

## ORIGINAL RESEARCH

# Discriminating geographical origins of green tea based on amino acid, polyphenol, and caffeine content through high-performance liquid chromatography: Taking Lu'an guapian tea as an example

Huan Su<sup>1</sup>  | Weiquan Wu<sup>2</sup> | Xiaochun Wan<sup>1</sup> | Jingming Ning<sup>1</sup>

<sup>1</sup>State Key Laboratory of Tea Plant Biology and Utilization, Anhui Agricultural University, Hefei, China

<sup>2</sup>Anhui Lu'an Guapian Tea Industry Co., Ltd, Lu'an, China

**Correspondence**

Jingming Ning, Anhui Lu'an Guapian tea industry co., LTD, Longhe Blvd West, Lu'an, China.  
Email: ningjm1998009@163.com

**Funding information**

Modern Agriculture (tea) Special System of China, Grant/Award Number: CARS-19; the National Key Research and Development Plan, Grant/Award Number: 2017YFD040800

**Abstract**

Seventy-three Lu'an guapian tea (LAGP) samples were collected from 11 growing locations in the city of Lu'an, Anhui Province, China. Through high-performance liquid chromatography, 18 amino acids, along with gallic acid, caffeine, and five catechins, were quantitatively detected. Hierarchical cluster, correlation and principal component analysis, and a support vector machine were used for geographical discrimination. The findings suggested that the differences in tea quality between the inner and outer mountain regions are related to isoleucine, leucine, phenylalanine, and valine contents, with a correlation coefficient of more than 0.85. Principal component analysis combining with support vector machine was a feasible method. The identification rates for the inner and outer mountains were 97.96% in the training set and 95.83% in the prediction set. Furthermore, the identification rates for the three counties were 91.84% and 95.83% in the training and prediction sets, respectively.

**KEYWORDS**

factor selection, geographical origin discrimination, high-performance liquid chromatography, Lu'an guapian tea, pattern recognition

## 1 | INTRODUCTION

Green tea, the most widely consumed tea in China, can benefit human health because of its chemical composition, which includes catechins, theanine, and caffeine (Cooper, 2012; Dulloo, Seydoux, Girardier, Chantre, & Vandermander, 2000; Sinija & Mishra, 2008). The chemical composition of tea depends on numerous factors, such as planting area, tea plantation management, processing practices, tea variety, and climate. However, the planting area plays the most important, direct role in the characterization and differentiation of tea quality and hence their market prices. Therefore, to evaluate tea quality, effective identification of geographical origin is crucial.

Lu'an guapian tea (LAGP), produced in Lu'an, Anhui Province, and the top 10 most famous teas in China (Bai et al., 2017), is the only tea in which tender leaves are processed without any buds or stems (Heiss, 2001). Due to the special process involved in its production, LAGP has a chestnut or even orchid-like aroma as well as a taste stronger than normal green tea. The area used for its cultivation has greatly expanded in the past few years.

Lu'an guapian tea development has, however, resulted in several problems, mostly regarding the lack of standards and increase in market irregularities (Zhan, Liu, & Yu, 2017). Differences in the prices of LAGP from different origins can be extreme, which could lead to the market confusion. In 2008, LAGP was awarded geographical

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Food Science & Nutrition* published by Wiley Periodicals, Inc.

indication by the General Administration of Quality Supervision, Inspection, and Quarantine of the People's Republic of China with the region of origin being the border areas of Jinzhai County, Yu'an District, and Huoshan County (administrative levels of district and county are equal, but with different features in city establishment in China), covering the entire city of Lu'an. However, the core region of LAGP only existed in the inner mountain not the outer. The inner mountain region covers Qishan Village, Xiang Hongdian Reservoir, Flowers Hill, and Gongdian Village in Jinzhai County; Dushan Town, Shuangfeng, Shi podian Town, and Shajia Cove in Yu'an District; and Zhufo Monastery Town in Huoshan County. The main outer mountain region includes Shi Banchong in Yu'an District's Lion Village (Fan, Xiaopei, Wang, & Han, 2017). At present, most technique for tea identification (quality and variety) relies on sensory evaluation of color, aroma, taste, and appearance by the trained tea specialists. The obvious disadvantage about the sensory evaluation could be easily influenced by the taster's experience or environment. Teas from the same geographical origin and with the same processing method probably have a similar or typical composition. And such kind of chemical composition could afford their distinctive characteristics and also enable LAGP to be discriminated according to the geographical origins.

To determine the differences in chemical compositions, chemical analysis (such as spectroscopy and chromatography) has been widely applied to tracing the tea origins. Chemical components have been identified using proton nuclear magnetic resonance and near-infrared spectroscopy (Li, Jin, Sun, Ye, & Liu, 2019; Meng et al., 2017), gas chromatography–mass spectrometry (Luo, Chen, Gao, Liu, & Wu, 2017), electronic tongue (Huo et al., 2014), and electronic nose (Chen, Liu, Zhao, & Ouyang, 2013). High-performance liquid chromatography (HPLC) could classify six tea categories (Ning et al., 2017) and green tea from China and Korea (Jang et al., 2008). Combined with discriminate analysis, HPLC has also been applied for authentic origins discrimination of many agricultural products, such as ginger (Yudthavorasit, Wongravee, & Leepipatpiboon, 2014), wine (Rastija & Srećnik, 2009), honey (Hermosín, Chicón, & Dolores Cabezudo, 2003), and olive oil (Cunha, Amaral, Fernandes, & Oliveira, 2006). Compared with secondary metabolites (e.g., catechins), the chemical properties of amino acids are more stable in green tea (Ananingsih, Sharma, & Zhou, 2013). Some relationship between the quality of green tea and amino acid content has been demonstrated (Ding, Yu, & Mou, 2002).

Total free amino acid, catechin, and caffeine contents account for approximately 20% of the dry weight of green tea. More accurate and available information can be provided by increasing the number of analyzed samples and combining this information with other analytical data and statistical treatment of the results (Hermosín et al., 2003). This study aimed to fill the knowledge gap regarding the chemical composition of LAGP. Besides, the results were examined to establish whether the chemical components of tea can discriminate its geographical origin. The following amino acids and catechins are referred to in this paper: alanine (Ala), arginine (Arg), aspartic acid (Asp), cysteine (Cys), glutamic acid (Glu), glycine (Gly), histidine (His), isoleucine (Ile), leucine (Leu), methionine (Met), lysine (Lys), phenylalanine (Phe), proline (Pro), serine (Ser), threonine (Thr),

tyrosine (Tyr), valine (Val), theanine (Thea), gallic acid (GA), (–)-epigallocatechin (EGC), (–)-epigallocatechin gallate (EGCG), (+)-catechin (C), (–)-epicatechin (EC), and (–)-epicatechin gallate (ECG).

## 2 | MATERIALS AND METHODS

### 2.1 | Samples and chemicals

In total, 73 LAGP samples were collected from 11 village points, two counties, one district, and the inner and outer mountain regions (Figure 1); the latitude and longitude of the map center is 116.12°N, 31.51°E (Google Maps). All tea samples were processed by tea farmer using the first to third leaf of the population species according to the techniques outlined in the Anhui provincial standard for LAGP DB34/T237-2017, and the author took samples once per day from 18 April 2017 to 24 April 2017. The name, altitude, and sample size of the sampling point are shown in Table S1. The main difference between inner mountain tea and outer mountain tea is the altitude of the plantation. The samples were kept in aluminum foil bags, stored in a freezer at 4°C, and ground into powder before being strained using a 600- $\mu$ m sieve for subsequent analysis.

High-performance liquid chromatography-grade acetonitrile was obtained from Tedia (Tedia Co.); GA, EGC, EGCG, C, EC, ECG (>99.99%), caffeine, and Thea were purchased from Sigma. A reagent kit was obtained from Waters that included a mixture of 17 amino acid hydrolysate standards (Waters AccQ-Tag Amino Acid Standard H), the derivatization reagent AQC (Waters AccQ-Fluor Reagent), and eluent A (Waters AccQ-Tag Eluent A, concentrate). Milli-Q-treated water was prepared to purity higher than 18 M $\Omega$  cm using a water purification system (Aquapro International Co.).

### 2.2 | Test solution preparation

#### 2.2.1 | Amino acids

According to the ISO/WD 19563 (2017), 1.00  $\pm$  0.01 g of sample powder in a 200-ml beaker and 100 ml of boiling water were added. The sample was brewed on a magnetic stirrer (500 rpm) for 5 min, filtered, and made up to a certain volume. Approximately, 1 ml of the sample solution was centrifuged at 16,000 g(rcf) for 10 min. And 10  $\mu$ l of the extract was reacted with 90  $\mu$ l derivatization reagent under 55°C to perform the precolumn derivatization (Reverter, Lundh, & Lindberg, 1997).

#### 2.2.2 | Polyphenols and caffeine

According to the ISO 14502-2: 2005 (2006), 0.20  $\pm$  0.01 g of powder was eluted in magnetic stirring apparatus (70°C) with 5 ml of methanol–water (70%) for 10 min. Then, it was centrifuged at 4,200 g(rcf) for 10 min and transferred into a 10-ml volumetric flask. Repeat extraction again and mix the two extracts to 10 ml. The extract was filtered through a 0.22- $\mu$ m Millipore membrane before injection.

### 2.3 | Chromatography conditions

Quantitative analysis of amino acids was performed on a Waters e2695 series HPLC system equipped with a Waters 2475 fluorescence detector ( $\lambda_{ex} = 250$  nm,  $\lambda_{em} = 395$  nm) and Waters AccQ-Tag reversed-phase HPLC column (4  $\mu$ m, 3.9  $\times$  150 mm, 37°C); Empower (version 3.0) was used for reporting and processing chromatographic information. The mobile phase comprised 1:11 (v/v) AccQ-Tag Eluent A and 10:11 (v/v) Milli-Q water (A), acetonitrile (B), and Milli-Q water (C). The gradient conditions of the mobile phase were as follows (1.0 ml/min): 17 min, 100% A turned to 91% A, 5% B; 24 min, 80% A, 17% B; 32–34 min, 68% A, 20% B; 35 min, 0% A, 40% B; 37 min, 0% A, 60% B; 38 min, returned to 100% A and maintained for 7 min to clear and stabilize the column. The peak retention times were displayed in the software (Figure 2a).

Polyphenols and caffeine were detected on a Waters 2489 Ultraviolet-Visible Detector ( $\lambda = 278$  nm) and the analytical column (Phenomenex C<sub>18</sub>, 5  $\mu$ m, 250  $\times$  4.6 mm, 35°C). The mobile phase consisted of solvents A (9% acetonitrile, 2% acetic acid, 0.2% EDTA, water) and B (80% acetonitrile, 2% acetic acid, 0.2% EDTA, water), followed by ISO 14502-2: 2005 (2006) (ISO). The gradient conditions of the mobile phase were as follows (1.0 ml/min): 0–10 min, 100% A; 25–35 min, 68% A; 40–45 min, 45% A; 50–60 min, 100% A. The peak retention times and areas are shown in the software (Figure 2b).

### 2.4 | Statistical analysis

In this study, hierarchical cluster, correlation and principal component analysis were used for factors selecting. Hierarchical cluster analysis (HCA), an unsupervised learning process, divides similar objects into groups or more subsets through static classification, and members

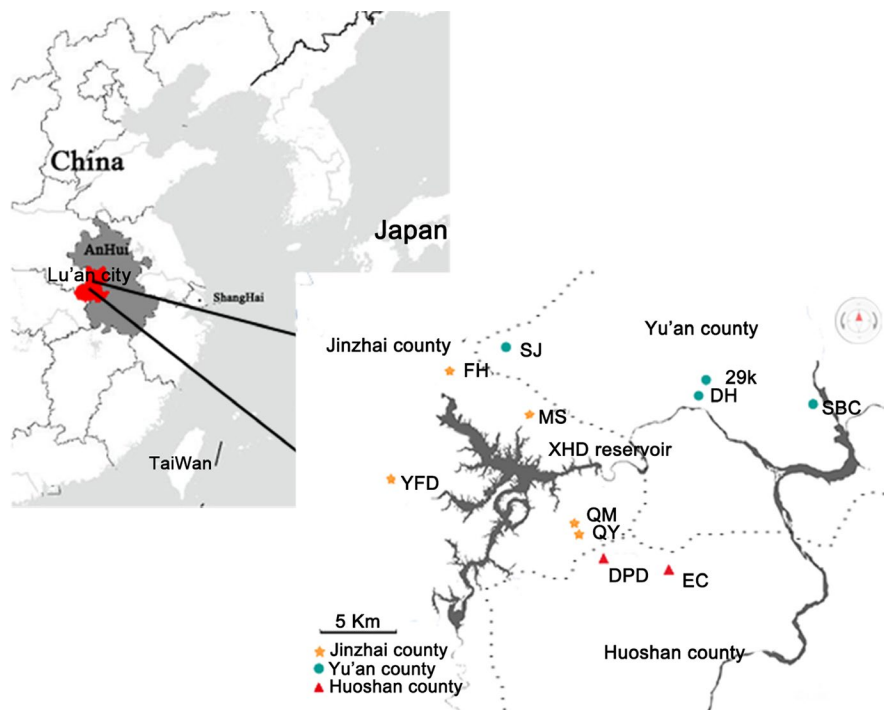
in the same subset have similar properties, which use Euclidean distance (Saito & Toriwaki, 1994) to calculate the distance (similarity) between different categories of data points. A more general alternative is the weighted Euclidean distance between two vectors  $x_i$  and  $x_j$ ,

$$d_{ij} = \left[ \sum_{k=1}^K w_k (x_{i,k} - x_{j,k})^2 \right]^{1/2} \quad (1)$$

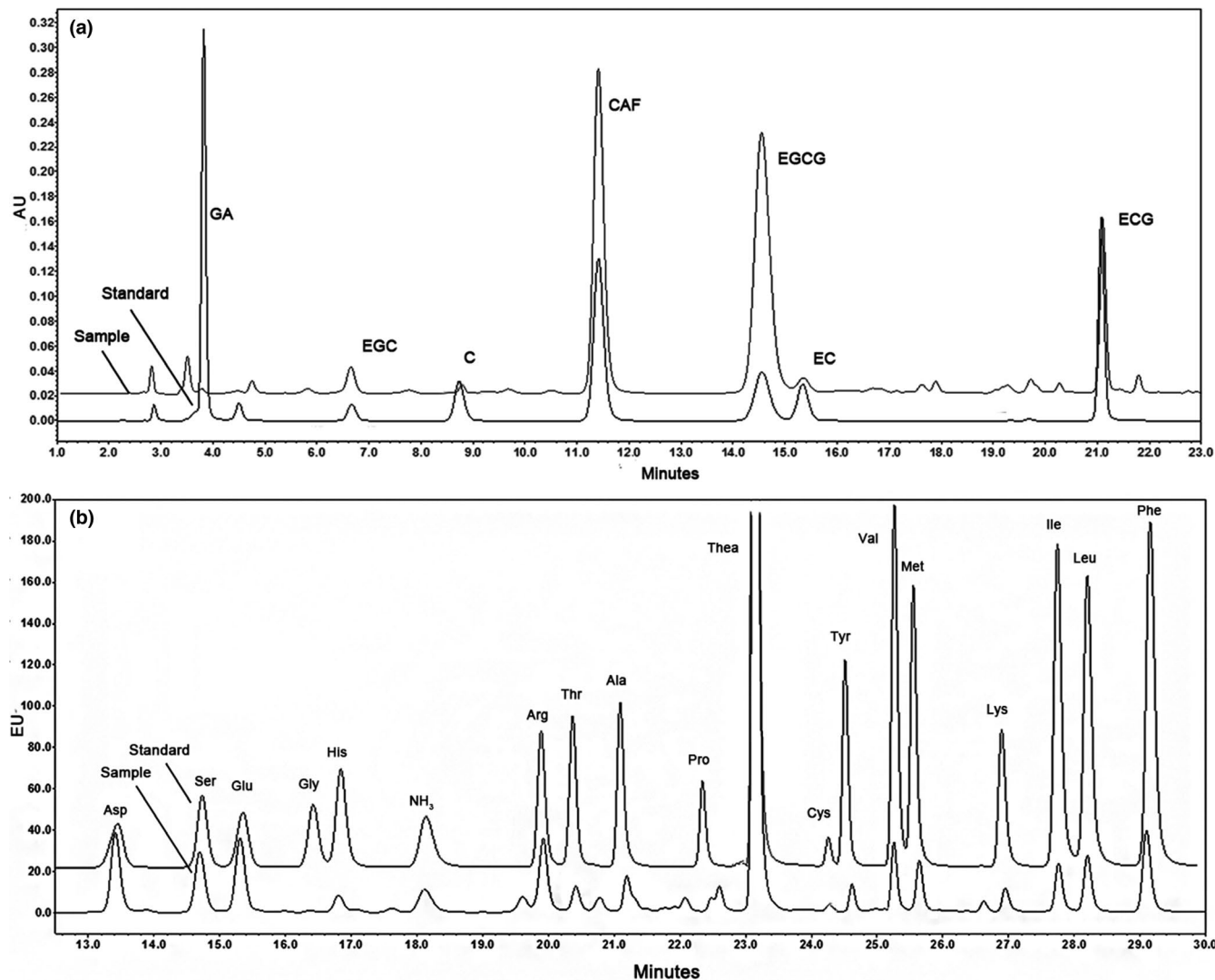
For  $w_k = 1$  for each  $k = 1, \dots, K$ , Equation (1) reduces to the ordinary Euclidean distance.

Correlation analysis refers to the analysis of two or more related variable elements to measure the closeness of the two variable factors (Ezekiel, 1941). The Pearson correlation coefficient, the covariance of the two variables divided by the product of their standard deviations, was used to measure the degree of correlation in this study. Principal component analysis (PCA) (Wold, Esbensen, & Geladi, 1987; Yu, Wang, Xiao, & Liu, 2009), the simplest method for analyzing multivariate statistical distributions with feature variables, is often used to reduce the dimensionality of data sets while maintaining the maximum contribution of the variance within. The obtained results provide a rough classification of all samples.

Support vector machine (SVM) (Vapnik, 2013) is a kind of generalized linear classifier that classifies data according to supervised learning. And the decision boundary is the maximum margin hyperplane for solving learning samples. The SVM can be nonlinearly classified by nuclear method and is one of the common nuclear learning methods, which uses the hinge loss function to calculate the empirical risk and adds a regularization term to the solution system to optimize the structural risk (Chen, Zhao, Fang, & Wang, 2007). Moreover, SVM can learn in high-dimensional feature space with fewer training data.



**FIGURE 1** Sample collection areas



**FIGURE 2** High-performance liquid chromatography (HPLC) chromatogram of catechins, caffeine, and amino acids. HPLC chromatograms of the mixed references of amino acids (a) and of catechins and caffeine (b)

The all-for-mentioned statistical analyses were performed using Matlab 2016a (MathWorks, Inc.). Correlation analysis and one-way ANOVA were performed using SPSS (version 22.0 for Windows).

### 3 | RESULTS AND DISCUSSION

#### 3.1 | Amino acid, GA, catechin, and caffeine detection and analysis

In this study, the contents of all tested substances were quantified using HPLC and provided in Table S2. This study examined two classes of tea: inner mountain LAGP (IMGP) and outer mountain LAGP (OMGP, including YFD, 29K, and SBC). Table 1 shows the differences in the three major metabolite contents for each area. It was remarkable that the concentrations of GA, Met, Leu, Ile, and Val had significant differences ( $p < 0.05$ ) among HSGP, JZGP, and YAGP; and the difference was almost 16%–60% between IMGP and OMGP in contents of EGCG, Ser, Lys, Phe, Val, etc. A heat map (Figure 3) was

created using Hempl (version 1.0) to visualize the content composition in different regions: Red, cyan, and blue areas indicate high, low, and moderate levels of chemical composition, respectively. All LAGP had the same relative contents: high EGCG, caffeine, EGC, ECG, and Thea but low Gly and Cys contents. Hierarchical cluster analysis, an unsupervised learning process, divides similar objects into groups or more subsets through static classification, and members in the same subset have similar properties, which use Euclidean distance to calculate the distance (similarity) between different categories of data points. Huoshan County guapian (HSGP; EC and DPD) and OMGP exhibited great difference with the other regions as shown in HCA.

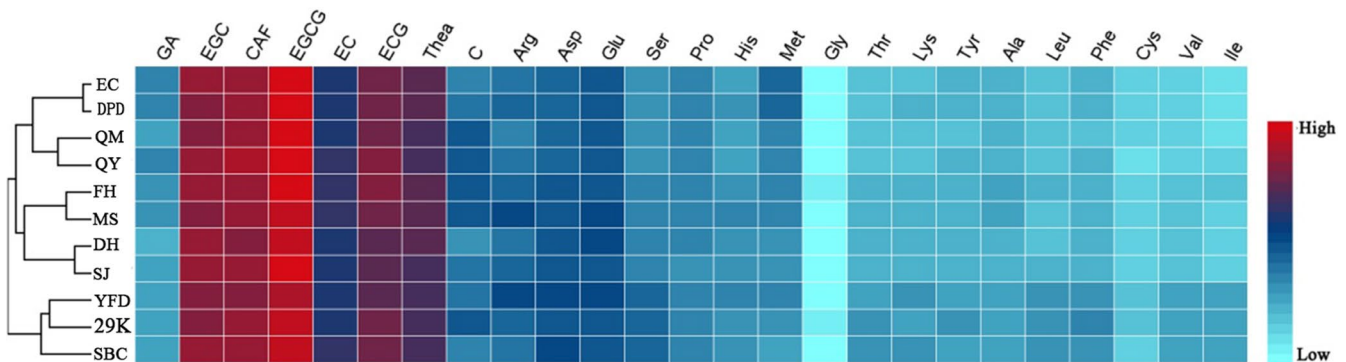
#### 3.2 | Correlation analysis

Correlation analysis between chemical composition and counties/districts based on Pearson correlation coefficient was used, and the Pearson correlation coefficient is shown in Table 2. The value was between +1 and -1, where the numerical value indicates the

**TABLE 1** The content of each substance at different regions (WT%, means  $\pm$  SD)

	Group 1			Group 2	
	HSGP	JZGP	YAGP	IMGP	OMSP
GA	0.049 $\pm$ 0.002 <sup>c</sup>	0.036 $\pm$ 0.002 <sup>b</sup>	0.027 $\pm$ 0.002 <sup>a</sup>	0.036 $\pm$ 0.002	0.033 $\pm$ 0.002
EGC	3.230 $\pm$ 0.173 <sup>b</sup>	2.773 $\pm$ 0.124 <sup>a</sup>	3.328 $\pm$ 0.131 <sup>b</sup>	3.095 $\pm$ 0.105	2.945 $\pm$ 0.144
C	0.062 $\pm$ 0.008 <sup>a</sup>	0.122 $\pm$ 0.008 <sup>b</sup>	0.073 $\pm$ 0.010 <sup>a</sup>	0.090 $\pm$ 0.008	0.105 $\pm$ 0.010
CAF	3.445 $\pm$ 0.117 <sup>ab</sup>	3.785 $\pm$ 0.133 <sup>b</sup>	3.196 $\pm$ 0.103 <sup>a</sup>	3.492 $\pm$ 0.101	3.572 $\pm$ 0.131
EC	0.492 $\pm$ 0.024 <sup>a</sup>	0.512 $\pm$ 0.022 <sup>a</sup>	0.534 $\pm$ 0.025 <sup>a</sup>	0.500 $\pm$ 0.014	0.556 $\pm$ 0.033
EGCG	11.069 $\pm$ 0.510 <sup>b</sup>	8.718 $\pm$ 0.373 <sup>a</sup>	8.023 $\pm$ 0.260 <sup>a</sup>	9.275 $\pm$ 0.315	7.933 $\pm$ 0.280
ECG	1.805 $\pm$ 0.070 <sup>b</sup>	1.939 $\pm$ 0.085 <sup>b</sup>	1.413 $\pm$ 0.040 <sup>a</sup>	1.750 $\pm$ 0.068	1.672 $\pm$ 0.061
Asp	0.109 $\pm$ 0.004 <sup>a</sup>	0.147 $\pm$ 0.009 <sup>b</sup>	0.156 $\pm$ 0.008 <sup>b</sup>	0.137 $\pm$ 0.006	0.159 $\pm$ 0.010
Ser	0.036 $\pm$ 0.001 <sup>a</sup>	0.059 $\pm$ 0.005 <sup>b</sup>	0.071 $\pm$ 0.006 <sup>b</sup>	0.051 $\pm$ 0.003	0.079 $\pm$ 0.007
Glu	0.149 $\pm$ 0.004 <sup>a</sup>	0.170 $\pm$ 0.008 <sup>a</sup>	0.165 $\pm$ 0.009 <sup>a</sup>	0.163 $\pm$ 0.006	0.167 $\pm$ 0.010
Gly	0.003 $\pm$ 0.000 <sup>a</sup>	0.003 $\pm$ 0.000 <sup>a</sup>	0.003 $\pm$ 0.000 <sup>a</sup>	0.003 $\pm$ 0.000	0.004 $\pm$ 0.000
His	0.031 $\pm$ 0.001 <sup>a</sup>	0.039 $\pm$ 0.002 <sup>a</sup>	0.039 $\pm$ 0.003 <sup>a</sup>	0.036 $\pm$ 0.002	0.042 $\pm$ 0.003
Arg	0.094 $\pm$ 0.006 <sup>ab</sup>	0.123 $\pm$ 0.012 <sup>b</sup>	0.088 $\pm$ 0.006 <sup>a</sup>	0.098 $\pm$ 0.007	0.123 $\pm$ 0.013
Thr	0.015 $\pm$ 0.000 <sup>a</sup>	0.019 $\pm$ 0.001 <sup>a</sup>	0.025 $\pm$ 0.002 <sup>b</sup>	0.017 $\pm$ 0.001	0.028 $\pm$ 0.002
Ala	0.017 $\pm$ 0.001 <sup>a</sup>	0.021 $\pm$ 0.001 <sup>b</sup>	0.022 $\pm$ 0.001 <sup>b</sup>	0.019 $\pm$ 0.001	0.024 $\pm$ 0.001
Pro	0.051 $\pm$ 0.002 <sup>a</sup>	0.054 $\pm$ 0.002 <sup>a</sup>	0.053 $\pm$ 0.002 <sup>a</sup>	0.051 $\pm$ 0.002	0.057 $\pm$ 0.002
Cys	0.009 $\pm$ 0.002 <sup>a</sup>	0.009 $\pm$ 0.002 <sup>a</sup>	0.011 $\pm$ 0.002 <sup>b</sup>	0.009 $\pm$ 0.002	0.011 $\pm$ 0.002
Tyr	0.017 $\pm$ 0.002 <sup>a</sup>	0.020 $\pm$ 0.002 <sup>a</sup>	0.026 $\pm$ 0.002 <sup>b</sup>	0.019 $\pm$ 0.002	0.027 $\pm$ 0.002
Val	0.008 $\pm$ 0.002 <sup>a</sup>	0.015 $\pm$ 0.002 <sup>b</sup>	0.021 $\pm$ 0.002 <sup>c</sup>	0.013 $\pm$ 0.002	0.023 $\pm$ 0.002
Met	0.097 $\pm$ 0.002 <sup>c</sup>	0.053 $\pm$ 0.002 <sup>b</sup>	0.037 $\pm$ 0.002 <sup>a</sup>	0.061 $\pm$ 0.002	0.039 $\pm$ 0.002
Lys	0.016 $\pm$ 0.002 <sup>a</sup>	0.020 $\pm$ 0.002 <sup>ab</sup>	0.024 $\pm$ 0.002 <sup>b</sup>	0.018 $\pm$ 0.002	0.027 $\pm$ 0.002
Ile	0.007 $\pm$ 0.002 <sup>a</sup>	0.012 $\pm$ 0.002 <sup>b</sup>	0.018 $\pm$ 0.002 <sup>c</sup>	0.011 $\pm$ 0.002	0.020 $\pm$ 0.002
Leu	0.012 $\pm$ 0.002 <sup>a</sup>	0.018 $\pm$ 0.00 <sup>b</sup>	0.024 $\pm$ 0.002 <sup>c</sup>	0.016 $\pm$ 0.002	0.026 $\pm$ 0.002
Phe	0.016 $\pm$ 0.002 <sup>a</sup>	0.021 $\pm$ 0.002 <sup>a</sup>	0.032 $\pm$ 0.002 <sup>b</sup>	0.019 $\pm$ 0.002	0.035 $\pm$ 0.002
Thea	1.076 $\pm$ 0.002 <sup>a</sup>	1.003 $\pm$ 0.002 <sup>a</sup>	0.993 $\pm$ 0.002 <sup>a</sup>	0.999 $\pm$ 0.002	1.045 $\pm$ 0.002

Note: Values in the same row of Group 1 that are labeled with different superscript letters (a–c) differ significantly ( $p < 0.05$ ). Statistical analysis was one-way ANOVA with pairwise post hoc comparisons by the method of Bonferroni.

**FIGURE 3** Heat map of 25 variables and the results of hierarchical clustering for origins

correlation strength, and “+” or “–” indicates the positive or negative correlation. The variables strongly correlated with IMG and OMSP were Phe, Leu, Ile, and Val, with all correlation coefficients being  $>0.85$ , whereas those strongly correlated with Jinzhai County LAGP (JZGP), Yu’an District LAGP (YAGP), and HSGP were GA and

Met, with all correlation coefficients being  $>0.5$ . The HCA results afforded great similarity among the LAGP types. A study of the found components revealed that the main bitter amino acids in the tea included L-Ile, L-Leu, L-Tyr, L-Phe, and L-Val; the more hydrophobic the R group of the amino acid side chain is, the more bitter the taste is

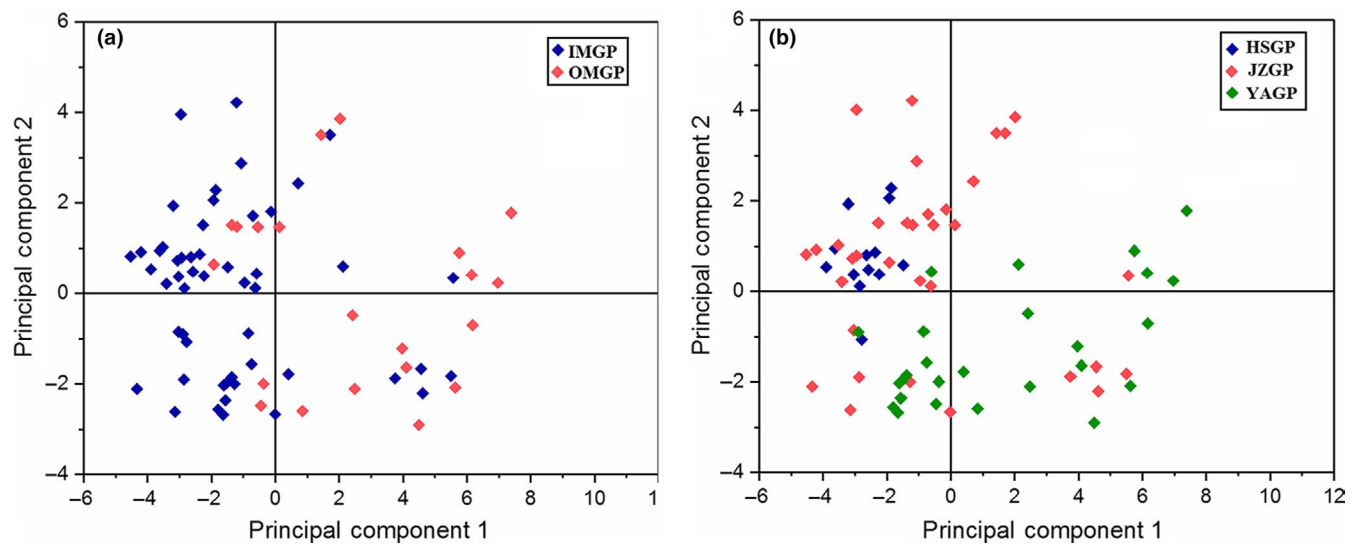
**TABLE 2** Correlation analysis for chemical composition and region

	GA	EGC	C	CAF	EC	EGCG	ECG	Asp	
Group 1	-0.319**	-0.250*	-0.139	-0.317**	-0.039	-0.527**	-0.451**	0.538**	
Group 2	-0.575**	0.130	-0.041	-0.208	0.126	-0.467**	-0.405**	0.313**	
	Ser	Glu	Gly	Phe	His	Arg	Thr	Ala	
Group 1	0.811**	0.125	0.221	0.877**	0.346**	0.191	0.830**	0.536**	
Group 2	0.405**	0.098	0.145	0.430**	0.163	-0.110	0.464**	0.363**	
	Pro	Cys	Tyr	Val	Met	Lys	Ile	Leu	Thea
Group 1	0.367**	0.519**	0.758**	0.887**	-0.471**	0.820**	0.897**	0.893**	-0.131
Group 2	0.042	0.270*	0.456**	0.481**	-0.779**	0.370**	0.484**	0.450**	-0.136

Note: Pearson correlation coefficient was used for analysis. Group 1 means IMGP and OMGP; Group 2 means JZGP, YAGP, and HSGP.

\*Correlation is significant at a 0.05 level (two-tailed);

\*\*Correlation is significant at a 0.01 level (two-tailed).

**FIGURE 4** Principal component analysis (PCA) scoring plots for IMGP, OMGP (a) and JZGP, YAGP, HSGP (b)

(Mao et al., 2018). GA content is significantly positively correlated with the quality grade of green tea, and GA is a characteristic component of green tea (Graham, 1992). These findings match with the results of sensory evaluation: Inner mountain tea has a less bitter, mellower taste than outer mountain tea does. However, the geographical tracing was negatively correlated with the geographical distance. Some sampling points, such as SJ and FH as well as QY and FH, border each other, and sometimes, they are even two sides of the same mountain. For example, the place adjacent to Jinzhai County and Huoshan County is the famous Dabie Mountains.

### 3.3 | Principal component analysis

Principal component analysis, as one of the commonly used multivariate analysis methods used for preliminary classification (Chen et al., 2015), was applied to extract the factors associated with production regions. Two principal components, explaining up to 96.87% of total variance, were revealed. Scoring plots are presented in

Figure 4, with Figure 4a,b showing the classification of Group 1 and Group 2, respectively. Among them, PC1 explained 92.59% of the total variance, with variation mainly in GA, EGCG, and ECG contents; by contrast, PC2 explained 4.28% of the total variance, with variation mainly in EGC, C, and ECG contents. The distinction between the inner mountain and the outer mountain is more obvious, but the county area is mixed and the catechins seem to play a major role from the results of PCA. HCA and PCA are two methods of unsupervised pattern recognition, which obtained the preliminary classification information. The further visualized analysis of samples by combining supervised pattern recognition SVM is necessary.

### 3.4 | Factor selection according to HCA, correlation analysis, and PCA

The final purpose of this study was to analyze the difference and discriminate LAGP from different origins using main chemical components combining pattern recognition. The results of factor selection

**TABLE 3** Results of factor selection and optimization

Methods	Screening criteria	Factors	
		Inner and outer mountains	Counties
Original factors	All components being tested	Polyphenols, amino acids, CAF	Polyphenols, amino acids, CAF
HCA	Euclidean distance	C, EC, ECG, Asp, Glu, Arg, Thea	EGC, C, Caf, EC, ECG, Thea, Arg, Asp, Glu
Correlation analysis	Pearson correlation coefficient	EGCG, Asp, Phe, Thr, Ala, Cys, Ser, Tyr, Val, Lys, Ile, Leu	GA, EGCG, ECG, Ser, Phe, Thr, Tyr, Val, Met, Ile, Leu
PCA	The principal components	GA, C, EGC, EGCG, ECG	GA, C, EGC, EGCG, ECG

**TABLE 4** Results of SVM with different factors

Samples	Factors selected (amount)	Correct discrimination rate	
		Training set	Prediction set
Inner and outer mountains	Origin (25)	100% (49/49)	91.67% (22/24)
	Origin–HCA (7)	100% (49/49)	95.83% (23/24)
	Origin–correlation analysis (12)	97.96% (48/49)	83.33% (20/24)
	Origin–PCA (5)	97.96% (48/49)	95.83% (23/24)
Counties	Origin (25)	93.88% (46/49)	95.83% (23/24)
	Origin–HCA (9)	87.76% (43/49)	87.50% (21/24)
	Origin–correlation analysis (11)	91.84% (45/49)	79.17% (19/24)
	Origin–PCA (5)	91.84% (45/49)	95.83% (23/24)

are listed in Table 3. The selection upshot of HCA depends on the relative magnitude of the Euclidean distance, and PCA relies on the coefficient values of the variables that make up PC1 and PC2, especially.

### 3.5 | Geographical tracing of LAGP through SVM combining with different factors

Support vector machine combining with the method of original variables and the factors selected by HCA, correlation analysis, and PCA all realized high discrimination statistically except for correlation analysis (Table 4). The results revealed that the model could obtain a high identification rate combining with PCA. The training and prediction sets of the correct classification for inner and outer mountains were 97.96% and 95.83%, with only two samples discriminated incorrectly. However, the discrimination between counties was worse than for IMGP and OMGP; the correct discrimination rates of the training set and prediction set were 91.84% and 95.83%. These results indicated that distinguishing LAGP by using SVM is feasible.

The geographical origins of tea are difficult to trace because of the many associated factors necessary to determine origin. The composition of contained matter of tea can be easily changed by climate (Wang et al., 2011), agricultural practices in tea plantations (Wang et al., 2016), and tea processing (Huang, Sheng, Yang, & Hu, 2007). Of these, geographical location is one of the most critical

factors. The discrimination rate in this study between the inner and outer mountain samples was high, possibly because of the relationship with altitude exhibited by the samples (Table S1). In addition, slight discrepancy of parameter caused by working operations among different production areas might affect geographical tracing, too. Huoshan County and Jinzhai County are traditional LAGP processing areas that emphasize “heat” during processing, which makes JZGP and HSGP highly dissimilar to YAGP and might also explain the reason that identification of YAGP was easier.

## 4 | CONCLUSIONS

Some studies have examined geographical tracing of tea products in terms of tea categories or tea-planting areas. This study shows that the analytical information extracted from chemical components and the method of pattern recognition can be used to discriminate the origin of LAGP. Hence, 25 component contents and their combinations with differing multivariate classifiers (HCA, correlation analysis, and PCA) were selected. Of these, Ile, Phe, Val, Leu, and Ser were key factors in the classification of IMGP and OMGP. Furthermore, GA and Met might be useful in classification of counties. The best result in classification was obtained by using SVM and PCA, which reached 95.83% classification accuracy in prediction set of inner and outer mountain and counties samples,

respectively. The proposed method successfully achieved origin discrimination of LAGP without sensory evaluation. However, any generalization of the proposed methodology requires more extensive and varied testing of LAGP samples with tea samples from numerous years.

## ACKNOWLEDGMENTS

This work was financially supported by the National Key Research and Development Plan (No. 2017YFD040800) and Modern Agriculture (tea) Special System of China (No. CARS-19).

## CONFLICT OF INTEREST

Huan Su, Weiquan Wu, Xiaochun Wan, and Jing ming Ning declare that they do not have any conflict of interest.

## ETHICAL REVIEW

This study does not involve any human or animal testing.

## INFORMED CONSENT

Written informed consent was obtained from all study participants.

## ORCID

Huan Su  <https://orcid.org/0000-0002-4257-5393>

## REFERENCES

- Ananingsih, V. K., Sharma, A., & Zhou, W. (2013). Green tea catechins during food processing and storage: A review on stability and detection. *Food Research International*, 50(2), 469–479. <https://doi.org/10.1016/j.foodres.2011.03.004>
- Bai, W.-X., Wang, C., Wang, Y.-J., Zheng, W.-J., Wang, W., Wan, X.-C., & Bao, G.-H. (2017). Novel acylated flavonol tetraglycoside with inhibitory effect on lipid accumulation in 3T3-L1 cells from Lu'an GuaPian tea and quantification of flavonoid glycosides in six major processing types of tea. *Journal of Agricultural and Food Chemistry*, 65(14), 2999–3005. <https://doi.org/10.1021/acs.jafc.7b00239>
- Chen, Q., Liu, A., Zhao, J., & Ouyang, Q. (2013). Classification of tea category using a portable electronic nose based on an odor imaging sensor array. *Journal of Pharmaceutical and Biomedical Analysis*, 84, 77–83. <https://doi.org/10.1016/j.jpba.2013.05.046>
- Chen, Q., Sun, C., Ouyang, Q., Wang, Y., Liu, A., Li, H., & Zhao, J. (2015). Classification of different varieties of Oolong tea using novel artificial sensing tools and data fusion. *Lwt—food Science and Technology*, 60(2), 781–787. <https://doi.org/10.1016/j.lwt.2014.10.017>
- Chen, Q., Zhao, J., Fang, C. H., & Wang, D. (2007). Feasibility study on identification of green, black and Oolong teas using near-infrared reflectance spectroscopy based on support vector machine (SVM). *Spectrochimica Acta Part A Molecular & Biomolecular Spectroscopy*, 66(3), 568–574. <https://doi.org/10.1016/j.saa.2006.03.038>
- Cooper, R. (2012). Green tea and theanine: Health benefits. *International Journal of Food Sciences and Nutrition*, 63(sup1), 90–97. <https://doi.org/10.3109/09637486.2011.629180>
- Cunha, S. C., Amaral, J. S., Fernandes, J. O., & Oliveira, M. B. P. (2006). Quantification of tocopherols and tocotrienols in Portuguese olive oils using HPLC with three different detection systems. *Journal of Agricultural and Food Chemistry*, 54(9), 3351–3356. <https://doi.org/10.1021/jf053102n>
- Ding, Y., Yu, H., & Mou, S. (2002). Direct determination of free amino acids and sugars in green tea by anion-exchange chromatography with integrated pulsed amperometric detection. *Journal of Chromatography A*, 982(2), 237–244. [https://doi.org/10.1016/S0021-9673\(02\)01650-3](https://doi.org/10.1016/S0021-9673(02)01650-3)
- Dulloo, A. G., Seydoux, J., Girardier, L., Chantre, P., & Vandermander, J. (2000). Green tea and thermogenesis: Interactions between catechin-polyphenols, caffeine and sympathetic activity. *International Journal of Obesity*, 24(2), 252. <https://doi.org/10.1038/sj.ijo.0801101>
- Ezekiel, M. (1941). *Methods of correlation analysis*. Hoboken, NJ: John Wiley & Sons.
- Fan, P., Xiaopei, B. O., Wang, M., & Han, B. (2017). Similarities and differences in composition of amino acids in four grades of Liu'an Neishan Guapian tea. *Journal of Anhui Agricultural University*, 44(1), 14–21. (in Chinese).
- Graham, H. N. (1992). Green tea composition, consumption, and polyphenol chemistry. *Preventive Medicine*, 21(3), 334–350. [https://doi.org/10.1016/0091-7435\(92\)90041-F](https://doi.org/10.1016/0091-7435(92)90041-F)
- Heiss, M. L. (2001). Chinese spring green tea. *Gastronomica*, 1(2), 80–83. <https://doi.org/10.1525/gfc.2001.1.2.80>
- Hermosín, I., Chicón, R. M., & Dolores Cabezudo, M. (2003). Free amino acid composition and botanical origin of honey. *Food Chemistry*, 83(2), 263–268. [https://doi.org/10.1016/S0308-8146\(03\)00089-X](https://doi.org/10.1016/S0308-8146(03)00089-X)
- Huang, Y., Sheng, J., Yang, F., & Hu, Q. H. (2007). Effect of enzyme inactivation by microwave and oven heating on preservation quality of green tea. *Journal of Food Engineering*, 78(2), 687–692. <https://doi.org/10.1016/j.jfoodeng.2005.11.007>
- Huo, D., Wu, Y., Yang, M., Fa, H., Luo, X., & Hou, C. (2014). Discrimination of Chinese green tea according to varieties and grade levels using artificial nose and tongue based on colorimetric sensor arrays. *Food Chemistry*, 145, 639–645. <https://doi.org/10.1016/j.jpba.2013.05.046>
- ISO, B. S. 14502-2: 2005 (2006). *Determination of substances characteristic of green and black tea*. (Part 2: Content of catechins in green tea method using high-performance liquid chromatography).
- Jang, J. H., Kim, E. S., Wu, S. Y., Lu, J. L., Liang, H. L., Du, Y. Y., ... Liang, Y. R. (2008). Assessing geographic origins of green teas using instruments. *Food Science & Biotechnology*, 17(5), 1016–1020.
- Li, X., Jin, J., Sun, C., Ye, D., & Liu, Y. (2019). Simultaneous determination of six main types of lipid-soluble pigments in green tea by visible and near-infrared spectroscopy. *Food Chemistry*, 270, 236–242. <https://doi.org/10.1016/j.foodchem.2018.07.039>
- Luo, D., Chen, J., Gao, L., Liu, Y., & Wu, J. (2017). Geographical origin identification and quality control of Chinese chrysanthemum flower teas using gas chromatography-mass spectrometry and olfactometry and electronic nose combined with principal component analysis. *International Journal of Food Science & Technology*, 52(3), 714–723. <https://doi.org/10.1111/ijfs.13326>
- Mao, A., Su, H., Fang, S., Chen, X., Ning, J., Ho, C., & Wan, X. (2018). Effects of roasting treatment on non-volatile compounds and taste of green tea. *International Journal of Food Science & Technology*, 53(11), 2586–2594. <https://doi.org/10.1111/ijfs.13853>
- Meng, W., Xu, X., Cheng, K.-K., Xu, J., Shen, G., Wu, Z., & Dong, J. (2017). Geographical origin discrimination of oolong tea (*TieGuanYin*, *Camellia sinensis* (L.) O. Kuntze) using proton nuclear magnetic resonance spectroscopy and near-infrared spectroscopy. *Food Analytical Methods*, 10(11), 3508–3522. <https://doi.org/10.1007/s12161-017-0920-4>
- Ning, J., Cao, Q., Su, H., Zhu, X., Wang, K., Wan, X., & Zhang, Z. (2017). Discrimination of six tea categories coming from different origins



- depending on polyphenols, caffeine and theanine combined with different discriminant analysis. *International Journal of Food Properties*, 20(sup2), 1838–1847. <https://doi.org/10.1080/10942912.2017.1354880>
- Rastija, V., & Srećnik, G. (2009). Polyphenolic composition of Croatian wines with different geographical origins. *Food Chemistry*, 115(1), 54–60. <https://doi.org/10.1016/j.foodchem.2008.11.071>
- Reverter, M., Lundh, T., & Lindberg, J. E. (1997). Determination of free amino acids in pig plasma by precolumn derivatization with 6-N-aminoquinolyl-N-hydroxysuccinimidyl carbamate and high-performance liquid chromatography. *Journal of Chromatography B: Biomedical Sciences and Applications*, 696(1), 1–8. [https://doi.org/10.1016/S0378-4347\(97\)00217-X](https://doi.org/10.1016/S0378-4347(97)00217-X)
- Saito, T., & Toriwaki, J. (1994). New algorithms for euclidean distance transformation of an n-dimensional digitized picture with applications. *Pattern Recognition*, 27(11), 1551–1565. [https://doi.org/10.1016/0031-3203\(94\)90133-3](https://doi.org/10.1016/0031-3203(94)90133-3)
- Sinija, V. R., & Mishra, H. N. (2008). Green tea: Health benefits. *Journal of Nutritional & Environmental Medicine*, 17(4), 232–242. <https://doi.org/10.1080/13590840802518785>
- Vapnik, V. (2013). *The nature of statistical learning theory*. Berlin, Germany: Springer Science & Business Media.
- Wang, L. Y., Wei, K., Jiang, Y. W., Cheng, H., Zhou, J., He, W., & Zhang, C. C. (2011). Seasonal climate effects on flavanols and purine alkaloids of tea (*Camellia sinensis* L.). *European Food Research & Technology*, 233(6), 1049–1055. <https://doi.org/10.1007/s00217-011-1588-4>
- Wang, W., Min, Q., Sardans, J., Wang, C., Asensio, D., Bartrons, M., & Peñuelas, J. (2016). Organic cultivation of jasmine and tea increases carbon sequestration by changing plant and soil stoichiometry. *Agronomy Journal*, 108(4), 1636–1648. <https://doi.org/10.2134/agronj2015.0559>
- Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1–3), 37–52. [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9)
- Yu, H., Wang, J., Xiao, H., & Liu, M. (2009). Quality grade identification of green tea using the eigenvalues of PCA based on the E-nose signals. *Sensors and Actuators B: Chemical*, 140(2), 378–382. <https://doi.org/10.1016/j.snb.2009.05.008>
- Yudthavorasit, S., Wongravee, K., & Leepipatpiboon, N. (2014). Characteristic fingerprint based on gingerol derivative analysis for discrimination of ginger (*Zingiber officinale*) according to geographical origin using HPLC–DAD combined with chemometrics. *Food Chemistry*, 158, 101–111. <https://doi.org/10.1016/j.foodchem.2014.02.086>
- Zhan, H., Liu, S., & Yu, J. (2017). Research on factors influencing consumers' loyalty towards geographical indication products based on grey incidence analysis. *Grey Systems: Theory and Application*, 7(3), 397–407. <https://doi.org/10.1108/GS-10-2016-0037>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Su H, Wu W, Wan X, Ning J. Discriminating geographical origins of green tea based on amino acid, polyphenol, and caffeine contents through high-performance liquid chromatography: Taking Lu'an guapian tea as an example. *Food Sci Nutr*. 2019;7:2167–2175. <https://doi.org/10.1002/fsn3.1062>