Article

Discriminant Analysis of Pu-Erh Tea of Different Raw Materials Based on Phytochemicals Using Chemometrics

Shao-Rong Zhang [1], Yu Shi [1], Jie-Lin Jiang [2], Li-Yong Luo [1,3] and Liang Zeng [1,3,*]

1   College of Food Science, Southwest University, No. 2 Tiansheng Road, Beibei District, Chongqing 400715, China; qianzhihe225093@126.com (S.-R.Z.); shi3796@icloud.com (Y.S.); liyongluo1979@126.com (L.-Y.L.)
2   Taetea Group, Menghai Tea Factory, Xishuangbanna Dai Autonomous Prefecture, Menghai 665000, China; dyjiangjielin@126.com
3   Tea Research Institute, Southwest University, No. 2 Tiansheng Road, Beibei District, Chongqing 400715, China
*   Correspondence: zengliangbaby@126.com

Abstract: Pu-erh tea processed from the sun-dried green tea leaves can be divided into ancient tea (AT) and terrace tea (TT) according to the source of raw material. However, their similar appearance makes AT present low market identification, resulting in a disruption in the tea market rules of fair trade. Therefore, this study analyzed the classification by principal component analysis/hierarchical clustering analysis and conducted the discriminant model through stepwise Fisher discriminant analysis and decision tree analysis based on the contents of water extract, phenolic components, alkaloid, and amino acids, aiming to investigate whether phytochemicals coupled with chemometric analyses distinguish AT and TT. Results showed that there were good separations between AT and TT, which was caused by 16 components with significant ($p < 0.05$) differences. The discriminant model of AT and TT was established based on six discriminant variables including water extract, (+)-catechin, (−)-epicatechin, (−)-epigallocatechin, theacrine, and theanine. Among them, water extract comprised multiple soluble solids, representing the thickness of tea infusion. The model had good generalization capability with 100% of performance indexes according to scores of the training set and model set. In conclusion, phytochemicals coupled with chemometrics analyses are a good approach for the identification of different raw materials.

Keywords: Pu-erh tea; raw material; phytochemical; chemometrics analyses; discriminant model; generalization capability

1. Introduction

Pu-erh tea is defined as a geographical indication product by the General Administration of Quality Supervision, Inspection, and Quarantine of the People's Republic of China (GB/T 22111-2008), and is one of the most popular tea beverages in Asian countries; in particular, southwestern China and South Asian countries attribute this to its unique flavors and beneficial effects on human health [1]. It is processed from the sun-dried green tea leaves and can be classified into ancient tea (Gu-shu cha, AT) and terrace tea (Tai-di cha, TT) based upon the source of raw material [2,3]. AT is collected from ancient tea gardens that have good economic and ecological efficiencies such as climate regulation, water and soil conservation whereas TT is gathered from terrace tea plantations that rely on the good management of tea fields including fertilization, pruning, and pesticide spraying. The differences in growth environments and management methods of both result in different flavors: compared with TT, AT has a richer taste with durability and a more distinctive aroma, which is widely considered to have more preservation value [4]. However, their similar appearance has led to low market identification of AT, which seriously damages the interests of consumers and the reputation of tea producers. Meanwhile, lack of yield and

the unclear relevant laws or market regulation of AT have also created an opportunity for tea adulteration [5].

Based on the phenomenon, numerous researchers have investigated the differences of AT and TT from phytochemicals, soil nutrients, and flavor, but most studies have simply compared them based on a single component because there is still a lack of methods to systematically identify the differences between AT and TT in the case of multiple samples and multiple indicators [6–9]. Lin et al. [10] established discriminant models of AT and TT through rare earth elements, but these components were inapplicable and difficult to understand for tea drinkers, especially consumers without professional knowledge in the tea market. Therefore, it is necessary to develop a valid method through more intuitive indicators to scientifically distinguish AT from TT. Chemometric analyses are a data analysis tool to extract effective information from multivariate chemical data. This is often combined with other methods such as high-performance liquid chromatography (HPLC) and mass spectrometry to analyze food adulteration in the market and explore whether there are similar characteristics of unknown samples [11,12]. Principal component analysis (PCA), hierarchical clustering analysis (HCA), and the establishment of discriminant model are the main chemometric techniques [13]. According to the results on discriminating Brazilian propolis using chemometrics by AF Mottese [14], it was found that DA/DTA was complementary to PCA/HCA and the important information of the discriminant model could be obtained based on them. At present, they have been widely applied in the identification of food adulteration in products such as butter, camellia oil, honey, and tea [15–21]. For example, Zhou et al. [18] correctly identified Pu-erh ripened teas of different production origins based on PCA and HCA; Aboulwafa et al. [22] studied the quality of green tea samples from the South and the East Asian regions, and the results showed that there were good separations of samples between them by establishing the model. Phytochemicals including phenolics, free amino acids, caffeine, and other components could intuitively reflect tea quality through flavor [23] whereas their combination could prove to be a useful tool to distinguish Pu-erh tea made from different raw materials.

Hence, with regard to the scientific classification of Pu-erh tea from different raw materials and guide tea producers and consumers to effectively identify tea products, 30 ATs and 50 TTs from Yunnan Province were collected in this work, and 26 phytochemicals including three mixtures, six catechin components, two purine alkaloids, and 15 free amino acids were determined by spectrophotometry method and high-performance liquid chromatography (HPLC), respectively. Further chemometric analyses were used to reveal classification accuracy and construct a model.

2. Materials and Methods

2.1. Chemicals

Used chemicals were of analytical grade unless otherwise stated. Folin–Ciocalteu reagent, sodium carbonate ($NaCO_3$), methanol, ninhydrin, stannous chloride ($SnCl_2$), disodium phosphate (Na_2HPO_4), potassium dihydrogen phosphate (KH_2PO_4), 5-sulfosalicylic acid, and glacial acetic acid were purchased from Taixin Chemical Company (Chongqing, China), among which methanol and glacial acetic acid were HPLC-grade. Gallic acid and glutamate were obtained from the Kelong Chemical Factory (Chengdu, China). Aspartic acid (Asp), serine (Ser), glutamic acid (Glu), glycine (Gly), alanine (Ala), cysteine (Cys), valine (Val), isoleucine (Ile), leucine (Leu), tyrosine (Tyr), phenylalanine (Phe), lysine (Lys), histidine (His), arginine (Arg), and theanine were guaranteed reagents purchased from Seebio Technology Co. Ltd. (Shanghai, China). Caffeine, (−)-epigallocatechin gallate (EGCG), (−)-gallocatechin gallate (GCG), (−)-epicatechin gallate (ECG), (−)-epigallocatechin (EGC), (−)-epicatechin (EC), and (+)-catechin (C) were guaranteed reagents purchased from Chengdu Biopurify Phytochemicals Ltd. (Chengdu, China). Theacrine was obtained from Better-in Com (Shanghai, China).

2.2. Tea Samples

As shown in Table 1, a total of 80 samples including 30 ATs and 50 TTs were collected from five ancient tea gardens (including Banpen, Laobanzhang, Hekai, Xinbanzhang, Laoman'e) and nine terrace tea plantations (including Nannuo, Bulang, Mensong, Lincang, Lancang, Xiding, Gelanghe, Menla, Dali) in Yunnan Province. The raw Pu-erh samples were processed through five steps including picking, withering, green removing, rolling and twisting, and sun-drying, and marked with the corresponding numbers according to the sample names before grinding. All samples were stored at $-4\,^\circ$C until further analysis.

Table 1. The sources, counts, and codes of different raw materials in Yunnan Province.

Raw Materials	Sources	Counts	Codes [1]
AT	Banpen, Laobanzhang, Hekai, Xinbanzhang, Laoman'e	30	A1~A30
TT	Nannuo, Bulang, Mensong, Lincang, Lancang, Xiding, Gelanghe, Menla, Dali	50	T1~T50

[1] The codes of AT and TT were respectively named as the uppercase of the first letter of them, followed by a number.

2.3. Determination of Phytochemicals

2.3.1. Water Extract (WE)

WE in the AT and TT was detected by the constant temperature drying method [24] with some modifications. Briefly, 1 g of the ground sample (m_0) was soaked in 150 mL boiled distilled water for 45 min. After washing and filtration using 75 mL boiled distilled water, tea grounds were put in an oven ($120 \pm 2\,^\circ$C) to bake for 4 h before weighing (m_1). The WE content was expressed as ($m_0 - m_1$) $\times 1000/m_0$ mg/g.

2.3.2. Total Phenolics (TPC)

The determination of TPC content was performed using Folin–Ciocalteu reagent [25]. Briefly, 0.2 g of the ground sample and 5 mL of 70% methanol (70 $^\circ$C preheat) were placed in a 70 $^\circ$C water bath pot for 10 min and centrifuged at 3500 rpm for 10 min after being cooled. The above operation was repeated, and all supernatants were merged to a constant volume of 10 mL. Then, 1 mL of the sample or water, appropriately diluted, was taken, followed by 5 mL of 10% Folin–Ciocalteu reagent. After 5–8 min at room temperature, 4 mL of 7.5% sodium carbonate solution was added and the mixture was placed at $25 \pm 2\,^\circ$C in the dark for 1 h. The absorbance was measured at 765 nm using a Synergy H1MG microplate reader (Synergy H1MG; BioTek Instruments Inc., Winooski, VT, USA). Gallic acid (0–0.0055 mg/g) was used as the reference standard, and the results were expressed as gallic acid equivalents per gram sample (mg/g).

2.3.3. Total Free Amino Acids (TFAAs)

The TFAAs contents of samples were identified using the ninhydrin colorimetric assay [26]. Briefly, 1 g of the ground sample and 150 mL water were placed in a boiling water bath for 45 min, and filtrated through decom-pressure filtration was added to water, yielding a 150 mL volume. Then, a 1 mL of sample was taken, followed by 0.5 mL phosphate buffer (pH 8) and 0.5 mL 2% ninhydrin. After shaking, the solution was placed in a boiling water bath for 15 min. Then samples were cooled to $25 \pm 2\,^\circ$C and water was added to a 25-mL volume before measurement at 570 nm using a Synergy H1MG microplate reader (Synergy H1MG; BioTek Instruments Inc., Winooski, VT, USA). Glu (0–0.6 mg/g) was used as the reference standard, and the results were expressed as Glu equivalents per gram sample (mg/g).

2.3.4. Catechins Components and Caffeine

Catechin components (EGCG, GCG, ECG, EC, EGC, C) and caffeine were identified using the high-performance liquid chromatography (HPLC) method [27]. Briefly, 0.2 g of the

ground sample and 5 mL of 70% methanol (70 °C preheat) were placed in a 70 °C water bath pot for 10 min, and centrifuged at 3500 rpm for 10 min after being cooled. The above operation was repeated, and all supernatants were merged to a constant volume of 10 mL. The above solution was filtered through a 0.45-μm organic membrane (Jinteng Experimental Equipment Co. Ltd., Tianjin, China) until further HPLC analysis. The chromatographic conditions were as follows: Agilent ZORBAX SB-C18 (5 μm, 4.6 mm × 250 mm); detection wavelength at 278 nm; mobile phase A, 0.2% glacial acetic acid in water; mobile phase B, methanol; flow rate, 0.9 mL/min; column temperature, 35 °C; and injection volume, 10 μL; gradient elution, 18–25% B, 0–25 min; 25–35% B, 25–30 min; 35–18% B, 30–32 min; and 18% B, 32–37 min. Catechin components and caffeine levels were expressed as mg/g.

2.3.5. Theacrine

The theacrine in the AT and TT was identified according to Li et al. [28] with some modifications. Briefly, the ground sample was extracted with 75% alcohol in a ratio of 1:20 (w/v) for 10 min and centrifuged at 12,000 rpm for 10 min. Then, the obtained supernatant was filtered through a 0.45-μm organic membrane (Jinteng Experimental Equipment Co. Ltd., Tianjin, China) until further HPLC analysis. The chromatographic conditions were as follows: Agilent ZORBAX SB-C18 (5 μm, 4.6 mm × 250 mm); detection wavelength at 280 nm; mobile phase A, acetonitrile/acetic acid/water (3:0.5:96.5, $v/v/v$); mobile phase B, acetonitrile/acetic acid/water (30:0.5:69.5, $v/v/v$); flow rate, 1.0 mL/min; column temperature, 32 °C; and injection volume, 10 μL; gradient elution, 20–80% B, 0–35 min; 20% B, 35.01 min; 20% B, 35.01–40 min. The theacrine level was expressed as mg/g.

2.3.6. Free Amino Acids

The free amino acids were detected according to Lu et al. [29] using an Amino Acid Analyzer (L-8900, Hitachi, Tokyo, Japan). Briefly, a 4 mL sample and 4 mL 10% sulfosalicylic acid were added to a 10-mL tube. After one night, the sample was filtered through a 0.45-μm organic membrane (Jinteng Experimental Equipment Co. Ltd., Tianjin, China) before analysis. The Amino Acid Analyzer system used a mobile phase involving lithium citrate and UV–Vis detection at 440 nm and 570 nm. The flow rates were 0.35 mL/min for the mobile phase and 0.3 mL/min for the derivatization reagent. The column temperature was set to 50 °C, and the post-column reaction equipment was maintained at 135 °C. The temperature of the autosampler was kept at 4 °C, and the injection volume was 20 μL for the standard and samples. The free amino acids levels were expressed as mg/g.

2.4. Chemometric Analyses

All data were subjected separately to PCA and HCA to detect whether phytochemicals could be used to classify AT and TT. HCA and PCA were respectively performed using IBM SPSS Statistics software (Version 22, SPSS Inc., Chicago, IL, USA). An analysis of variance (ANOVA) was also used by IBM SPSS Statistics software (Version 22, SPSS Inc., Chicago, IL, USA) to determine significant differences ($p < 0.05$). Regarding the establishment of discriminant models, stepwise Fisher discriminant analysis (SFDA) was performed based on the model set. The model set consisted of 20 ATs and 40 TTs chosen randomly from 80 samples, and the other samples (10 ATs and 10 TTs) were the training set to check the accuracy of the model by the leave-one-out method (LOO). Based on discriminant variables from SFDA, DTA was further used to detect the classification of AT and TT, thereby establishing the model. The optimal model of AT and TT was selected by performance indexes and distance measurement. Both SFDA and DTA were applied by IBM SPSS Statistics software (Version 22, SPSS Inc., Chicago, IL, USA), and the evaluation performance indexes including accuracy, precision, recall, and F-score were calculated through the confusion matrix, as shown in Table 2. The formulas were as follows [30]:

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{FN} + \text{FP} + \text{FN}) \tag{1}$$

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \tag{2}$$

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \tag{3}$$

$$\text{F-score} = 2\text{TP}/(2\text{TP} + \text{FP} + \text{FN}) \tag{4}$$

Table 2. Confusion matrix table.

		True Classification [1]	
		Type 1	**Type 2**
Predictive classification	Type 1	TP	TN
	Type 2	FP	FN

[1] TP: true positives, represented the correct numbers of type 1 from the model prediction; TN: true negatives, represented the number of mistakes in type 1 from the model prediction; FP: false positives, represented the correct numbers of type 2 from the model prediction; FN: false negatives, represented the number of mistakes in type 2 from the model prediction.

3. Results and Discussion

3.1. Feasibility of Phytochemicals to Classification of AT and TT

Based on WE, TPC, TFAAs, six catechin components, two purine alkaloids, and fifteen free amino acids, PCA was used to explore whether the AT and TT could be distinguished through phytochemicals in this study. As shown in Table 3, the seven principal components (PCs) had an eigenvalue greater than 1 and explained 80.352% of the cumulative percentage of variance, among which PC1, PC2, and PC3 extracted according to the Kaiser criterion represented 24.542%, 19.348%, and 11.261%, respectively, of the variability of raw materials. Obviously, the first three PCs were the main components [31]. Figure 1a shows the score plots of PC1 versus PC3 and it could be observed that all samples were distributed in three areas, among which 11 TTs (T6, T7, T14, T15, T16, T17, T22, T23, T32, T46, and T48) were distributed in an area exhibiting positive scores of PC1, and other samples with cross distribution in the remaining two areas showed negative scores of PC1. Further combined with 3D score plots of PC1 versus PC2 versus PC3 (Figure 1b), it was found that 30 ATs were distributed around with TTs as the center, among which the 12 ATs (A29, A30, A13, A24, A27, A22, A28, A26, A21, A23, A20, and A25) in the left-upper corner showed negative scores of PC2 and positive scores of PC3; five ATs (A17, A18, A15, A16, and A19) in the right-upper corner showed positive scores of PC2 and PC3; and 13 ATs (A14, A11, A9, A3, A10, A12, A1, A6, A4, A7, A2, A8, and A5) in the right-lower corner showed positive scores of PC2 and negative scores of PC3.

Table 3. Eigenvalue and variance contribution rates of PCA in AT and TT.

PCs	Eigenvalue	Percentage of Variance (%)	Cumulative Percentage of Variance (%)
PC1	6.381	24.542	24.542
PC2	5.030	19.348	43.889
PC3	2.928	11.261	55.151
PC4	2.534	9.745	64.896
PC5	1.745	6.711	71.607
PC6	1.205	4.635	76.242
PC7	1.069	4.110	80.352
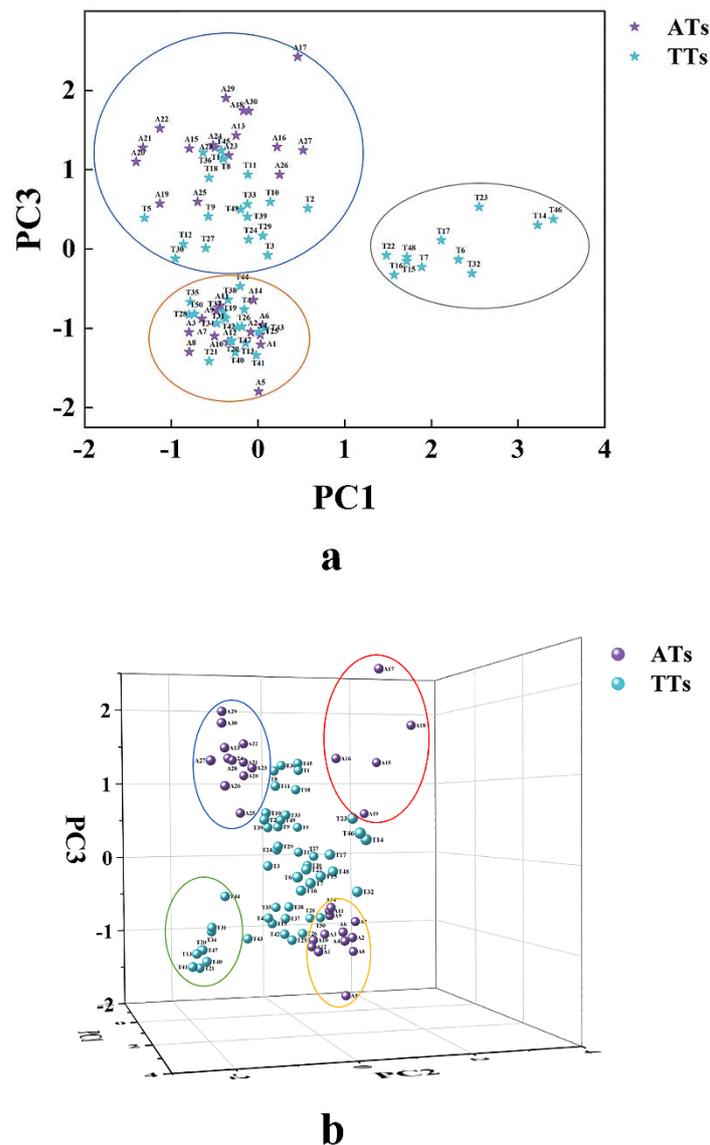
Figure 1. The score plots from the PCA of the phytochemicals in the ancient tea (AT, purple) and terrace tea (TT, cyan). (**a**) PC1 vs. PC3; (**b**) PC1 vs. PC2 vs. PC3. (**a**) had three clusters (blue, orange, and gray circle), among which the gray circle only contained TT. The blue, red, and orange circles only contained AT and the green circle only contained TT in (**b**).

Meanwhile, HCA also showed that when the square Euclidean distance was set at 14, all samples could be divided into eight clusters (clusters-A, B, C, D, E, F, G, and H) (Figure 2). 30 ATs were distributed in clusters-A, D, and H, corresponding to the sample distribution in the blue, orange, and red circles in Figure 1b, respectively. However, A19 was distributed in the red circle (cluster-H) in Figure 1b in the PCA while it was classified into cluster-D in the HCA. The reason might be that the new variables generated in the dimensionality reduction process of PCA produced different results. The clusters-B, C, E, F, and G included seven, twenty-one, two, nine, and eleven TTs, respectively, among which clusters-F and G corresponded to the sample distribution in the gray and green circles in Figure 1a,b, respectively; while other samples in Figure 1b corresponded to clusters-B, C, and E. In short, PCA and HCA can be used to explain whether teas could be distinguished according to season, year, production place, processing technologies, and category variations [32–35]. In this study, the results revealed that phytochemicals were the dominating influence factors in the classification of AT and TT, which could be used as identification indicators of both.
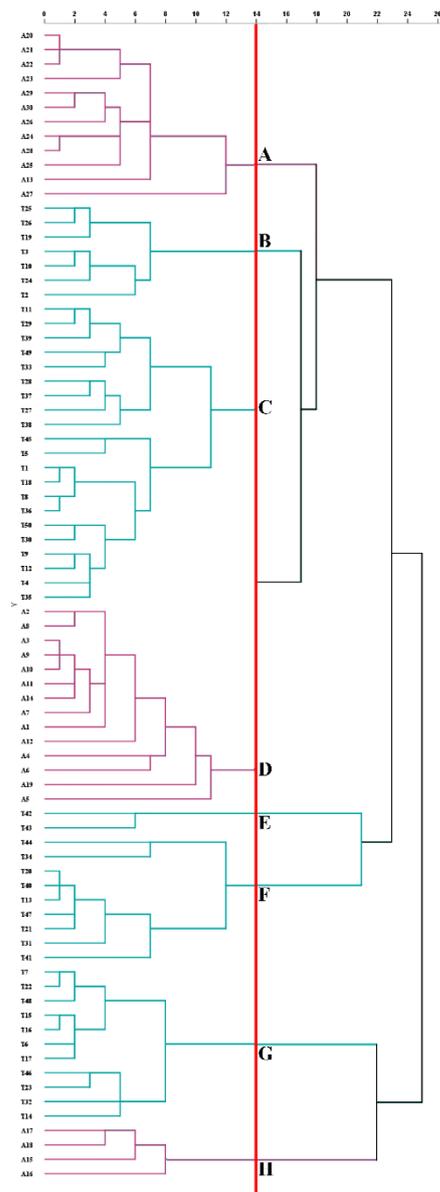
Figure 2. Cluster results on phytochemicals of ancient tea (AT, purple) and terrace tea (TT, cyan). The red line illustrates that samples were divided into eight clusters when the Euclidean distance squared was 14 including clusters-A, B, C, D, E, F, G, and H. Thirty ATs were distributed in clusters-A, D, and H and 50 TTs were distributed in clusters-B, C, E, F, G, and H.

3.2. Comparison of Phytochemicals Differences Related with a Classification of AT and TT

The phytochemical contents related to the classification of AT and TT are presented in Table S1 (Supplementary Materials), and the differences were analyzed based on the average levels of both. As shown in Figure 3a, the average levels of WE and TFAAs in AT were significantly ($p > 0.05$) higher than those in TT, whereas there was no significant difference in the average level of TPC between AT and TT. In terms of catechin components (Figure 3b), the average level of EC in AT was up to 34.3 mg/g, which was significantly higher than that in TT, while the contents of EGC, C, and GCG in AT were significantly lower than that in TT. Theacrine, as a kind of purine alkaloid, had a similar chemical structure to caffeine. It is the key component of Yunnan Kucha (also called bitter Pu-erh tea or Pu-erh Kucha tea), which not only has a beneficial effect on the human body but also a more bitter taste than caffeine [36,37]. As presented in Figure 3b, the level of theacrine in AT was significantly higher than that of TT, which indicated that the bitterness of AT

was more remarkable than that of TT. Compared with the content of theacrine, the level of caffeine of AT and TT reached 22.6 and 61.6 folds, respectively, among which the content of caffeine in AT was significantly higher than that in TT. Additionally, free amino acids of AT and TT were dominant with theanine and Glu, according to Figure 3c. Theanine in AT and TT was up to 12.0 mg/g and 7.96 mg/g, respectively, and there was a significant difference between both; Glu in AT and TT had no significant difference. Asp and Ser were detected only in TT. Those findings were similar to those of Zhang [38] but opposite to the results on the levels of WE and TFAAs in AT and TT as reported by Liang et al. [6]. The disagreement might be due to the influence of processing technology.
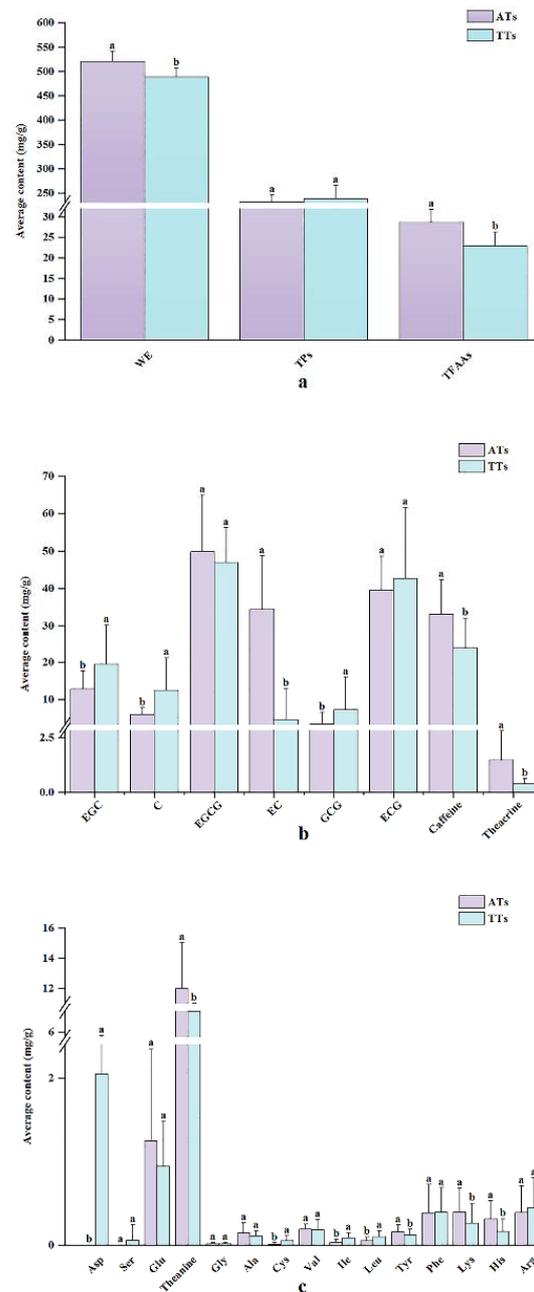
Figure 3. The average levels of the main phytochemical components of ancient tea (AT, purple) and terrace tea (TT, cyan) from Yunnan Province. (**a**) Mixtures; (**b**) catechin components and alkaloids; (**c**) free amino acids. The height of the error bar is ±standard deviation. Different letters with different superscripts on the same index represent significant differences (*p* < 0.05).

Overall, growth environments and management methods had definite impacts on phytochemicals, resulting in significant differences in WE, TFAAs, four catechin components, two purine alkaloids, and eight free amino acids between AT and TT. Previous studies have also indicated that chemical fertilizers and environment could affect the contents of polyphenols, WE, caffeine, and free amino acids [7,39–41]. This indicates that the above differential components of AT and TT were representative, which could be considered as major components of classifying AT and TT in this study.

3.3. Establishment and Optimization of Discriminant Model of AT and TT

3.3.1. Establishing a Discriminant Model through SFDA

SFDA is a popular recognition method by dimensionality reduction based on variance analysis and is one of the most effective methods for feature extraction [42]. To scientifically classify AT and TT, this study extracted key identification indicators by establishing a discriminant model based on 16 phytochemicals. Twenty ATs and forty TTs selected from all samples formed a model set and then the feature components of the model set were extracted, thereby establishing a discriminant model. The results indicated that there were eight phytochemicals that significantly ($p < 0.05$) affected the discriminant effect, and the order of influence degree was as follows, according to F value: EC > C > theanine > WE > EGC > theacrine > Ala > Arg (Table 4). However, due to the lack of detection of Ala and Arg in some samples, which could affect the accuracy of identification between AT and TT, six phytochemicals were finally used to establish a discriminant model in this study. The model is shown in Equation (5): the unknown sample can be determined as AT if Y > 0; otherwise, it was judged to be TT.

$$Y = -21.685 + 0.038 \times WE - 0.094 \times C + 0.12 \times EC - 0.042 \times EGC + 0.371 \times \text{theacrine} + 0.217 \times \text{theanine} \quad (5)$$

Table 4. Results and significance of variables affecting AT and TT discriminations by SFDA extraction.

Numbers	Variables	Statistics	df1	df2	Significance
1	EC	117.094	1	58	1.54×10^{-15}
2	C	118.649	2	57	4.81×10^{-21}
3	Theanine	106.949	3	56	3.67×10^{-23}
4	WE	94.932	4	55	5.10×10^{-24}
5	EGC	88.306	5	54	9.73×10^{-25}
6	Theacrine	83.408	6	53	3.24×10^{-25}
7	Ala	77.816	7	52	2.74×10^{-25}
8	Arg	80.700	8	51	3.10×10^{-26}

3.3.2. Optimizing Classification Model through DTA

To reflect the classification of samples more intuitively, DTA was used to detect the classification of 80 samples based upon six variables of the discriminant model. As shown in Figure 4, the tree involved a three-level structure with a total of eight decision nodes and four classification rules created by only using three elements. Among these, the decision rules were based on two concentration ranges of EC, four concentration ranges of C, and two concentration ranges of WE, which correctly identified 30 ATs and 50 TTs. According to the DTA results, the discriminant model of AT and TT was optimized. As presented in Equation (6), the unknown sample could be determined as AT if Y > −1; on the contrary, it was judged to be TT.

$$Y = -21.685 + 0.038 \times WE - 0.094 \times C + 0.12 \times EC \quad (6)$$
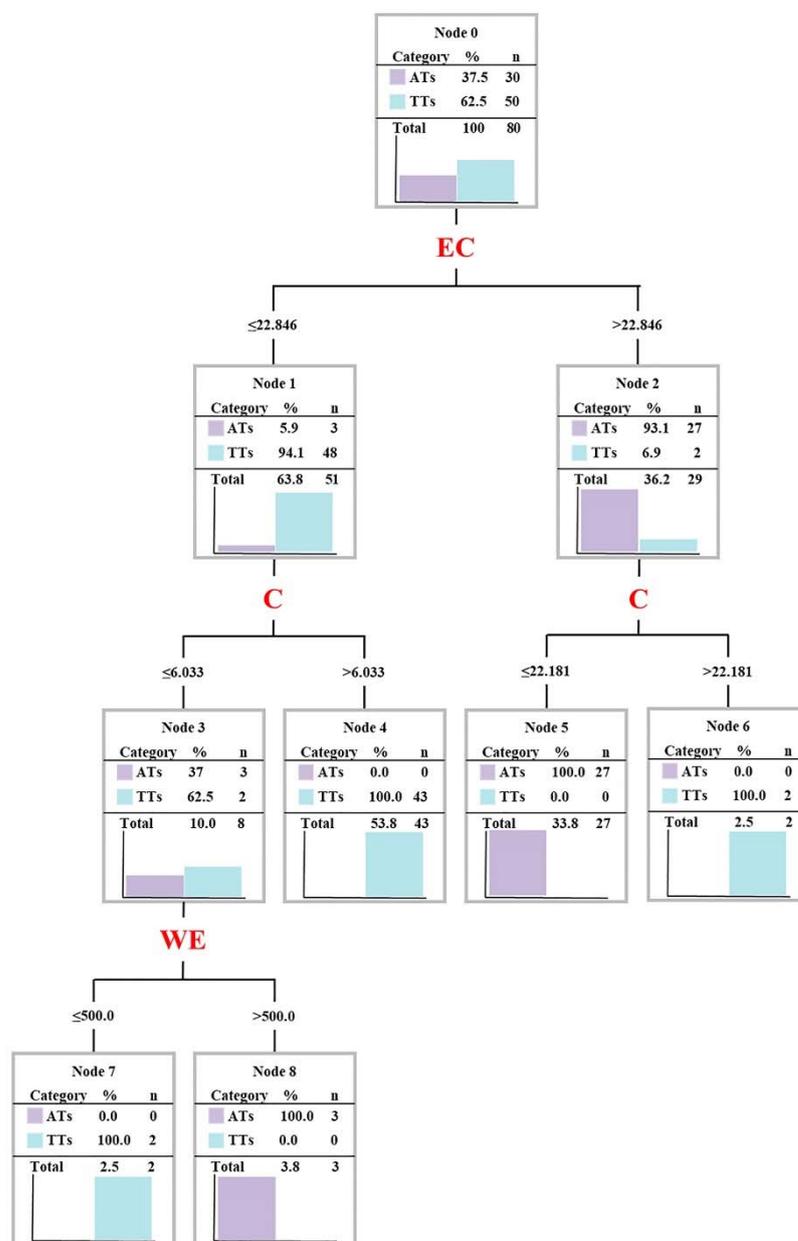
Figure 4. Decision classification tree resulting from the application of the DTA algorithm for the classification of ancient tea (AT) and terrace tea (TT) using the element profile. Concentrations are reported in mg/g.

3.3.3. Evaluating Classification Models

Upon the above two models, this study tried to evaluate the model externally and internally through a training set (10 ATs and 10 TT, *n* = 20) and model set (20 ATs and 40 TTs, *n* = 60) to explore the generalization capability of two discriminant models and choose the optimal model. As shown in Table 5, there were no misjudgments in the training set and model set by model calculation and the leave-one-out method (LOO), regardless of Equation (1) or Equation (2), that is to say, the accuracy, precision, recall, and F-score reached 100%. This suggested that the two discriminant models had a good generalization capability [43]. However, the above performance indexes were unable to help us choose the most suitable discriminant model of AT and TT. The distance measurement could be used to evaluate the separation effect of a model by a minimum distance of two categories: The larger the distance, the easier the classification and a lower error rate [44]. According to Figure 5a,b, the separation effect of AT and TT based on Equation (1) was better than

that based on Equation (2) (1.60 > 1.11). This indicates that Equation (1), established by six components, was optimal, which could be used as the discriminant model of AT and TT in this study.

Table 5. Confusion matrix obtained for the discriminant model AT and TT and evaluation of the model performance indexes.

Verification Modes			True Classification		Performance Index (%)			
			ATs	TTs	Accuracy	Precision	Recall	F-Score
Predictive classification	Training set	ATs	10	0	100	100	100	100
		TTs	0	10				
	Model set (LOO)	ATs	20	0	100	100	100	100
		TTs	0	40				
	DTA	ATs	30	0	100	100	100	100
		TTs	0	50				
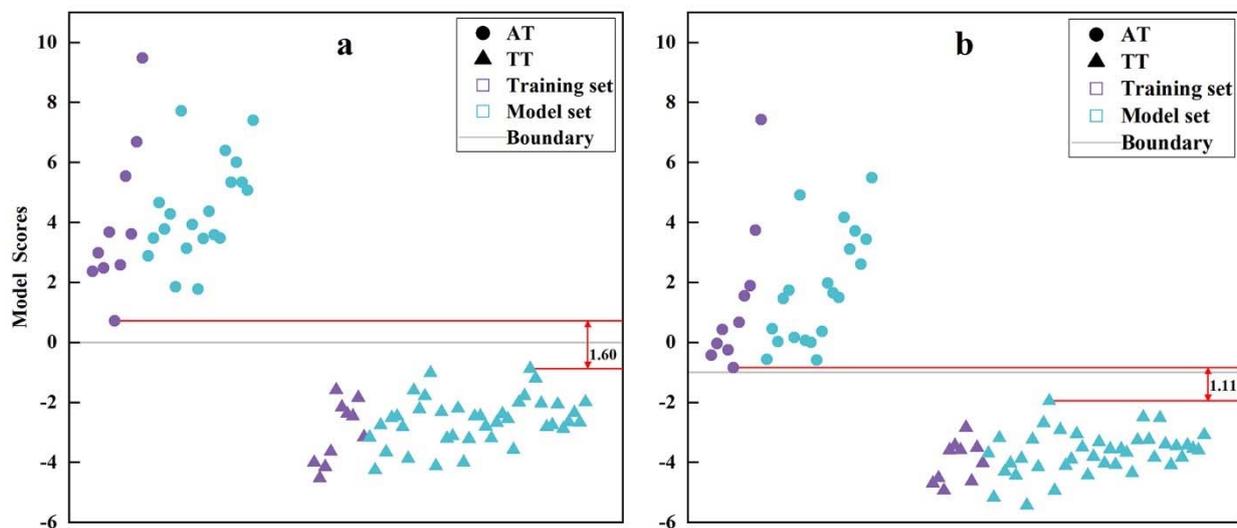
Figure 5. Model scores of the training set and model set calculated through classification model. (**a**) The classification model established through stepwise Fisher discriminant analysis (SFDA). (**b**) The classification model established through decision tree analysis (DTA). The circle represents AT and the triangle represents TT. The purple represents the training set and the cyan represents the model set. The gray line represents the dividing line of AT and TT. The interval between two red lines represents the minimum separation distance of AT and TT ((**a**), 1.60; (**b**), 1.11).

Regarding the six discriminant variables, previous studies have shown that WE is made up of soluble substances such as phenolics and alkaloids, reflecting the thickness of tea infusion [45]. Theanine and nonester catechins (EC, C, and EGC) were related to umami/sweetness and bitterness/astringency, respectively, which were significantly affected by some factors such as geographical environment, light, cultivar, and fertilizer [46,47]. For instance, the climate had an impact on the chlorophyll contents, thereby regulating the contents of nonester catechins; the contents of catechins were higher in northern areas while the contents of free amino acids were higher in the southeast. Theacrine significantly affected the bitter taste of tea infusion and had many excellent pharmacological effects such as sedative and hypnotic [48–50]. At present, our laboratory has already published relevant studies on the recognition threshold and leaching rule of theacrine [51]. The results showed that the taste of AT was more bitter than that of TT due to its high level of theacrine at the same recognition threshold. Meanwhile, the slow leaching rate of theacrine contributed to the endurance property in brewing AT. Overall, the six components could well reflect the differences between AT and TT, and an in-depth study

on flavor contribution will be conducted based on these in the future, thereby providing a reference for regulating the sale of AT.

4. Conclusions

All in all, according to the 26 phytochemicals determined by spectrophotometry methods and HPLC, the PCA and HCA could divide 80 samples into AT and TT, and ANOVA showed the growth environment and management method caused the significant ($p < 0.05$) differences of the 16 phytochemicals as the principal factor including WE, TFAAs, four catechin components, two purine alkaloids, and eight free amino acids. Based on the ANOVA results, the discriminant model of AT and TT was eventually established based on six components including WE, EC, C, EGC, theacrine, and theanine by comparing the separation effect of SFDA and DTA. The accuracy, precision, recall, and F-score of the model were up to 100%, which illustrates the good generalization capability of the discriminant model. This study offers data support for Pu-erh tea from different raw materials from the perspective of phytochemical components and an analytical thinking of classification, which achieved a great effect. In future production practice, the classification method could be applied to classify and distinguish unknown samples. In addition, the chemometric analyses will also be a powerful tool in food fraud such as tea origin, storage time, and organic tea.

References

1. Zhang, Z.; He, F.; Yang, W.; Yang, L.; Huang, S.; Mao, H.; Hou, Y.; Xiao, R. Pu-erh tea extraction alleviates intestinal inflammation in mice with flora disorder by regulating gut microbiota. *Food Sci. Nutr.* **2021**, *9*, 4883–4892. [CrossRef] [PubMed]
2. Ahmed, S.; Unachukwu, U.; Stepp, J.R.; Peters, C.M.; Long, C.; Kennelly, E. Pu-erh tea tasting in Yunnan, China: Correlation of drinkers' perceptions to phytochemistry. *J. Ethnopharmacol.* **2010**, *132*, 176–185. [CrossRef] [PubMed]
3. Qi, D.-H.; Guo, H.-J.; Sheng, C.-Y. Assessment of plant species diversity of ancient tea garden communities in Yunnan, Southwest of China. *Agrofor. Syst.* **2012**, *87*, 465–474. [CrossRef]
4. Bao, X.; Dong, X.; Pan, S. A comparative study between wild ancient tea and tea garden tea. *Zhejiang J. Agric. Sci.* **2012**, *24*, 603–608. (In Chinese)
5. Wang, T.; Li, X.; Yang, H.; Wang, F.; Kong, J.; Qiu, D.; Li, Z. Mass spectrometry-based metabolomics and chemometric analysis of Pu-erh teas of various origins. *Food Chem.* **2018**, *268*, 271–278. [CrossRef]

6. Liang, M.; Xia, L.; Zhang, J.; Fang, C.; Chen, J.; Chen, L.; Sun, R. Study on quality comparison of old plant tea and tableland tea. *J. Yunnan Agric. Univ.* **2006**, *4*, 493–497. (In Chinese)

7. Chowaniak, M.; Niemiec, M.; Zhu, Z.; Rashidov, N.; Gródek-Szostak, Z.; Szeląg-Sikora, A.; Sikora, J.; Kuboń, M.; Fayzullo, S.; Mahmadyorzoda, U.; et al. Quality Assessment of Wild and Cultivated Green Tea from Different Regions of China. *Molecules* **2021**, *26*, 3620. [CrossRef]

8. Gao, F.; Jiahua, L.I.; Zhao, P. Research and Analysis of Soil Nutrients Status of the Modern Tea Gardens and Ancient Tea Gardens in Pu'er City. *J. Yunnan Agric. Univ.* **2016**, *31*, 511–518. (In Chinese) [CrossRef]

9. Yang, G.; Wang, X.; Xie, J. Analysis of the Relationship between Soil Nutrients and Tea Main Quality Components of Ancient Tea Arboretum and Modern Tea Garden in Yunnan Province. *J. Tea Sci.* **2015**, *35*, 574–582.

10. Lin, X.; Li, Q.; He, L.; Lan, S.; Lin, T.; Liu, H. Application of Heavy Rare Earth Element Fingerprints in Discrimination of Pu'er Old Plant Tea and Tableland Tea. *Mod. Food Sci. Technol.* **2013**, *29*, 2921–2925.

11. Sarker, S.D.; Nahar, L. Applications of High-Performance Liquid Chromatography in the Analysis of Herbal Products. In *Evidence-Based Validation of Herbal Medicine*; Mukherjee, P.K., Ed.; Elsevier: Boston, MA, USA, 2015; pp. 405–425.

12. dos Santos, P.M.; Filho, E.P.; Rodriguez-Saona, L. Rapid detection and quantification of milk adulteration using infrared microspectroscopy and chemometrics analysis. *Food Chem.* **2013**, *138*, 19–24. [CrossRef] [PubMed]

13. Christou, C.; Agapiou, A.; Kokkinofta, R. Use of FTIR spectroscopy and chemometrics for the classification of carobs origin. *J. Adv. Res.* **2017**, *10*, 1–8. [CrossRef]

14. Tomazzoli, M.M.; Neto, R.D.P.; Moresco, R.; Westphal, L.; Zeggio, A.R.S.; Specht, L.; Costa, C.; Rocha, M.; Maraschin, M. Discrimination of Brazilian propolis according to the seasoning using chemometrics and machine learning based on UV-Vis scanning data. *J. Integr. Bioinform.* **2015**, *12*, 15–26. [CrossRef]

15. Wang, H.; Cao, X.; Yuan, Z.; Guo, G. Untargeted metabolomics coupled with chemometrics approach for Xinyang Maojian green tea with cultivar, elevation and processing variations. *Food Chem.* **2021**, *352*, 129359. [CrossRef] [PubMed]

16. Ku, K.M.; Kim, J.; Park, H.-J.; Liu, K.-H.; Lee, C.H. Application of Metabolomics in the Analysis of Manufacturing Type of Pu-erh Tea and Composition Changes with Different Postfermentation Year. *J. Agric. Food Chem.* **2009**, *58*, 345–352. [CrossRef] [PubMed]

17. Zhou, B.; Ma, C.; Wu, T.; Xu, C.; Wang, J.; Xia, T. Classification of raw Pu-erh teas with different storage time based on characteristic compounds and effect of storage environment. *LWT* **2020**, *133*, 109914. [CrossRef]

18. Zhou, B.; Ma, B.; Ma, C.; Xu, C.; Wang, J.; Wang, Z.; Yin, D.; Xia, T. Classification of Pu-erh ripened teas and their differences in chemical constituents and antioxidant capacity. *LWT* **2021**, *153*, 112370. [CrossRef]

19. Sun, H.; Lu, W.; Gao, B. Non-targeted detection of butter adulteration using pointwise UHPLC-ELSD and UHPLC-UV fingerprints with chemometrics. *Food Chem.* **2021**, *356*, 129604. [CrossRef]

20. Shi, T.; Wu, G.; Jin, Q.; Wang, X. Detection of camellia oil adulteration using chemometrics based on fatty acids GC fingerprints and phytosterols GC–MS fingerprints. *Food Chem.* **2021**, *352*, 129422. [CrossRef]

21. Zheng, Y.-F.; Wu, M.-C.; Chien, H.-J.; Wang, W.-C.; Kuo, C.-Y.; Lai, C.-C. Honey proteomic signatures for the identification of honey adulterated with syrup, producing country, and nectar source using SWATH-MS approach. *Food Chem.* **2021**, *354*, 129590. [CrossRef]

22. Aboulwafa, M.M.; Youssef, F.S.; Gad, H.A.; Sarker, S.D.; Nahar, L.; Al-Azizi, M.M.; Ashour, M.L. Authentication and discrimination of green tea samples using UV–vis, FTIR and HPLC techniques coupled with chemometrics analysis. *J. Pharm. Biomed. Anal.* **2019**, *164*, 653–658. [CrossRef] [PubMed]

23. Yu, J.; Liu, Y.; Zhang, S.; Luo, L.; Zeng, L. Effect of brewing conditions on phytochemicals and sensory profiles of black tea infusions: A primary study on the effects of geraniol and β-ionone on taste perception of black tea infusions. *Food Chem.* **2021**, *354*, 129504. [CrossRef] [PubMed]

24. Wang, Z.-H.; Yue, C.-N.; Tong, H.-R. Analysis of taste characteristics and identification of key chemical components of fifteen Chinese yellow tea samples. *J. Food Sci. Technol.* **2020**, *58*, 1378–1388. [CrossRef]

25. Astill, C.; Birch, M.R.; Dacombe, C.; Humphrey, P.G.; Martin, P.T. Factors Affecting the Caffeine and Polyphenol Contents of Black and Green Tea Infusions. *J. Agric. Food Chem.* **2001**, *49*, 5340–5347. [CrossRef]

26. Magné, C.; Larher, F. High sugar content of extracts interferes with colorimetric determination of amino acids and free proline. *Anal. Biochem.* **1992**, *200*, 115–118. [CrossRef]

27. Liu, Y.; Luo, L.; Liao, C.; Chen, L.; Wang, J.; Zeng, L. Effects of brewing conditions on the phytochemical composition, sensory qualities and antioxidant activity of green tea infusion: A study using response surface methodology. *Food Chem.* **2018**, *269*, 24–34. [CrossRef]

28. Li, H.; Fang, K.; Qin, D.; Jiang, X.; Wang, Q.; Pan, C.; Li, B.; Wang, Q.; Wu, H. Comparative Transcriptome Analysis Reveals Putative Genes Responsible for High Theacrine Content in Kucha (Camellia Kucha (Chang et Wang) Chang). *Trop. Plant Biol.* **2021**, *14*, 82–92. [CrossRef]

29. Lu, M.; Han, J.; Zhu, B.; Jia, H.; Yang, T.; Wang, R.; Deng, W.-W.; Zhang, Z.-Z. Significantly increased amino acid accumulation in a novel albino branch of the tea plant (*Camellia sinensis*). *Planta* **2018**, *249*, 363–376. [CrossRef]

30. Rodríguez, L.C.; Pérez-Castaño, E.; Ruiz-Samblás, C. Quality performance metrics in multivariate classification methods for qualitative analysis. *TrAC Trends Anal. Chem.* **2016**, *80*, 612–624. [CrossRef]

31. Yeung, K.Y.; Ruzzo, W.L. Principal component analysis for clustering gene expression data. *Bioinformatics* **2001**, *17*, 763–774. [CrossRef]

32. Zhou, P.; Hu, O.; Fu, H.; Ouyang, L.; Gong, X.; Meng, P.; Wang, Z.; Dai, M.; Guo, X.; Wang, Y. UPLC–Q-TOF/MS-based untargeted metabolomics coupled with chemometrics approach for Tieguanyin tea with seasonal and year variations. *Food Chem.* **2019**, *283*, 73–82. [CrossRef] [PubMed]

33. Wang, Z.; Ma, B.; Ma, C.; Zheng, C.; Zhou, B.; Guo, G.; Xia, T. Region identification of Xinyang Maojian tea using UHPLC-Q-TOF/MS-based metabolomics coupled with multivariate statistical analyses. *J. Food Sci.* **2021**, *86*, 1681–1691. [CrossRef]

34. Wang, Y.; Kan, Z.; Thompson, H.J.; Ling, T.-J.; Ho, C.-T.; Li, D.; Wan, X. Impact of Six Typical Processing Methods on the Chemical Composition of Tea Leaves Using a Single *Camellia sinensis* Cultivar, Longjing 43. *J. Agric. Food Chem.* **2018**, *67*, 5423–5436. [CrossRef]

35. Zhou, Y.; Yu, J.; Wang, L.; Wu, D.; Qi, D.; Sha, Y.; Liu, B. Identification of Volatile Components in Tea Infusions by Headspace–Programmed Temperature Vaporization–Gas Chromatography–Mass Spectrometry (HS–PTV–GC–MS) with Chemometrics. *Anal. Lett.* **2020**, *53*, 1902–1918. [CrossRef]

36. Ouyang, S.-H.; Zhai, Y.-J.; Wu, Y.-P.; Xie, G.; Wang, G.-E.; Mao, Z.-F.; Hu, H.-H.; Luo, X.-H.; Sun, W.-Y.; Liang, L.; et al. Theacrine, a Potent Antidepressant Purine Alkaloid from a Special Chinese Tea, Promotes Adult Hippocampal Neurogenesis in Stressed Mice. *J. Agric. Food Chem.* **2021**, *69*, 7016–7027. [CrossRef] [PubMed]

37. Lin, P.-R.; Kuo, P.-C.; Li, Y.-C.; Jhuo, C.-F.; Hsu, W.-L.; Tzen, J.T. Theacrine and strictinin, two major ingredients for the anti-influenza activity of Yunnan Kucha tea. *J. Ethnopharmacol.* **2020**, *262*, 113190. [CrossRef] [PubMed]

38. Zhang, Q. Quickly distinguishing ancient tea from terrace tea. *Mod. Agric.* **2018**, *8*, 99. (In Chinese)

39. Gao, S.-L.; Hu, S.-S.; He, P.; Feng, K.; Pan, R.-Y.; Zhang, S.; Guo, B.; Lee, T.-C.; Lin, J.-K. Effects of reducing chemical fertilizer on the quality components of Tieguanyin tea leaves. *IOP Conf. Ser. Earth Environ. Sci.* **2020**, *559*. [CrossRef]

40. Yang, T.; Li, H.; Hu, X.; Li, J.; Hu, J.; Liu, R.; Deng, Z.-Y. Effects of Fertilizing with N, P, Se, and Zn on Regulating the Element and Functional Component Contents and Antioxidant Activity of Tea Leaves Planted in Red Soil. *J. Agric. Food Chem.* **2014**, *62*, 3823–3830. [CrossRef]

41. Huang, S.; Zuo, T.; Zheng, X.; Zhuo, C.; Hou, Q.; Yao, L.; Wang, X.; Wang, J.; Ni, W. Foliar application of glycinebetaine and Zn fertilizer improves both the apparent and functional qualities of albino tea [*Camellia sinensis* (L.) O. Kuntze]. *Food Funct.* **2021**. [CrossRef]

42. Huang, H.; Li, J.; Liu, J. Enhanced semi-supervised local Fisher discriminant analysis for face recognition. *Futur. Gener. Comput. Syst.* **2012**, *28*, 244–253. [CrossRef]

43. Landis, J.R.; Koch, G.G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, *33*, 159–174. [CrossRef] [PubMed]

44. Li, L.; Ma, X. Study on specificity of colon carcinoma-associated serum markers and establishment of SVM prediction model. *Saudi J. Biol. Sci.* **2017**, *24*, 644–648. [CrossRef] [PubMed]

45. Yao, L.; Liu, X.; Jiang, Y.; Caffin, N.; D'Arcy, B.; Singanusong, R.; Xu, Y. Compositional analysis of teas from Australian su-permarkets. *Food Chem.* **2006**, *94*, 115–122. [CrossRef]

46. Wei, K.; Wang, L.; Zhou, J.; He, W.; Zeng, J.; Jiang, Y.; Cheng, H. Catechin contents in tea (*Camellia sinensis*) as affected by cultivar and environment and their relation to chlorophyll contents. *Food Chem.* **2011**, *125*, 44–48. [CrossRef]

47. Wen, B.; Ren, S.; Zhang, Y.; Duan, Y.; Shen, J.; Zhu, X.; Wang, Y.; Ma, Y.; Zou, Z.; Fang, W. Effects of geographic locations and topographical factors on secondary metabolites distribution in green tea at a regional scale. *Food Control.* **2019**, *110*, 106979. [CrossRef]

48. Qin, D.; Wang, Q.; Li, H.; Jiang, X.; Fang, K.; Wang, Q.; Li, B.; Pan, C.; Wu, H. Identification of key metabolites based on non-targeted metabolomics and chemometrics analyses provides insights into bitterness in Kucha [Camellia Kucha (Chang et Wang) Chang]. *Food Res. Int.* **2020**, *138*, 109789. [CrossRef]

49. Xu, J.-K.; Kurihara, H.; Zhao, L.; Yao, X.-S. Theacrine, a special purine alkaloid with sedative and hypnotic properties from Cammelia assamica var. kucha in mice. *J. Asian Nat. Prod. Res.* **2007**, *9*, 665–672. [CrossRef]

50. Wang, Y.; Yang, X.; Zheng, X.; Li, J.; Ye, C.; Song, X. Theacrine, a purine alkaloid with anti-inflammatory and analgesic ac-tivities. *Fitoterapia* **2010**, *81*, 627–631. [CrossRef]

51. Shi, Y.; Zhang, S.-R.; Sun, K.; Wang, X.-H.; Jiang, J.-L.; Luo, L.-Y.; Zeng, L. Characterization of bitter taste theacrine in Pu-erh tea. *J. Food Compos. Anal.* **2021**, *106*, 104331. [CrossRef]