundstrom, I., Ostergren, K., & Skoglund, T. (2005). An electronic tongue in the dairy industry. *Sensors and Actuators B: Chemical, 299*, 111–112.

Zhang, Q., Chen, W., Zhao, J., & Xi, W. (2016). Functional constituents and antioxidant activities of eight Chinese native goji genotypes. *Food Chemistry, 200*, 230–236.

Zhang, X., Liu, Y., Li, Y., & 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.

Zhang, H., Wang, J., Tian, X., Yu, H., & Yu, Y. (2007). Optimization of sensor array and detection of stored duration of wheat by electronic nose. *Journal of Food Engineering, 82*, 403–408.

Zheng, J., Ding, C., Wang, L., Li, G., Shi, J., Li, H., et al. (2011). Anthocyanins composition and antioxidant activity of wild *Lycium ruthenicum* Murr. from Qinghai-Tibet Plateau. *Food Chemistry, 126*, 859–865.

Łata, B. (2007). Relationship between apple peel and the whole fruit antioxidant content: Year and cultivar variation. *Journal of Agricultural and Food Chemistry, 55*, 663–671.