Discriminant Analysis of the Geographical Origin of Asian Red Pepper Powders Using Second-Derivative FT-IR Spectroscopy

Miso Kim [1,†], **Junyoung Hong** [1,†], **Dongwon Lee** [1], **Sohyun Kim** [1], **Hyang Sook Chun** [2], **Yoon-Ho Cho** [3], **Byung Hee Kim** [4] and **Sangdoo Ahn** [1,*]

1. Department of Chemistry, Chung-Ang University, Seoul 06974, Korea; rlaalth1328@naver.com (M.K.); hjuny94@hanmail.net (J.H.); idleplanet@naver.com (D.L.); dragon725@naver.com (S.K.)
2. Department of Food Science & Technology, Chung-Ang University, Ansung 17546, Korea; hschun@cau.ac.kr
3. Department of Civil and Environmental Engineering, Chung-Ang University, Seoul 06974, Korea; yhcho@cau.ac.kr
4. Department of Food and Nutrition, Sookmyung Women's University, Seoul 04310, Korea; bhkim@sookmyung.ac.kr
* Correspondence: sangdoo@cau.ac.kr; Tel.: +82-2-820-5230
† These authors contributed equally to this work.

Abstract: This study aimed to discriminate between the geographical origins of Asian red pepper powders distributed in Korea using Fourier-transform infrared (FT-IR) spectroscopy coupled with multivariate statistical analyses. Second-derivative spectral data were obtained from a total of 105 red pepper powder samples, 86 of which were used for statistical analysis, and the remaining 19 were used for blind testing. A one-way analysis of variance (ANOVA) test confirmed that eight peak variables exhibited significant origin-dependent differences, and the canonical discriminant functions derived from these variables were used to correctly classify all the red pepper powder samples based on their origins. The applicability of the canonical discriminant functions was examined by performing a blind test wherein the origins of 19 new red pepper powder samples were correctly classified. For simplicity, the four most significant variables were selected as discriminant indicator variables, and the applicable range for each indicator variable was set for each geographical origin. By applying the indicator variable ranges, the origins of the red pepper powders of all the statistical and blind samples were correctly identified. The study findings indicate the feasibility of using FT-IR spectroscopy in combination with multivariate analysis for identifying the geographical origins of red pepper powders.

Keywords: fourier-transform infrared (FT-IR) spectroscopy; second-derivative spectrum; red pepper powder; geographical origin; discriminant analysis

1. Introduction

Red peppers (*Capsicum annuum L.*) are perennial plants of the family Solanaceae and are widely grown worldwide. The capsaicinoids specifically contained in red peppers are pungent alkaloids and are known to promote energy metabolism [1]. In addition, carotenoids and vitamin C, which are abundant in red peppers, have been reported to have anti-cancer effects [2,3]. Red peppers are mainly used for their hot spicy flavor and red color. They are predominantly processed into a dried powder form for easy transport to markets worldwide. The quality and cost of red pepper powders vary considerably depending on their country of origin. For instance, the quality of imported peppers is reduced owing to freezing or other pretreatment processes [4]. Consequently, consumers typically prefer domestic products [5]. In some cases, retailers deceive consumers by omitting the country of origin of the red pepper powders to inflate their margins [6]. Therefore, it is necessary to develop an accurate and rapid method for identifying the origin of red pepper powders.

Several factors contribute to the differences between plants of different geographical origins [7,8]. Each country has a different crop cultivation environment, such as soil

and climate, which lead to differences in the metabolite compositions of plants. Therefore, many studies have been conducted to discriminate foods and agricultural products based on their geographical origins by analyzing their metabolite profiles using mass spectrometry (MS) [9], nuclear magnetic resonance (NMR) spectroscopy [10,11], and chromatography [12]. Recently, simple non-destructive Fourier-transform infrared (FT-IR) spectroscopic techniques have been used for metabolite research [13,14]. The FT-IR method has been applied in combination with multivariate statistical analysis for the geographical discrimination of Korean and Chinese soybeans [15], European saffron [16], European olive oils [17], and European and South American honey [18]. FT-IR, which provides a variety of information on chemical bonds and functional groups, involves the use of inexpensive equipment and does not require the special pretreatment of samples regardless of their state. Diverse FT-IR information has also been used to study the adulteration and authenticity of foods [19]. Moreover, recent studies have combined FT-IR with other instrumental analysis methods to improve the efficiency of food analysis [20,21]; in particular, multivariate statistical analyses have been applied to effectively process instrumental analytical data, requiring the observation of minute differences in the compositions or spectral profiles of various metabolites [22].

The geographical classification of Asian red peppers has also been studied [23–26]. Yin et al. proposed a simple method to determine the geographical origins of red peppers from various regions in China through multivariate analysis of sensory characteristics, such as color, taste, and smell [24]. Zhang et al. conducted a study to effectively distinguish the regions of origin of red peppers in China by analyzing the multi-elements obtained using the inductively coupled plasma (ICP) methods with various chemometric tools [25]. Song et al. demonstrated the possibility of discriminating the geographical origins of Korean red peppers through multivariate statistical analysis of the stable isotope ratio ($^{87}Sr/^{86}Sr$) affected by cultivation environments [26]. Recently, it has been demonstrated that ^1H NMR spectroscopy, combined with multivariate statistical analyses, has a significant predictive potential for identifying the geographical origins of Asian red pepper powders [23]. As is well known, NMR spectroscopy is an extremely beneficial analytical method for metabolite studies because NMR measurements are highly reproducible and can provide both qualitative and quantitative information simultaneously. However, NMR spectroscopy incurs a high instrumental cost and requires pretreatment processes to extract the metabolite components from solid samples, such as red pepper powders.

In this study, FT-IR spectroscopy was used to develop an alternative, simple, and convenient experimental method combined with multivariate statistical analysis for determining the geographical origins of red pepper powders from Korea, China, and Vietnam. The second derivative of the FT-IR spectrum was used to enhance the resolution of broadly overlapping peaks and improve peak quantification by removing baseline errors [27–30]. A one-way analysis of variance (ANOVA) test was used to identify whether the peak variables were significantly different depending on their origins [31]. Canonical discriminant analysis, as a multivariate statistical analysis, was used to obtain the discriminant functions for classifying the geographical origins of red pepper powder samples [32]. The obtained canonical discriminant functions represent linear combinations of the meaningful variables identified by the ANOVA test, and exhibit the highest possible multiple correlations with the origins. The applicability of the discriminant functions was verified through a blind test. Additionally, to easily discriminate new red pepper samples based on their geographical origins without complicated statistical processes, several variables having significant influences on the discrimination were selected as the discriminant indicator variables, and their applicable ranges were set for each geographical origin. The feasibility of this method was also verified by using it to correctly identify the origins of the blind samples. When appropriate discriminant indicator variables and ranges are set, the application becomes very simple. Thus, the proposed method may be considered as a more effective and convenient method for discriminating new samples in comparison with other discrimination methods that require statistical processing.

2. Materials and Methods

2.1. Red Pepper Powder Samples

A total of 105 Asian (Korean, Chinese, and Vietnamese) red pepper powders (or dried red peppers) distributed in Korea were collected as samples. Korean red pepper powders were obtained from local producers or reliable suppliers, such as agricultural cooperatives. Chinese and Vietnamese red pepper powders imported to Korea, through the Korea Agro-Fisheries Trade Corporation, were purchased from local markets. Among the total samples, 86 red pepper powders (Korean = 50, Chinese = 23, Vietnamese = 13) were used for statistical analyses to establish the discriminant functions and indicator variables, which could be used to distinguish their geographical origins. The remaining 19 red pepper powders (Korean = 9, Chinese = 5, Vietnamese = 5) were used as blind test samples to verify the applicability of the established discriminant functions and indicator variables. All the dried red pepper powder samples were stored in a refrigerator at 4 °C. However, commonly sold red pepper powders are mixtures of peel and seed fragments with a length of 1–3 mm, making it difficult to reflect all the component information in the FT-IR spectra and ensure reproducibility of measurements. Therefore, the powder was further ground into a fine powder (with particle diameters of $\leq$ 200 μm) in a food grinder before measurement. The prepared fine powder samples exhibited good reproducibility in repeated measurements. Three Korean samples purchased in the form of dried red pepper were first ground into powders using a crusher, and then further ground into finer powders as in the other samples.

2.2. FT-IR Measurement

Finely ground red pepper powder samples were loaded onto an FT-IR spectrometer (TENSOR-27; Bruker Optics GmbH, Karlsruhe, Germany) equipped with a diamond attenuated total reflectance (ATR) accessory (A225/Q Platinum ATR; Bruker Optics GmbH, Karlsruhe, Germany). All the spectra were acquired in absorbance mode, in the wavenumber range of 4000–400 cm^{-1}, with 32 repeated scans and a resolution of 4 cm^{-1}. The acquisition time was less than 1 min for each measurement. To ensure the representativeness and reproducibility of the obtained FT-IR spectra, measurements were repeated five times for each sample and statistically averaged. The ATR crystal was cleaned with ethyl alcohol before every measurement. Atmospheric correction was also performed for each measurement to eliminate the effects of CO_2 and H_2O in the atmosphere. OMINC software (version 8.2, ThermoFisher Scientific Inc. Waltham, MA, USA) was used to process the obtained spectra of the red pepper powder samples. Second derivatives of the processed FT-IR absorbance spectra were derived using the Savitzky–Golay (SG) numerical algorithm with third-order polynomials at seven smoothing points [33–35]. The SG method is a commonly used filtering and smoothing technique to remove background effects and any possible noise in the spectrum during second-order differentiation. Through the differentiation process, the sensitivity and resolution of the spectrum were improved by correcting the baseline drift and separating the overlapped peaks [33–35]. The normalized value of each peak in the second-derivative spectra was used for statistical analyses to establish the origin discriminant functions and indicator variables.

2.3. Multivariate Statistical Analysis

Statistical analyses for the second-derivative FT-IR spectral data were performed using IBM SPSS Statistics software (version 26, SPSS Inc., Chicago, IL, USA). Tests of homogeneity of variance were conducted to determine if each peak variable was equally distributed according to the origin group. For the variables with equal variance, a one-way ANOVA test was used to determine significant differences in the peak variables depending on their origin group (significance level, $p < 0.05$). Canonical discriminant analyses were performed with the selected variables to determine the discriminant functions capable of effectively classifying the geographical origins of the red pepper powder samples. Additionally, by selecting several indicator variables that contribute significantly to the discriminant

functions and setting the ranges of their values, we determined whether the geographical origin could be easily identified without the statistical dataset. The applicability of both the discriminant functions and the indicator variables obtained were tested through a blind test [28,36–39].

3. Results and Discussion

3.1. FT-IR Spectrum of Red Pepper Powder

Figure 1 illustrates a representative FT-IR spectrum of a red pepper powder sample and its second-derivative spectrum. Using the second derivative of the FT-IR spectrum, more sophisticated spectral data were obtained, while broadly overlapping peaks in the original absorption spectrum could be isolated. The second-derivative process also improved the peak quantification by removing the baseline errors. As summarized in Table 1, 19 distinguishable peaks were selected and labeled in the second-derivative FT-IR spectrum. The peaks were assigned by referring to previous studies [15,16,28,40–45].
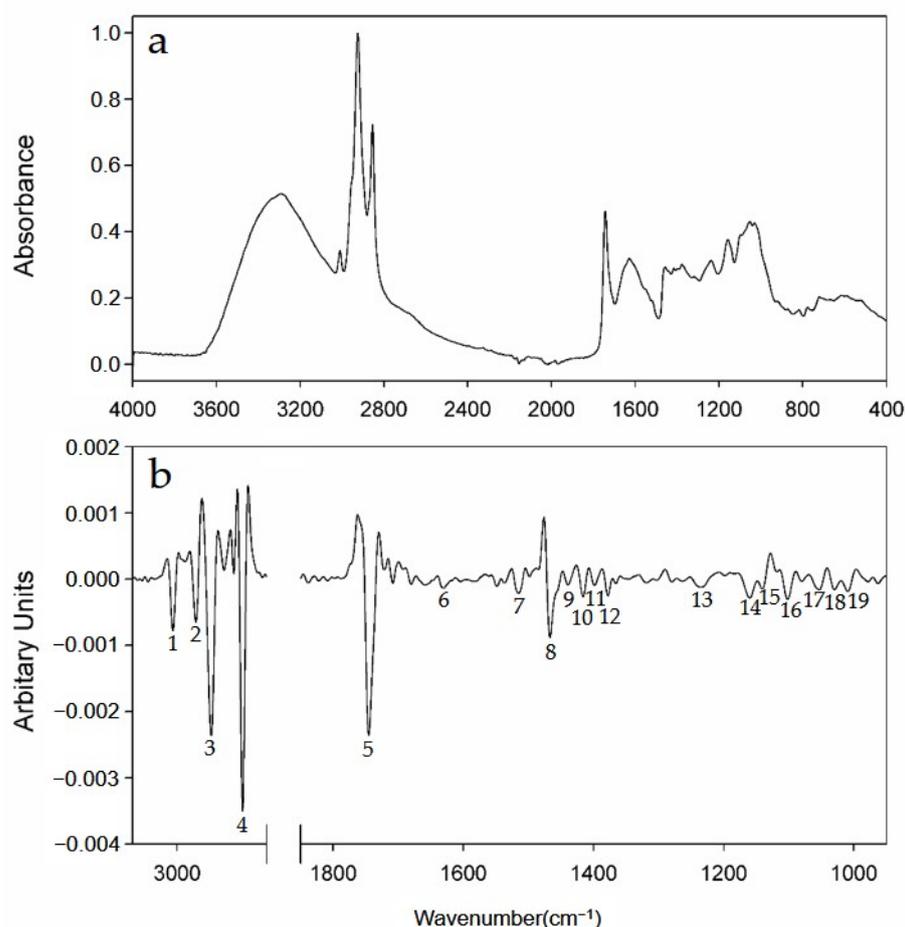
Figure 1. Representative (**a**) FT-IR absorption spectrum of a Korean red pepper powder sample, and (**b**) its second-derivative spectrum (partially expanded at the variables) with numbered peaks (corresponding to the variables in Table 1).

Table 1. Assignment of peaks in the second-derivative FT-IR spectrum of red pepper powders.

Variable	Wavenumber (cm^{-1})	Functional Group	Mode of Vibration
P1	3010	=C–H (*cis*-)	Stretching [40,41]
P2	2958	–C–H (CH$_3$)	Stretching [40,41]
P3	2924	–C–H (*asym* CH$_2$)	Stretching [40,41,44]
P4	2852	–C–H (*sym* CH$_2$)	Stretching [40,41,44]
P5	1745	–C=O (ester)	Stretching [40,41,44]
P6	1653	–C=C– (*cis*-)	Stretching [40,41]
P7	1516	–C–C– (aromatic)	Stretching [42]
P8	1468	–C–H (CH$_2$, CH$_3$)	Bending [40,41,44]
P9	1439	–C–H (CH$_2$)	Bending [45]
P10	1415	=C–H (cis-)	Bending [28,41,42]
P11	1398	–C–H (CH$_2$, CH$_3$)	Bending [15,16,41]
P12	1377	–C–H (CH$_3$)	Bending [16,40,41,44]
P13	1238	–C–O (ester), –C–H (CH$_2$)	Stretching, Bending [41,44]
P14	1159	–C–O (ester)	Stretching [40,41,44,45]
P15	1142	–C–O	Stretching [43,45]
P16	1101	–C–O	Stretching [43,45]
P17	1053	–C–O	Stretching [43,45]
P18	1028	–C–O	Stretching [43,45]
P19	1008	–C–O	Stretching [43,45]

The broad band at approximately 3400 cm^{-1} is mainly due to the stretching of the O–H bonds, because red pepper powder has a low protein content [7] and easily absorbs moisture [46]. The peak at 3010 cm^{-1} (P1) is attributed to sp^2 C–H stretching, while the peaks at 2958 cm^{-1}, 2924 cm^{-1}, and 2852 cm^{-1} (P2, P3, P4) are attributed to the sp^3 C–H stretching of metabolites in the red pepper powders. The strong peak at 1745 cm^{-1} (P5) is due to the C=O stretching, and the weak peak at 1653 cm^{-1} (P6) is due to the C=C stretching. The aromatic C-C stretching band appears at 1516 cm^{-1} (P7), and several C-H bending bands appear at 1516–1238 cm^{-1} (P8–P13). The various C–O stretching bands of the ester and ether groups appear at 1238–1008 cm^{-1} (P3–P19) which are mainly attributed to the lipids and carbohydrates in the red pepper powders [43,45]. The intensities of the peaks differ slightly depending on the distribution of various metabolites in the red pepper powders. Hence, the statistical analysis of this information could be used to discriminate between the geographical origins. For further statistical analysis, the absolute peak values normalized by the intensity of the C–O stretching peak at 1008 cm^{-1} were used.

3.2. Statistical Analysis

3.2.1. Canonical Discriminant Analysis

Canonical discriminant analysis was performed as a multivariate statistical analysis to achieve the most discriminative peak variables for the arrangement of red pepper powder samples in a lower dimensional space by maximizing the distances between the origin groups. To ensure the robustness of these statistical processes, the homogeneity of the variance of each variable must be considered [31]. Therefore, to select suitable variables for the statistical analysis, a variance homogeneity test was conducted first. As a result of testing 18 peaks, it was confirmed that eight peak variables, namely P5, P7, P8, P10, P12, P14, P16, and P17, had equal variance ($p > 0.05$), while the 10 remaining peaks did not exhibit equal variance ($p < 0.05$) (Table S1).

In this study, an ANOVA test was performed to determine the second-derivative FT-IR peak variables with meaningful differences among the Korean, Chinese, and Vietnamese red pepper powder groups. The ANOVA test verified the equality of the group means of variables using the *F* test, and determined whether the means of three or more groups were different [31]. Since the ANOVA test is a parametric test, only the eight peaks with equal variance identified in the previous test of homogeneity of variance were considered [31]. All the eight peak variables exhibited significant differences in the origins ($p < 0.001$) with

large *F* values (Table 2). As can be seen in Table 2, a smaller Wilks' lambda value (i.e., a larger *F*-value) implies a higher significance in the discrimination analysis.

Table 2. Tests of equality of group means.

Variable	Wilks' Lambda	*F*	df_1	df_2	Significance Level, *p*
P05	0.480	44.881	2	83	<0.001
P07	0.776	11.946	2	83	<0.001
P08	0.401	61.900	2	83	<0.001
P10	0.550	34.005	2	83	<0.001
P12	0.264	115.938	2	83	<0.001
P14	0.383	66.762	2	83	<0.001
P16	0.623	25.106	2	83	<0.001
P17	0.246	127.075	2	83	<0.001

These eight significant variables were used for the canonical discriminant analysis to establish the discriminant functions. Two canonical discriminant functions were derived for identifying the red pepper powder samples from different origins, and accounted for 100% of the variance. Functions 1 and 2 accounted for 65.2% and 34.8% of the total variance, respectively. The separation between the red pepper powder samples of different geographical origins in the discriminant space was investigated by scatter plotting the discriminant function scores. The score plot showed good separation among the samples from three different origins (Figure 2), suggesting that the variables used to derive the discriminant functions provided sufficient information to identify the geographical origins of red pepper powders. The Korean and Vietnamese samples were found to be completely distinguishable from each other, while the Chinese samples appeared relatively widely scattered between the Korean and Vietnamese samples. This may be attributed to the diversity of the Chinese samples, reflecting the characteristics of China's large geographical area.
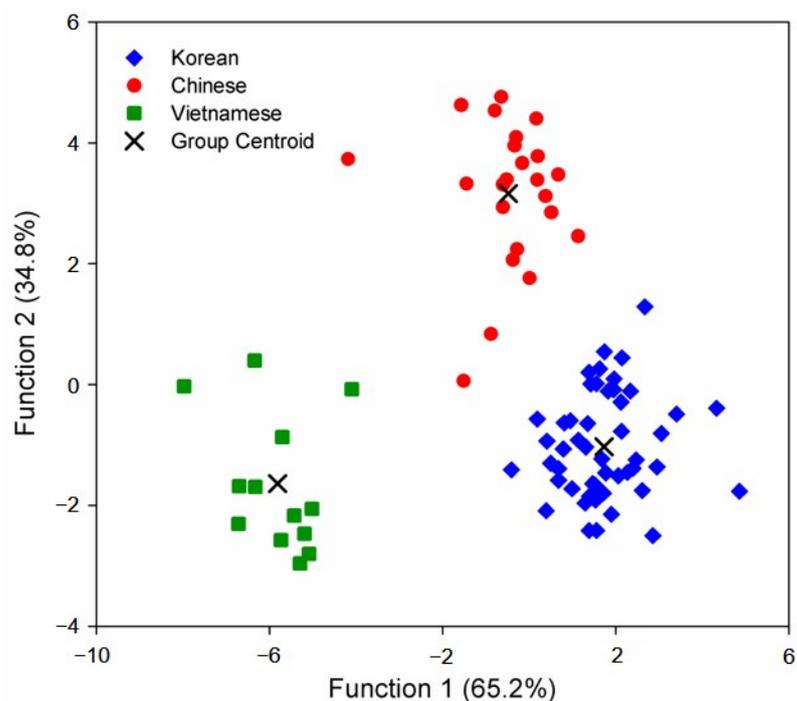
Figure 2. Scatter plot of two discriminant scores for the geographic origins of Korean, Chinese, and Vietnamese red pepper powders.

To verify and examine the predictive discrimination capability of the established canonical discriminant functions, we reclassified the red pepper powder samples used in the multivariate statistical analysis, according to their geographical origins. Table 3 indicates that the canonical discriminant functions correctly classified all 86 red pepper powder samples (50 Korean, 23 Chinese, and 13 Vietnamese) according to their geographical origins (100% of the original group cases were correctly classified), while only one Chinese sample was incorrectly classified in the cross-validation (98.8% of the original group cases were correctly classified). These results were similar to the discrimination results of the origins of 62 Asian red pepper powder (36 Korean, 17 Chinese, and 9 Vietnamese) samples using ^{1}H NMR spectroscopy [23]. In particular, this result was of significance considering that various metabolite components even with minor contents could be used as individual indicators in the ^{1}H NMR analysis. By comparing the analysis results of the mineral elements [25] and sensor characteristics [24] of red peppers from other regions in China using various multivariate statistical analysis methods, it can be observed that their regional scopes were different. However, it can be confirmed that the second-derivative FT-IR method can be sufficiently utilized to discriminate the origins of red pepper powders. In addition, similar discrimination abilities can be confirmed by comparing previous results of the origins of other foods, such as olive oil and honey, using the FT-IR technique [17,18]. Overall, these results indicate that second-derivative FT-IR spectroscopy combined with canonical discriminant analysis has the potential to discriminate Asian red pepper powders according to their geographical origins.

Table 3. Reclassification results for the origins of red pepper powder samples using the canonical discriminant functions.

Origin		Predicted Group			Total	Accuracy (%)
		Korean	Chinese	Vietnamese		
Original	Korean	50	0	0	50	100
	Chinese	0	23	0	23	100
	Vietnamese	0	0	13	13	100
Cross-validated [a]	Korean	50	0	0	50	100
	Chinese	1	22	0	23	95.6
	Vietnamese	0	0	13	13	100

[a] Cross-validation was performed only for those cases in the analysis. In cross-validation, each case is classified using the functions derived from all other cases except that case.

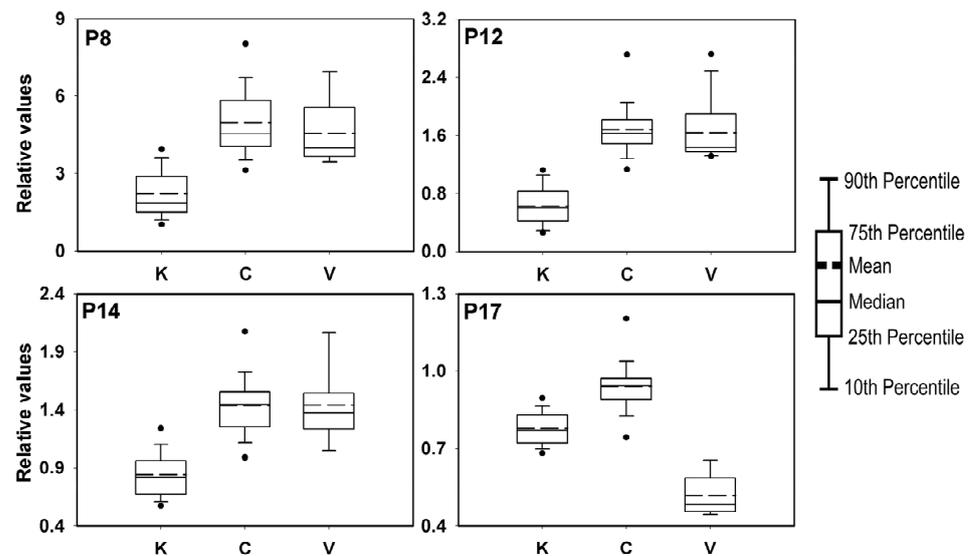### 3.2.2. Discriminant Indicator Variables

It was confirmed that Asian red pepper powders could be effectively discriminated according to their geographical origins by canonical discriminant analysis of the signals obtained from the second-derivative FT-IR spectra. This protocol can also be applied to the discrimination of new red pepper powder samples through statistical processes. If several indicator variables suitable for discriminating the origin of red pepper samples are selected and appropriate ranges are set for them, rapid and facile discrimination of the geographical origins of new red pepper powder samples is possible without the need for a specific statistical program or process.

The Pearson coefficients are summarized in the structure matrix table (Table 4), which shows the correlation of each variable with each canonical discriminant function [47–49]. This table reveals that P12 and P17 are the most significant variables in discriminant Functions 1 and 2 (with correlations of −0.475 and 0.714), respectively. P14, and P8 also show high significance in both functions.

Table 4. Structure matrix table with coefficients for the peak variables used in discrimination analysis.

Variable	Structure Matrix	
	Function 1	**Function 2**
P5	−0.303	0.333
P7	−0.184	0.11
P8	−0.326	0.437
P10	−0.202	0.371
P12	−0.475	0.555
P14	−0.368	0.409
P16	−0.185	0.306
P17	0.394	0.714

These four peak variables (P8, P12, P14, and P17) were also found to have high significance in the mean difference, with an *F*-value of 60 or more in the one-way ANOVA test (Table 2). The distribution of data between the geographical origin groups of these four variables were compared as box plots (Figure 3), confirming that P12 and P17 were the most effective variables for discriminating the Korean and Vietnamese samples, respectively, from those of other geographical origins. Additionally, the distribution characteristics of P8, P12, and P14 were similar, whereas those for P17 were different. This was also confirmed in the Pearson correlation matrix, which shows the correlations among variables (Table S2).

Figure 3. Box plots for variables that are highly correlated with the canonical discriminant functions (K = Korean, C = Chinese, and V = Vietnamese). The dots indicate the 5th and 95th percentiles.

Considering their correlation with the discriminant functions, mean difference, and difference in distribution values, P8 and P14, along with the most significant variables P12 and P17, were selected as indicator variables for discriminating the origins of Asian red pepper powder samples. To discriminate the geographical origins using the specific indicator variables, they must have ranges differentiated according to the origins.

For the Korean red pepper samples, the distribution values of P8 and P12 were smaller than those of the others. These signals can be attributed to C–H stretching vibrations, which are derived from various metabolites containing alkyl groups, and are likely largely influenced by the hydrocarbon chains of fatty acids. Because the fatty acid content is relatively higher in seeds than in the peel of red pepper [50], it can be estimated that the Korean red pepper powder samples contain relatively fewer seeds than the Chinese or Vietnamese samples. Moreover, the P17 signal attributed to the C–O stretching vibration arising mainly from the fructosyl unit [45] was observed to be small in the Vietnamese

samples. This implies that the Vietnamese red pepper powders had relatively lower fructose content than those of the Korean and Chinese peppers, which was also confirmed in previous NMR experiments (Figure S1) [23]. For the Chinese red pepper powder samples, all four variables exhibited relatively higher means than the others. However, owing to the diversity of the Chinese samples, the ranges of all the indicator variables significantly overlapped with the ranges of those for other origins; hence, establishing independent variable ranges for Chinese samples was not possible.

Based on these observations, the ranges of the discriminant indicator variables that could discriminate between Korean and Vietnamese red pepper powder samples were set as presented in Table 5.

Table 5. Ranges of the indicator variables for Korean and Vietnamese samples.

Peak No.	Wavelength (cm^{-1})	Vibration	Range	
			Korean	Vietnamese
P8	1468	C–H (CH$_2$, CH$_3$) bending	<4.945	>3.445
P12	1377	–C–H (CH$_3$) stretching	<1.155	>1.305
P14	1159	–C–O (ester) stretching	<1.555	0.985–2.085
P17	1053	–C–O stretching	0.620–0.945	<0.699

The range of each discriminant variable was set based on their maximum or minimum values, or by considering values between the minimum and maximum based on the relative distribution characteristics of each variable value [27,37,38]. For example, in the case of the P8 variable, because Korean red pepper powders had the lowest distribution, its range was set below the maximum value for Korean samples. On the contrary, the Vietnamese samples had a relatively high distribution and, thus, were set above the minimum value for Vietnamese samples. It is worth noting that if each variable value obtained the analysis of more samples satisfied the normal distribution sufficiently, the ranges could be established using a statistical technique as well.

To confirm the suitability of the selected indicator variables and their range settings, we reclassified the red pepper powder samples used in the multivariate statistical analysis, based on their geographical origins. A sample was attributed to a specific origin only if the values of all the indicator variables for the sample were within the discriminant ranges for that origin; the results are summarized in Table 6. When the ranges of the indicator variables for the Korean red pepper powder samples were applied, all 50 Korean samples were identified as "Korean," and the remaining 36 samples (23 Chinese and 13 Vietnamese) were all classified as "not Korean." When applying the ranges of the indicator variables for the Vietnamese red pepper powder samples to the 36 "not Korean" samples, all 13 Vietnamese samples were identified as "Vietnamese" and the remaining 23 Chinese samples were identified as "not Vietnamese." Changing the order of applying the indicator variable ranges for the Korean and Vietnamese samples produced the same results, indicating that the two sets of ranges were well separated.

Setting the range of discriminant indicator variables aids in determining the authenticity of food, based on the content of intrinsic ingredients (such as metabolites and minerals) [28,37–39]. However, it is not easy to apply this method to discriminate between the origins of the same food. Therefore, it is meaningful that the geographical origin was correctly classified by setting several discriminant indicators and their ranges. Recently, FT-IR spectroscopy combined with statistical analysis has been actively applied to determine the authenticity, adulteration, and geographical origins of various foods. If the discriminant indicator variables and their ranges are set suitably, more effective and practical use of such results can be realized.

Table 6. Reclassification results for red pepper powder samples using the ranges of the indicator variables.

Applied Ranges	Origin	Predicted Results			
		Predicted	Not Predicted	Total	Accuracy (%)
Korean	Korean	50	0	50	100
	Not Korean	0	36	36	100
Vietnamese	Vietnamese	13	0	13	100
	Not Vietnamese	13	0	13	100

3.3. Blind Tests

To evaluate the applicability of the developed statistical discrimination method and the discriminant indicator variables to new samples, a blind test was performed on 19 new red pepper powder samples (9 Korean, 5 Chinese, and 5 Vietnamese), which were not used in the previous statistical analyses. The geographical origins were correctly classified for all the 19 blind red pepper powder samples using the established canonical discriminant functions (Table 7).

Table 7. Classification of the geographical origins of the blind samples using the established canonical discriminant functions.

Sample No.	Function 1	Function 2	Origin	Predicted	Probability
1	−1.075	−2.451	Korean	Korean	0.998
2	0.995	−0.257	Korean	Korean	0.991
3	2.643	−1.205	Korean	Korean	1
4	2.118	−0.816	Korean	Korean	1
5	1.979	−1.449	Korean	Korean	1
6	1.018	−0.926	Korean	Korean	0.999
7	1.49	−0.881	Korean	Korean	1
8	1.66	0.506	Korean	Korean	0.963
9	0.757	−0.917	Korean	Korean	0.998
10	−0.434	2.797	Chinese	Chinese	1
11	−1.023	2.405	Chinese	Chinese	1
12	−1.402	0.23	Chinese	Chinese	0.946
13	−1.825	7.135	Chinese	Chinese	1
14	−0.862	4.166	Chinese	Chinese	1
15	−5.508	1.616	Vietnamese	Vietnamese	1
16	−4.216	−0.415	Vietnamese	Vietnamese	1
17	−4.529	1.186	Vietnamese	Vietnamese	0.977
18	−5.688	0.018	Vietnamese	Vietnamese	1
19	−7.532	−0.289	Vietnamese	Vietnamese	1

Table 8 presents the classification results of comparing the values of the indicator peak variables obtained from the second-derivative FT-IR spectra of the blind samples with the discriminant ranges for the Korean and Vietnamese red peppers. When the ranges of the indicator variables for the Korean red pepper powder were applied, nine blind samples were correctly identified as "Korean", and the remaining 10 blind samples were classified as "not Korean". When applying the ranges of indicator variables for Vietnamese pepper to 10 blind "not Korean" samples, five samples were correctly identified as "Vietnamese". As in the canonical discriminant analysis, the other five samples that were classified as neither Korean nor Vietnamese can be assumed to be Chinese red pepper powder samples. These results indicate that the indicator ranges can be conveniently used to classify the geographical origins of new red pepper powder samples, even if they are established using a limited number of samples.

Table 8. Classification of the blind samples based on their geographical origins using the ranges of the indicator variables for Korean and Vietnamese samples.

Sample No.	P8	P12	P14	P17	Origin	Predicted
1	2.703	0.804	1.097	0.624	Korean	Korean
2	2.637	0.749	1.081	0.794	Korean	Korean
3	2.379	0.549	0.96	0.858	Korean	Korean
4	1.328	0.414	0.712	0.808	Korean	Korean
5	0.955	0.268	0.577	0.753	Korean	Korean
6	1.562	0.465	0.709	0.729	Korean	Korean
7	1.623	0.522	0.772	0.782	Korean	Korean
8	2.393	0.781	0.912	0.873	Korean	Korean
9	1.903	0.637	0.753	0.752	Korean	Korean
10	4.197	1.525	1.437	0.935	Chinese	(Chinese) *
11	4.544	1.515	1.284	0.841	Chinese	(Chinese) *
12	3.684	1.287	1.4	0.759	Chinese	(Chinese) *
13	7.074	2.778	2.22	1.117	Chinese	(Chinese) *
14	5.164	1.858	1.534	0.921	Chinese	(Chinese) *
15	4.362	1.777	1.648	0.69	Vietnamese	Vietnamese
16	5.107	1.818	1.564	0.604	Vietnamese	Vietnamese
17	5.32	1.919	1.608	0.694	Vietnamese	Vietnamese
18	5.67	2.1	1.824	0.617	Vietnamese	Vietnamese
19	6.971	2.468	1.955	0.559	Vietnamese	Vietnamese

* Samples were classified as neither Korean nor Vietnamese.

4. Conclusions

In this study, we investigated the feasibility of second-derivative FT-IR spectroscopy, combined with multivariate statistical analysis, to discriminate red pepper samples from Korea, China, and Vietnam, based on their geographical origins. Canonical discriminant functions for classifying Asian red pepper powders based on geographical origins were derived from the discriminant analysis, and the discriminating capability of the functions was verified by 100% correct reclassification of the origins of the powder samples used in the analysis. The results of the blind test to classify new red pepper powder samples according to geographical origins confirmed that the derived discriminant functions could correctly classify all new test samples. Although the classification method using the canonical discrimination functions is highly accurate, it requires the statistical data and program used to create the functions to discriminate the origins of new samples. To compensate for these limitations and simply determine the geographical origin without a special statistical program, four indicator variables with large differences in values according to their origins were selected from the variables used in the statistical analysis, and their origin-specific ranges were set. These indicator ranges were successfully used to correctly classify the geographical origins of all statistical samples and blind samples. Although applied to a limited number of samples, the use of the ranges of discriminant indicator variables provides a simple classification method for new samples. Further analyses of more red pepper powder samples, including samples from other countries, may enhance the capability and accuracy of the method of using both the canonical discriminant functions and the discriminant indicator variable ranges. In addition, the discriminant method that uses set ranges of the discriminant indicator variables may be useful in terms of experimental methodology; however, it can be expected to have more applications useful in fields that manage the traceability of foods.

In conclusion, the findings of this study indicate that the second-derivative FT-IR spectroscopy is a reliable, low-cost, and convenient analytical method for discriminating Asian red pepper powders according to their geographical origins.

sum of all signals) ^{1}H NMR spectra of Asian red pepper powders at 600 MHz NMR. Table S1: Test of homogeneity of variance between the groups of Korean, Chinese, and Vietnamese red pepper powders using second derivative values of FT-IR spectra. Table S2: Pearson's Correlation Matrix.

Author Contributions: Conceptualization, M.K. and S.A.; methodology, H.S.C., B.H.K., and S.A.; software, M.K., J.H., D.L., and S.K.; validation, M.K., J.H., D.L., and S.A.; formal analysis, M.K. and J.H.; investigation, M.K., J.H., D.L., and S.K.; resources, H.S.C., Y.-H.C., and S.A.; data curation, J.H. and S.A.; writing—original draft preparation, M.K. and J.H.; writing—review and editing, J.H., B.H.K., and S.A.; visualization, M.K. and J.H.; supervision, S.A.; project administration, B.H.K. and S.A.; funding acquisition, Y.-H.C., H.S.C., and S.A. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shin, K.O.; MoriTani, T. Alterations of Autonomic Nervous Activity and Energy Metabolism by Capsaicin Ingestion during Aerobic Exercise in Healthy Men. *J. Nutr. Sci. Vitaminol.* **2007**, *53*, 124–132. [CrossRef] [PubMed]
2. Martinez, S.; Lopez, M.; Gonzalez-Raurich, M.; Bernardo Alvarez, A. The effects of ripening stage and processing systems on vitamin C content in sweet peppers (*Capsicum annuum* L.). *Int. J. Food Sci. Nutr.* **2005**, *56*, 45–51. [CrossRef] [PubMed]
3. Srinivasan, K. Biological Activities of Red Pepper (*Capsicum annuum*) and Its Pungent Principle Capsaicin: A Review. *Crit. Rev. Food Sci Nutr.* **2016**, *56*, 1488–1500. [CrossRef] [PubMed]
4. Sanatombi, K.; Rajkumari, S. Effect of Processing on Quality of Pepper: A Review. *Food Rev. Int.* **2019**, *36*, 626–643. [CrossRef]
5. Schupp, A.; Gillespie, J. Consumer Attitudes Toward Potential Country-of-Origin Labeling of Fresh or Frozen Beef. *J. Food Distrib. Res.* **2001**, *32*, 1–11. [CrossRef]
6. Hong, E.; Lee, S.Y.; Jeong, J.Y.; Park, J.M.; Kim, B.H.; Kwon, K.; Chun, H.S. Modern analytical methods for the detection of food fraud and adulteration by food category. *J. Sci Food Agric.* **2017**, *97*, 3877–3896. [CrossRef]
7. Choi, J.-Y.; Bang, K.-H.; Han, K.-Y.; Noh, B.-S. Discrimination Analysis of the Geographical Origin of Foods. *Korean J. Food Sci. Technol.* **2012**, *44*, 503–525. [CrossRef]
8. Kim, E.; Lee, S.; Baek, D.; Park, S.; Lee, S.; Ryu, T.; Lee, S.; Kang, H.; Kwon, O.; Kil, M.; et al. A comparison of the nutrient composition and statistical profile in red pepper fruits (*Capsicums annuum* L.) based on genetic and environmental factors. *Appl. Biol. Chem.* **2019**, *62*, 48. [CrossRef]
9. Willenberg, I.; Parma, A.; Bonte, A.; Matthaus, B. Development of Chemometric Models Based on a LC-qToF-MS Approach to Verify the Geographic Origin of Virgin Olive Oil. *Foods* **2021**, *10*, 479. [CrossRef] [PubMed]
10. Zhou, Y.; Kim, S.Y.; Lee, J.S.; Shin, B.K.; Seo, J.A.; Kim, Y.S.; Lee, D.Y.; Choi, H.K. Discrimination of the Geographical Origin of Soybeans Using NMR-Based Metabolomics. *Foods* **2021**, *10*, 435. [CrossRef] [PubMed]
11. Becerra-Martinez, E.; Florentino-Ramos, E.; Perez-Hernandez, N.; Gerardo Zepeda-Vallejo, L.; Villa-Ruano, N.; Velazquez-Ponce, M.; Garcia-Mendoza, F.; Banuelos-Hernandez, A.E. 1H NMR-based metabolomic fingerprinting to determine metabolite levels in serrano peppers (*Capsicum annum* L.) grown in two different regions. *Food Res. Int.* **2017**, *102*, 163–170. [CrossRef]
12. Ma, Y.; Tian, J.; Wang, X.; Huang, C.; Tian, M.; Wei, A. Fatty Acid Profiling and Chemometric Analyses for Zanthoxylum Pericarps from Different Geographic Origin and Genotype. *Foods* **2020**, *9*, 1676. [CrossRef] [PubMed]
13. Li, A.; Yat, Y.W.; Yap, W.K.; Lim, C.W.; Chan, S.H. Discriminating authentic Nostoc flagelliforme from its counterfeits by applying alternative ED-XRF and FTIR techniques. *Food Chem.* **2011**, *129*, 528–532. [CrossRef] [PubMed]
14. Ruoff, K.; Luginbühl, W.; Künzli, R.; Iglesias, M.T.; Bogdanov, S.; Bosset, J.O.; Ohe, K.v.d.; Ohe, W.v.d.; Amado, R. Authentication of the Botanical and Geographical Origin of Honey by Mid-Infrared Spectroscopy. *J. Agric. Food Chem.* **2006**, *54*, 6873–6880. [CrossRef] [PubMed]
15. Lee, B.J.; Zhou, Y.; Lee, J.S.; Shin, B.K.; Seo, J.A.; Lee, D.; Kim, Y.S.; Choi, H.K. Discrimination and prediction of the origin of Chinese and Korean soybeans using Fourier transform infrared spectrometry (FT-IR) with multivariate statistical analysis. *PLoS ONE* **2018**, *13*, e0196315. [CrossRef]
16. Ordoudi, S.A.; de los Mozos Pascual, M.; Tsimidou, M.Z. On the quality control of traded saffron by means of transmission Fourier-transform mid-infrared (FT-MIR) spectroscopy and chemometrics. *Food Chem.* **2014**, *150*, 414–421. [CrossRef]
17. Tapp, H.S.; Defernez, M.; Kemsley, E.K. FTIR Spectroscopy and Multivariate Analysis Can Distinguish the Geographic Origin of Extra Virgin Olive Oils. *J. Agric. food Chem.* **2003**, *51*, 6110–6115. [CrossRef] [PubMed]
18. Hennessy, S.; Downey, G.; O'Donnell, C. Multivariate analysis of attenuated total reflection-Fourier transform infrared spectroscopic data to confirm the origin of honeys. *Appl. Spectrosc.* **2008**, *62*, 1115–1123. [CrossRef] [PubMed]
19. Valand, R.; Tanna, S.; Lawson, G.; Bengtstrom, L. A review of Fourier Transform Infrared (FTIR) spectroscopy used in food adulteration and authenticity investigations. *Food Addit Contam Part. A Chem Anal. Control. Expo. Risk Assess.* **2020**, *37*, 19–38. [CrossRef]

20. Giannetti, V.; Mariani, M.B.; Marini, F.; Torrelli, P.; Biancolillo, A. Grappa and Italian spirits: Multi-platform investigation based on GC–MS, MIR and NIR spectroscopies for the authentication of the Geographical Indication. *Microchem. J.* **2020**, *157*, 104896. [CrossRef]

21. Ordoudi, S.A.; Papapostolou, M.; Kokkini, S.; Tsimidou, M.Z. Diagnostic Potential of FT-IR Fingerprinting in Botanical Origin Evaluation of Laurus nobilis L. Essential Oil is Supported by GC-FID-MS Data. *Molecules* **2020**, *25*, 583. [CrossRef]

22. Maione, C.; Barbosa, F.; Barbosa, R.M. Predicting the botanical and geographical origin of honey with multivariate data analysis and machine learning techniques: A review. *Comput. Electron. Agric.* **2019**, *157*, 436–446. [CrossRef]

23. Lee, D.; Kim, M.; Kim, B.H.; Ahn, S. Identification of the Geographical Origin of Asian Red Pepper (*Capsicum annuum* L.) Powders Using 1H NMR Spectroscopy. *Bull. Korean Chem. Soc.* **2020**, *41*, 317–322. [CrossRef]

24. Yin, X.; Xu, X.; Zhang, Q.; Xu, J. Rapid Determination of the Geographical Origin of Chinese Red Peppers (*Zanthoxylum Bungeanum* Maxim.) Based on Sensory Characteristics and Chemometric Techniques. *Molecules* **2018**, *23*, 1001. [CrossRef] [PubMed]

25. Zhang, J.; Yang, R.; Chen, R.; Li, Y.C.; Peng, Y.; Wen, X. Geographical origin discrimination of pepper (*Capsicum annuum* L.) based on multi-elemental concentrations combined with chemometrics. *Food Sci Biotechnol.* **2019**, *28*, 1627–1635. [CrossRef]

26. Song, B.Y.; Ryu, J.S.; Shin, H.S.; Lee, K.S. Determination of the source of bioavailable Sr using ^{87}Sr/^{86}Sr tracers: A case study of hot pepper and rice. *J. Agric. Food Chem.* **2014**, *62*, 9232–9238. [CrossRef] [PubMed]

27. Kohler, A.; Bertrand, D.; Martens, H.; Hannesson, K.; Kirschner, C.; Ofstad, R. Multivariate image analysis of a set of FTIR microspectroscopy images of aged bovine muscle tissue combining image and design information. *Anal. Bioanal. Chem.* **2007**, *389*, 1143–1153. [CrossRef] [PubMed]

28. Park, S.M.; Yu, H.Y.; Chun, H.S.; Kim, B.H.; Ahn, S. A Second Derivative Fourier-Transform Infrared Spectroscopy Method to Discriminate Perilla Oil Authenticity. *J. Oleo. Sci.* **2019**, *68*, 389–398. [CrossRef] [PubMed]

29. Rieppo, L.; Saarakkala, S.; Narhi, T.; Helminen, H.J.; Jurvelin, J.S.; Rieppo, J. Application of second derivative spectroscopy for increasing molecular specificity of Fourier transform infrared spectroscopic imaging of articular cartilage. *Osteoarthr. Cartil.* **2012**, *20*, 451–459. [CrossRef]

30. Mossoba, M.M.; Milosevic, V.; Milosevic, M.; Kramer, J.K.; Azizian, H. Determination of total trans fats and oils by infrared spectroscopy for regulatory compliance. *Anal. Bioanal Chem.* **2007**, *389*, 87–92. [CrossRef] [PubMed]

31. Kim, H.Y. Analysis of variance (ANOVA) comparing means of more than two groups. *Restor. Dent. Endod.* **2014**, *39*, 74–77. [CrossRef] [PubMed]

32. Rencher, A.C. Interpretation of Canonical Discriminant Functions, Canonical Variates, and Principal Components. *Am. Stat.* **1992**, *46*, 217–225. [CrossRef]

33. Gautam, R.; Vanga, S.; Ariese, F.; Umapathy, S. Review of multidimensional data processing approaches for Raman and infrared spectroscopy. *EPJ Tech. Instrum.* **2015**, *2*, 8. [CrossRef]

34. Baker, M.J.; Trevisan, J.; Bassan, P.; Bhargava, R.; Butler, H.J.; Dorling, K.M.; Fielden, P.R.; Fogarty, S.W.; Fullwood, N.J.; Heys, K.A.; et al. Using Fourier transform IR spectroscopy to analyze biological materials. *Nat. Protoc.* **2014**, *9*, 1771–1791. [CrossRef]

35. Zhou, W.; Guo, P.; Chen, J.; Lei, Y. A rapid analytical method for the quantitative determination of the sugar in acarbose fermentation by infrared spectroscopy and chemometrics. *Spectrochim. Acta A Mol. Biomol Spectrosc.* **2020**, *240*, 118571. [CrossRef] [PubMed]

36. Jeon, H.; Kim, I.-H.; Lee, C.; Choi, H.-D.; Kim, B.H.; Akoh, C.C. Discrimination of Origin of Sesame Oils Using Fatty Acid and Lignan Profiles in Combination with Canonical Discriminant Analysis. *J. Am. Oil Chem. Soc.* **2012**, *90*, 337–347. [CrossRef]

37. Kim, J.; Jin, G.; Lee, Y.; Chun, H.S.; Ahn, S.; Kim, B.H. Combined Analysis of Stable Isotope, ^{1}H NMR, and Fatty Acid To Verify Sesame Oil Authenticity. *J. Agric. Food Chem.* **2015**, *63*, 8955–8965. [CrossRef]

38. Lee, J.E.; Hwang, J.; Choi, E.; Shin, M.J.; Chun, H.S.; Ahn, S.; Kim, B.H. Rubidium analysis as a possible approach for discriminating between Korean and Chinese perilla seeds distributed in Korea. *Food Chem.* **2020**, *312*, 126067. [CrossRef] [PubMed]

39. Kim, J.; Yang, S.; Jo, C.; Choi, J.; Kwon, K.; Ahn, S.; Sook Chun, H.; Hee Kim, B. Comparison of Carbon Stable Isotope and Fatty Acid Analyses for the Authentication of Perilla Oil. *Eur. J. Lipid Sci. Technol.* **2018**, *120*, 1700480. [CrossRef]

40. Christy, A.A.; Egeberg, P.K. Quantitative determination of saturated and unsaturated fatty acids in edible oils by infrared spectroscopy and chemometrics. *Chemom. Intell. Lab. Syst.* **2006**, *82*, 130–136. [CrossRef]

41. Guillen, M.D.; Cabo, a.N. Infrared Spectroscopy in the Study of Edible Oils and Fats. *J. Sci Food Agric.* **1997**, *75*, 1–11. [CrossRef]

42. Lammers, K.; Arbuckle-Keil, G.; Dighton, J. FT-IR study of the changes in carbohydrate chemistry of three New Jersey pine barrens leaf litters during simulated control burning. *Soil Biol. Biochem.* **2009**, *41*, 340–347. [CrossRef]

43. Smith, B.C. An IR Spectral Interpretation Potpourri: Carbohydrates and Alkynes. *Spectroscopy* **2017**, *32*, 18–24. Available online: http://www.spectroscopyonline.com/ir-spectral-interpretation-potpourri-carbohydrates-and-alkynes (accessed on 22 April 2021).

44. Wang, L.; Yang, Q.; Zhao, H. Sub-regional identification of peanuts from Shandong Province of China based on Fourier transform infrared (FT-IR) spectroscopy. *Food Control.* **2021**, *124*, 107879. [CrossRef]

45. Wiercigroch, E.; Szafraniec, E.; Czamara, K.; Pacia, M.Z.; Majzner, K.; Kochan, K.; Kaczor, A.; Baranska, M.; Malek, K. Raman and infrared spectroscopy of carbohydrates: A review. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2017**, *185*, 317–335. [CrossRef] [PubMed]

46. Coenen, K.; Gallucci, F.; Mezari, B.; Hensen, E.; van Sint Annaland, M. An in-situ IR study on the adsorption of CO_2 and H_2O on hydrotalcites. *J. CO_2 Util.* **2018**, *24*, 228–239. [CrossRef]

47. Huberty, C.J. Discriminant Analysis_Carl J Huberty. *Rev. Educ. Res.* **1975**, *45*, 543–598. [CrossRef]
48. Schellenberg, A.; Chmielus, S.; Schlicht, C.; Camin, F.; Perini, M.; Bontempo, L.; Heinrich, K.; Kelly, S.D.; Rossmann, A.; Thomas, F.; et al. Multielement stable isotope ratios (H, C, N, S) of honey from different European regions. *Food Chem.* **2010**, *121*, 770–777. [CrossRef]
49. Shi, X.; Li, J.; Wang, S.; Zhang, L.; Qiu, L.; Han, T.; Wang, Q.; Chang, S.K.; Guo, S. Flavor characteristic analysis of soymilk prepared by different soybean cultivars and establishment of evaluation method of soybean cultivars suitable for soymilk processing. *Food Chem.* **2015**, *185*, 422–429. [CrossRef] [PubMed]
50. Ananthan, R.; Subhash, K.; Longvah, T. Capsaicinoids, amino acid and fatty acid profiles in different fruit components of the world hottest Naga king chilli (*Capsicum chinense* Jacq). *Food Chem.* **2018**, *238*, 51–57. [CrossRef] [PubMed]