Research Article

Received: 3 June 2020 Revised: 25 September 2020 Accepted article published: 2 October 2020 Published online in Wiley Online Library: 21 October 2020

(wileyonlinelibrary.com) DOI 10.1002/jsfa.10845

Rapid detection of green-pea adulteration in pistachio nuts using Raman spectroscopy and chemometrics

Osman Taylan,[a] Nur Cebi,[b*] Mustafa Tahsin Yilmaz,[a] Osman Sagdic,[b] Durmus Ozdemir[c] and Mohammed Balubaid[a]

Abstract

BACKGROUND: Ground pistachio nut is prone to adulteration because of its high economic value and wide usage. Green pea is known as the main adulterant in frauds involving pistachio nuts. The present study developed a new, rapid, reliable and low-cost methodology by using a portable Raman spectrometer in combination with chemometrics for the detection of green pea in pistachio nuts.

RESULTS: Three different methods of Raman spectroscopy-based chemometrics analysis were developed for the determination of green-pea adulteration in pistachio nuts. The first method involved the development of hierarchical cluster analysis (HCA) and principal component analysis (PCA), which differentiated authentic pistachio nuts from green pea and green pea-adulterated samples. The best classification pattern was observed in the adulteration range of 20–80% (w/w). In addition to classification methods, partial least squares regression (PLSR) and genetic algorithm-based inverse least squares (GILS) were also used to develop multivariate calibration models to determine quantitatively the degree of green-pea adulteration in grounded pistachio nuts. The spectral range of 1790–283 cm^{-1} was used in the case of multivariate data analysis. A green-pea adulteration level of 5–80% (w/w) was successfully identified by PLSR and GILS. The correlation coefficient of determination (R^2) was determined as 0.91 and 0.94 for the PLSR and GILS analyses, respectively.

CONCLUSION: A Raman spectrometer combined with chemometrics has a high capability with regard to the detection of adulteration in pistachio nuts, combined with low cost, strong reliability, a high level of accuracy, rapidity of analysis, and minimum sample preparation.

Keywords: portable Raman; pistachio nut; hierarchical cluster analysis (HCA); principal component analysis (PCA); partial least squares (PLSR); genetic inverse least squares (GILS)

INTRODUCTION

In recent years, economically motivated adulteration has attracted attention as an important risk in industry, government, and food analysis laboratories. In general, food fraud occurs in the form of replacement, addition, and removal. The Food and Drug Administration (FDA) defined 'economically motivated adulteration' as the intentional substitution or addition of a substance in a product to increase the apparent value or reduce its cost.[1] Replacement is the complete or partial replacement of the valuable authentic ingredient with a cheaper substitute, and is probably the most common type of adulteration worldwide.[2] Cheap or substitute materials may have a serious impact on the health of consumers. Any adulteration also causes changes in the identity and authenticity of the original, authentic food product.[3] The protection of authenticity is a crucial requirement for maintaining food safety, from raw materials to final product, in all sections of industry, for producers, suppliers, retailers, consumers, and regulators.[4]

The pistachio nut is one of the most popular nuts in the world due to its unique organoleptic characteristics.[5] According to the International Nut and Dried Fruit Statistical Yearbook, the world pistachio production was 586 207 metric tons in the 2017/2018 season, and the leading producing countries were the USA and Iran, with production rates of 47% and 38%, respectively. Turkey is the third biggest producer with 9% of world production in the 2017/2018 season.[6] In addition to the favorable organoleptic properties, pistachio nuts are of high nutritional importance due to the healthy compounds in their composition, including protein,

* *Correspondence to: N Cebi, Davutpaşa Campus, Chemical and Metallurgical Engineering Faculty, Food Engineering Department, Yıldız Technical University, 34210, Istanbul, Turkey. E-mail: nurcebi@yildiz.edu.tr*

a *Department of Industrial Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah, Saudi Arabia*

b *Davutpaşa Campus, Chemical and Metallurgical Engineering Faculty, Food Engineering Department, Yıldız Technical University, Istanbul, Turkey*

c *Faculty of Science, Department of Chemistry, İzmir Institute of Technology, Izmir, Turkey*

dietary fiber, potassium, vitamin K, γ-tocopherol, and valuable phytochemicals.[7]

In Turkey, pistachio nuts are used in confectionery products and in popular traditional Turkish desserts such as baklava, kadayif, kunefe, and burma. The pistachio nut is subject to adulteration for economic reasons such as its high price and wide areas of application. However, it is quite challenging to discriminate ground pistachio nuts from adulterated ones in terms of their sensory and textural properties. Moreover, pistachio nuts are frequently supplied by the confectionary sector and by Turkish dessert factories in a ground form. Previous studies have reported that green peas, spinach, and peanuts with food dyes have been used as adulterants in the case of ground pistachio nuts.[8,9] A market search showed that a markedly cheaper product referred to as 'decorative peanut' – a mixture of peanut flour, sugar, starch, and green food dye – is used in confectionery products instead of authentic pistachio nuts. This is especially so in the case of popular Turkish desserts such as baklava, kadayıf, burma, etc., in which pistachio nut is used, and which have a high economic value as both local and export products. Thus, the detection of adulteration in pistachio nuts, or the ability to discriminate authentic pistachio nuts from adulterated ones, is important as a means of preventing food fraud, ensuring food safety, and maintaining food authenticity.

Currently, authenticity is a significant concern from the field to the fork, including in the food industry, in the case of producers, suppliers, regulators, and consumers. Government regulations worldwide have determined quality standards with regard to authentic food products. However, foods may be adulterated using cheap ground or chemical substances, and this situation may result in serious economic disruption as well as serious health problems.[4] There is an absolute need for controlling the authenticity of food in order to maintain food quality and safety, as well as ensuring consumer health and satisfaction. The worldwide occurrence of high levels adulteration requires fast, reliable, effective, and robust analytical methods that can cope with adulteration problems and authenticity challenges. Previous studies have reported that polymerase chain reaction (PCR), high pressure liquid chromatography (HPLC), mass spectroscopy (MS), nuclear magnetic resonance (NMR), gas chromatography (GC), isotope ratio-mass spectroscopy (IR-MS), mid-infrared spectroscopy (MIR), near-infrared spectroscopy (NIR), and the Raman technique have been used as well-known analytical approaches for the determination of adulteration in a wide variety of food categories such as animal origin food, seafood, beverages, spices, sweetness and others (organic foods, dietary supplements).[4] These technologies are reliable and have been implemented successfully to determine the quality of food products affected by adulteration. However, most of them are generally costly, time-consuming, and destructive, and require the employment of highly-skilled operators.[10]

Demand exists for rapid, inexpensive, and practical techniques for adulteration testing in foods. Infrared spectroscopies combined with chemometrics have shown a high potential for the rapid, strong, effective, and reliable detection of adulterants in complex food matrices. Additionally, infrared (IR) techniques (MIR, NIR, Raman, hyperspectral imaging) are sensitive, green, non-destructive, and comparatively low-cost[11] approaches. Raman spectroscopy has some advantages over NIR and MIR spectroscopic techniques, including minimal interference from water and easy analysis through glass or polymer packaging.[12] Moreover, Raman spectroscopy with handheld options can be used as an effective tool for the at-line or on-line monitoring of food systems.[10]

Several studies have documented the determination of adulteration in pistachio nuts. Rigano *et al.* (2019) developed an iknife (intelligent knife)-coupled REIMS (rapid evaporative ionization mass spectrometry) method for determining the authenticity of pistachio nuts based on geographical origin.[13] Boukid *et al.* (2019) utilized the physical and chemical features of pistachio nuts (*Pistacia vera* L.) for geographical origin discrimination.[14] Cavus *et al.* (2018) developed a methodology for the determination of green-pea adulteration in pistachio nuts (*Pistacia vera* L.) by using liquid chromatography-mass spectrometry combined with chemometrics.[9] In another previous study, Sezer *et al.* (2019) successfully detected green pea and spinach adulteration in pistachio nuts using laser-induced breakdown spectroscopy.[8]

The current study has created a new, rapid, effective, reliable and low-cost methodology for the determination of green-pea adulteration in pistachio nuts using a portable Raman spectrometer combined with classifications such as hierarchical cluster analysis (HCA) and principal component analysis (PCA) in conjunction with two multivariate calibration methods: partial least squares regression (PLSR) and genetic algorithm-based inverse least squares (GILS).

MATERIALS AND METHODS

Apparatus and reagents

A Raman Progeny X2 spectrometer (Rigaku Analytical Devices, Wilmington, MA, USA) was used for spectral acquisition. All measurements were obtained in the 200–2000 cm^{-1} spectral range. Instrument control and data acquisition were accomplished by using Xantus V3.0.0.0 software. Benzonitrile standard solution was used for verification prior to the Raman measurements.

Sample preparation for the Raman analysis

Samples of raw pistachio nuts (n = 12) were obtained from local producers from Gaziantep and Siirt (Turkey). Green pea samples (n = 2) were obtained from well-known producers in Turkey and were freeze-dried prior to the analysis. The dried green peas and raw pistachio nuts were ground using a coffee grinder machine (Fakir Hausgeräte, Stuttgart, Germany). The average particle sizes of the pistachio nut and green pea samples were 212 µm and 180 µm, respectively. Ground raw pistachio nuts were homogenously mixed with ground green peas at ratios of between 5% and 60%. They were mixed manually perfectly for 5 min and placed in glass vials (5 cc) prior to the Raman measurements. The green pea content of the adulterated pistachio nut samples (n = 108) employed in this study are presented in Table 1. Three replicates of each sample were prepared.

In the present study, 21 commercial ground pistachio nut samples were obtained from Turkish dessert producers (İstanbul) and pistachio nut producers (Gaziantep) in Turkey. These samples were used as test samples for confirmation of the identification results from the Raman method. Five of the 21 samples (Table 2) were obtained from reputable pistachio nut producers and used as authentic test samples (numbered as 3,10,14,18 and 21). The remaining commercial samples (n = 16) were obtained from Turkish dessert producers in Istanbul to evaluate the models.

Raman measurements

Raman measurements were performed using a portable Raman spectrometer, Progeny (Rigaku Analytical Devices, Wilmington,

Table 1. Green pea contents of 12 different pistachio nut samples used in the present study

		Green Pea (GP) Content (%)									
Sample Name	*Sample Code*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Pistachio Nut 1	P1	5	10	15	20	25	30	40	50	60	0
Pistachio Nut 2	P2	5	10	15	20	25	30	40	50	60	0
Pistachio Nut 3	P3	5	10	15	20	25	30	40	50	60	0
Pistachio Nut 4	P4	5	10	15	20	25	30	40	50	60	0
Pistachio Nut 5	P5	5	10	15	20	25	30	40	50	60	0
Pistachio Nut 6	P6	5	10	15	20	25	30	40	50	60	0
Pistachio Nut 7	P7	5	10	15	20	25	30	40	50	60	0
Pistachio Nut 8	P8	5	10	15	20	25	30	40	50	60	0
Pistachio Nut 9	P9	5	10	15	20	25	30	40	50	60	0
Pistachio Nut 10	P10	5	10	15	20	25	30	40	50	60	0
Pistachio Nut 11	P11	5	10	15	20	25	30	40	50	60	0
Pistachio Nut 12	P12	5	10	15	20	25	30	40	50	60	0

MA, USA), equipped with 1064 nm yttrium aluminium garnet (YAG) laser. A Peltier cooled indium gallium arsenide (InGaAs) detector was used for the detection of the scattered Raman light. All spectra were recorded in the 2000–200 cm^{-1} spectral range. Laser power and exposure time were selected as 0.25 watts and 0.851 s, respectively. The spectrum acquisition of each sample was repeated three times in identical conditions, and the average spectrum was measured. Measurements were performed through glass vials. The same procedure was applied for all samples.

Chemometric analysis

Classification and discrimination

Hierarchical cluster analysis and PCA were conducted utilizing the OPUS Version 7.2 software (Bruker, Germany). Hierarchical cluster analysis was performed in order to reveal the intrinsic relationship between the authentic pistachio nut samples and the green-pea adulterated ones using their Raman spectra. Clusters and sub-clusters in which pistachio nut, green pea, and adulterated samples scattered were displayed using OPUS Version 7.2 software. Raman spectra were preprocessed in the 1790–283 cm^{-1} spectral range through vector-normalization, and first derivatization (25 smoothing points) was carried out prior to the hierarchical cluster analysis and PCA. The PCA was employed to observe the classification pattern of the authentic pistachio nuts, the green pea, and the adulterated samples in a three-dimensional PCA plot. Dendrograms were obtained by HCA analysis, with factorization being selected as the appropriate algorithm. Three-dimensional cluster analysis plots were achieved by using the identity test method of the OPUS Version 7.2 software.

Multivariate calibration (PLSR and GILS)

Multivariate calibration models were generated with PLSR and GILS methods applied to the Raman data of the pistachio nuts, the adulterated samples, and the green peas. Calibration set, validation set, and prediction set were created by using the Microsoft Excel (MS Office 2010, Microsoft Corporation) program. Multivariate calibration models were developed by using calibrations set where leave one-out cross validation was applied in order to avoid overfitting of the models and also to determine the optimum number of PLSR components. In addition, the predictive capability of the developed calibration models was determined through independent validation sets. The GILS and PLSR methods were both coded in the MATLAB programming language (MATLAB R2016a-MathWorks Inc., Natick, MA, USA). The GILS algorithm was set to run with 30 genes, where each gene represented randomly selected Raman spectral data points from the whole spectral range used in this study. The length of each gene (the maximum size) depended on the number of calibration samples, as GILS would be over-fitted if the spectral data point exceeded the calibration samples. The genes were selected from the whole spectral range, with an initial selection criterion of R^2 having a value of at least 0.50, meaning that any initial gene that was used to build an GILS model was not accepted if the R^2 value of the

Table 2. Test samples of different product types employed for validation of Raman methodology

Number of test samples	Product type	Raman identification (HCA analysis, PLSR, GILS)
1	Decoration on Kadayıf	Adulterated Pistachio Nut
2	Original Pistachio Nut	Pistachio Nut
3	Original Pistachio Nut	Pistachio Nut
4	Original Pistachio Nut	Pistachio Nut
5	Decoration on Baklava	Adulterated Pistachio Nut
6	Original Pistachio Nut	Pistachio Nut
7	Decoration on Baklava	Adulterated Pistachio Nut
8	Decoration on Baklava	Adulterated Pistachio Nut
9	Original Pistachio Nut	Pistachio Nut
10	Decoration on Kunefe	Pistachio Nut
11	Decoration on Baklava	Pistachio Nut
12	Grinded Pistachio Nuts	Adulterated Pistachio Nut
13	Decoration on Pudding	Adulterated Pistachio Nut
14	Decoration for Ice Cream	Pistachio Nut
15	Decorative pea nuts	Adulterated Pistachio Nut
16	Decoration on Baklava	Adulterated Pistachio Nut
17	Decoration on Kunefe	Adulterated Pistachio Nut
18	Original Pistachio Nut	Pistachio Nut
19	Decoration on Kunefe	Adulterated Pistachio Nut
20	Decoration on Baklava	Adulterated Pistachio Nut
21	Decoration on Kunefe	Pistachio Nut

model was not at least 0.50 and was thrown out. Once the initial genes were selected, the program was set to run 50 iterations so that the genes were put into a single point cross over 50 times, with the best gene with the lowest standard error of cross validation (SECV) being selected to build the final model of the run. As GILS is heavily governed by random processes, there will always be different results whenever the program is rerun. As a result, the program was set to run 100 times, with the number

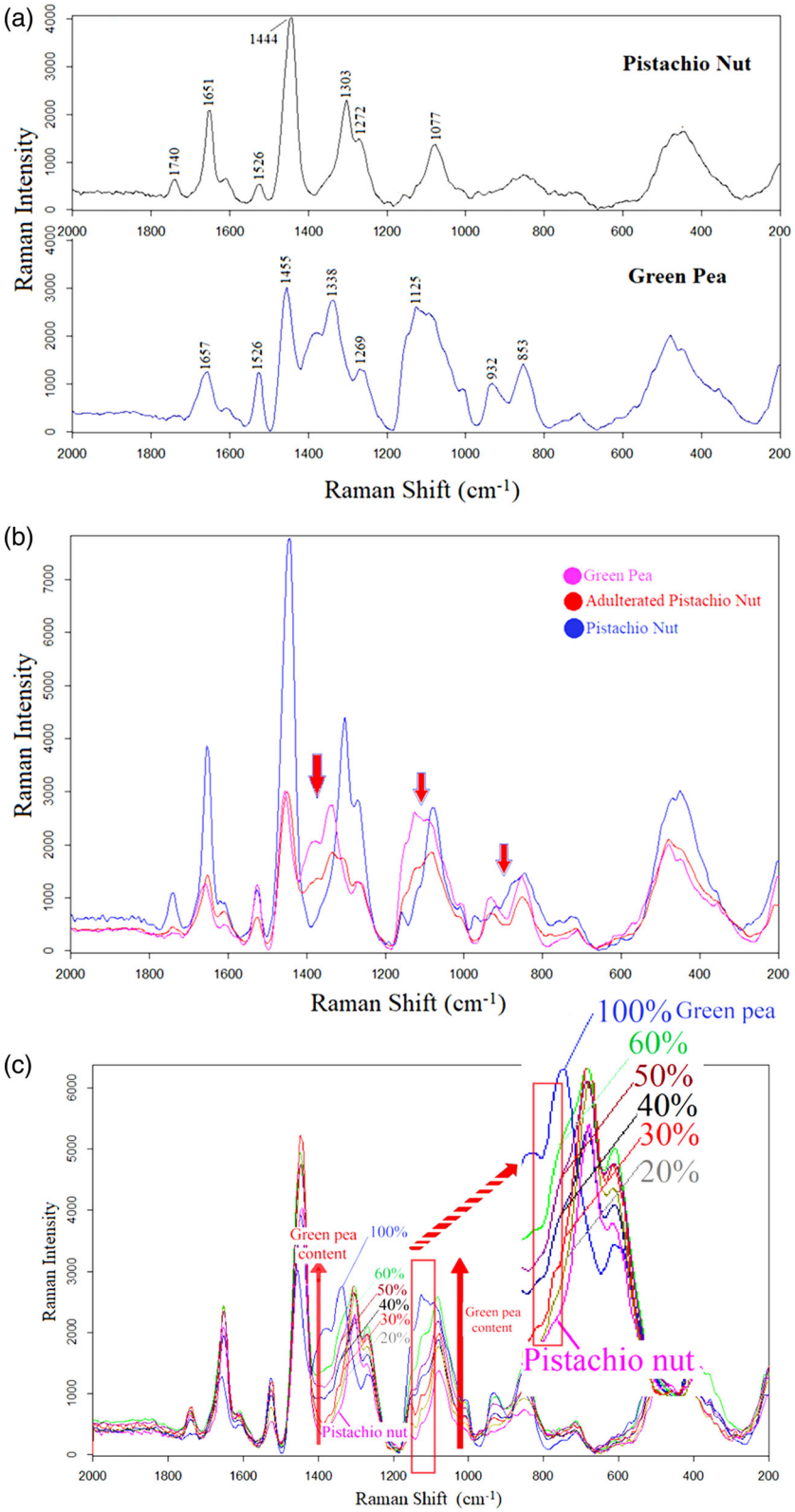

Figure 1. (a) Raman spectra of pistachio nut and green pea samples. (b) Overlapped Raman spectra of green pea, adulterated pistachio nut and pistachio nut samples (2000–200 cm^{-1}). (c) Adulteration level (0%, 20%, 30%, 40%, 50%, 60%, 100%) related Raman spectra of samples.

of genes and iterations being set to 30 and 50 respectively, in each run. At the end, there was a total of 100 best models, and the average predictions of these models were used to predict an independent validation set, and the standard error of prediction (SEP) was determined. In addition, the same average prediction approach was used to predict the additional test set given in Table 2.

RESULTS

Characterization of the Raman spectra of pistachio nuts and green peas

The Raman spectra of all the samples were obtained in the spectral range of 200–2000 cm^{-1}. Spectral bands are associated with the chemical groups of components present in the chemical structure of the pistachio nuts and green peas. The Raman spectra of the pistachio nuts and the green peas are presented in Fig. 1(a). Significant vibrational bands were observed at 1740, 1651, 1526, 1444, 1303, 1272, 1077 and 456 cm^{-1} in the characteristic Raman spectrum of the pistachio nuts. The band with peak point at 1740 cm^{-1} corresponds to C═O stretching vibrations of lipids in the composition of the pistachio nuts.[15,16] The strong band at 1651 cm^{-1} results from the Amid I vibrations of proteins.[17,18] The strongest band is observed at 1444 cm^{-1} and is associated with the vibrations arising from the CH_2 functional group of lipids in the composition of the pistachio nuts.[15,18]

According to the Turkish Food Composition Database (Turkish FCDB), TürKomp, raw pistachio nuts include 511.5 $g.kg^{-1}$ fat, 176.4 $g.kg^{-1}$ protein and 105.5 $g.kg^{-1}$ carbohydrates in their composition.[19] Lipids are major compounds in pistachio nuts. Consequently, the bands at 1301, 1272 and 1077 cm^{-1} correspond to the CH scissoring, =C-H scissoring and CH_3 bending vibrations of lipids, respectively.[15,18]

Clear vibrational bands were observed at 1657, 1526, 1455, 1338, 1125, 932, and 853 cm^{-1} in the characteristic Raman spectrum of the green peas. According to the Turkish FCDB, the major constituents are carbohydrates (429.8 $g.kg^{-1}$), fiber (236.5 $g.kg^{-1}$), protein (198.2 $g.kg^{-1}$) and lipids (11.5 $g.kg^{-1}$) in the composition of dried green peas.[19] The strong band at 1657 cm^{-1} corresponds to the Amid I vibrations of proteins.[17,18] Additionally, the band with a peak point at 1526 cm^{-1} may be attributed to the Amid II in protein structure.[20,21] The spectral range between 1200 and 1500 cm^{-1} may include spectral bands arising from carbohydrates.[21,22] The band at 1455 cm^{-1} corresponds to the CH, CH_2, and COH deformation vibrations of carbohydrates.[21,22] The starch content of the peas has been reported to be quite high in previous studies.[23] The bands at 1338, 1125, 932, and 853 cm^{-1} result from the ν(C—O) & δ(C-O-H), ν (C—O) + ν(C—C) + δ(C-O-H), δ(C-O-C) + δ(C-O-H) + ν(C—O), and δ(C-C-H) + δ(C-O-C) stretching and bending vibrations of starch molecules, respectively.[22,24]

As mentioned previously, although lipids are the major ingredient of pistachio nuts, carbohydrates are the major ingredient of green peas. The compositional disparity between pistachio nuts and green peas creates spectral differences because Raman scattering provides an intrinsic fingerprint in terms of chemical composition.[25] The Raman spectrum provides a unique fingerprint in terms of molecules.[26] As can be seen from Fig. 1(b), there are spectral differences between pistachio nuts and adulterated pistachio nuts in the whole spectral range of 200–2000 cm^{-1}. Most of the spectral differences arise from the starch content of the green peas, and the major ones (the most significant) were shown using red arrows at 1338, 1125, 932, and 853 cm^{-1} (Fig. 1(b)). The intensity of these bands increases slightly with increases in the green pea content of the adulterated samples (Fig. 1(b)). The intensity of starch-related vibrational bands clearly increases at the spectral ranges of 1090–1160 cm^{-1} and 1330–1360 cm^{-1} in relation to the increase in the green pea content of the adulterated samples. The spectral ranges mentioned above were marked with red rectangles in Fig. 1(b). These regions were not the only regions in which adulteration-related alteration was observed. Spectral changes were observed in many Raman spectra

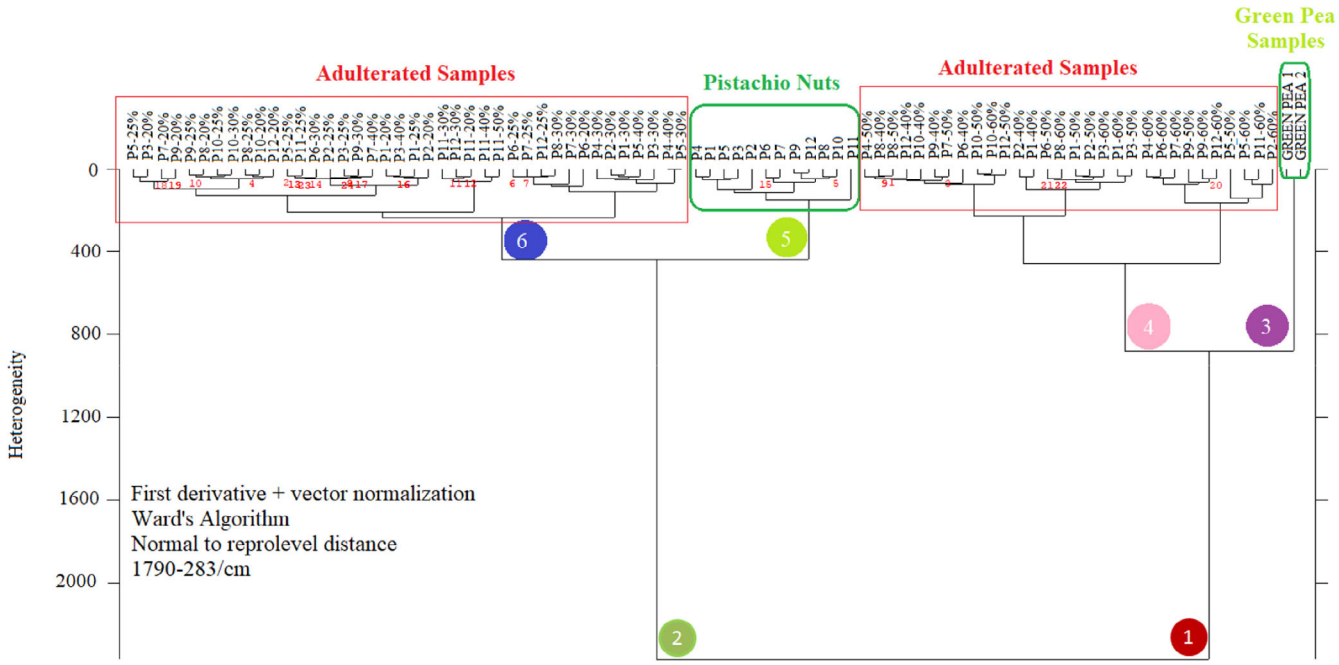

Figure 2. Dendrogram of HCA analysis (Ward's algorithm) of Raman spectra from a total of 84 different samples of pistachio nuts, green peas and adulterated samples.

frequencies. However, the most visually significant ones were marked for clear presentation.

Discrimination and classification of pure and adulterated samples

The main goal of the current study was to discriminate authentic pistachio nuts from adulterated ones involving green peas using Raman spectroscopy combined with chemometrics multivariate classification and clustering methods such as HCA and PCA. The overlapped Raman spectra of green peas, adulterated pistachio nuts (30%), and pistachio nuts are presented in Fig. 1(b). Slight spectral differences were observed between all samples in the spectral range of 200–2000 cm^{-1}. However, the most significant distinction was observed in the spectral range between 800 and 1500 cm^{-1}. This spectral range mostly includes vibrational bands arising from the starch ingredient of the green peas. Explicit regions were marked with a red arrow in Fig. 1(b). The spectral bands and their assignments were described in the previous section.

A number of studies reported that the spectral region between 400 and 2000 cm^{-1} has been known as the fingerprint region, and this region includes the very specific and characteristic spectral properties of the investigated compounds.[27–29] It has also been reported that the window between 900 and 600 cm^{-1} is referred to as the true fingerprint region with very specific spectral patterns.[30] The HCA and PCA cluster analyses were performed by taking first the derivative spectra with vector normalization preprocessing. Hierarchical cluster analysis is an algorithmic approach that is used for the observation of a hierarchy of clusters. In hierarchical cluster analysis, clusters and sub-clusters are visualized definitively in dendrogram graphs.[26] Hierarchical was performed on the 1790–283 cm^{-1} spectral region. The dendrogram was constructed using Ward's algorithm. Known as the minimum variance method, the Ward's method joins at each stage of the cluster pair whose merger minimizes the increase in the total within group error sum of squares. In other words, Ward's method minimizes the increase in the total within-cluster sum of the squared error. This increase is proportional to the squared Euclidean distance between cluster centers.[31]

Spectral distances were calculated using the normal- to repro-level distance. Normal- to repro-level algorithms calculated the spectral distances separately for each frequency range.[32] The most crucial advantage of hierarchical cluster analysis is its ability to provide detailed information about the relationship between clusters and sub-clusters. Hierarchical cluster analysis dendrograms of pistachio nuts, adulterated samples, and green pea samples are presented in Fig. 2. As can be seen from the dendrograms, pistachio nut samples (n = 12) were clearly discriminated from adulterated samples and green pea samples by using selected algorithms and chemometrics parameters. The green pea content of the pistachio nut samples are presented in detail in Table 1. The superiority of this methodology (HCA dendrogram) is that it has the ability to discriminate authentic pistachio nut samples from adulterated ones and from green pea samples. Well-separated clusters were observed with a high heterogeneity value of 2000. Authentic pistachio nut samples clustered on the arm, which is numbered as 5 in Fig. 2. All the samples were divided into two distinct clusters (illustrated as 1 and 2 on the dendrogram). While the arm numbered '1' was divided into two clusters, in which green pea samples were significantly discriminated from adulterated samples, arm 2 was divided into two clusters, with the authentic pistachios being clearly distinguished from the adulterated ones.

Principal component analysis was performed in the same spectral region (1790–283 cm^{-1}). The construction of the PCA 3-D plot was performed using the factorization method. The PCA plot is presented in Fig. 3(a). The PCA results show that the successful and explicit differentiation between the pistachio nut, green pea, and adulterated samples is possible. All samples were classified as belonging to three distinct clusters. It is possible to observe the clear and strong discrimination between the green pea and the pistachio nut samples in the 3-D PCA plot. Although the discrimination between the pistachio nut and the adulterated samples was complicated, a distinct classification was accomplished by using the multivariate chemometrics method. Additionally, the first derivative and vector normalized Raman spectra (processed spectra) of the pistachio nut, the adulterated and the pure samples are presented in Fig. 3(b). As a result, the classification and discrimination of all samples were accomplished correctly by using two different chemometric techniques – HCA and PCA. The best classification pattern was obtained at an adulteration level ranging from 20% to 60% (w/w). Below a 20% (w/w) concentration, perfect discrimination could not be obtained by PCA and HCA analyses using the chemometric software of OPUS Version 7.2. Twenty-one samples (Table 2) were tested using the HCA dendrogram. The authenticity of the samples was determined

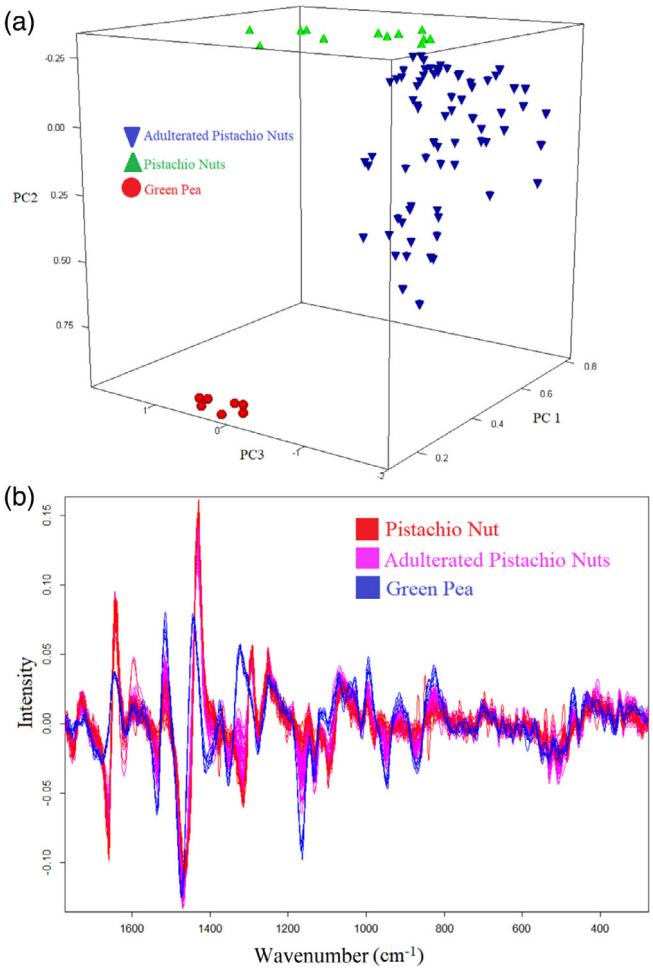

Figure 3. (a) Three-dimensional PCA analysis maps for a total of 84 samples of green peas, pistachio nuts, and adulterated samples. (b) Overlaid first derivative Raman spectra of pistachio nut, adulterated, and green pea samples in the spectral region of 1790–283 cm^{-1}.

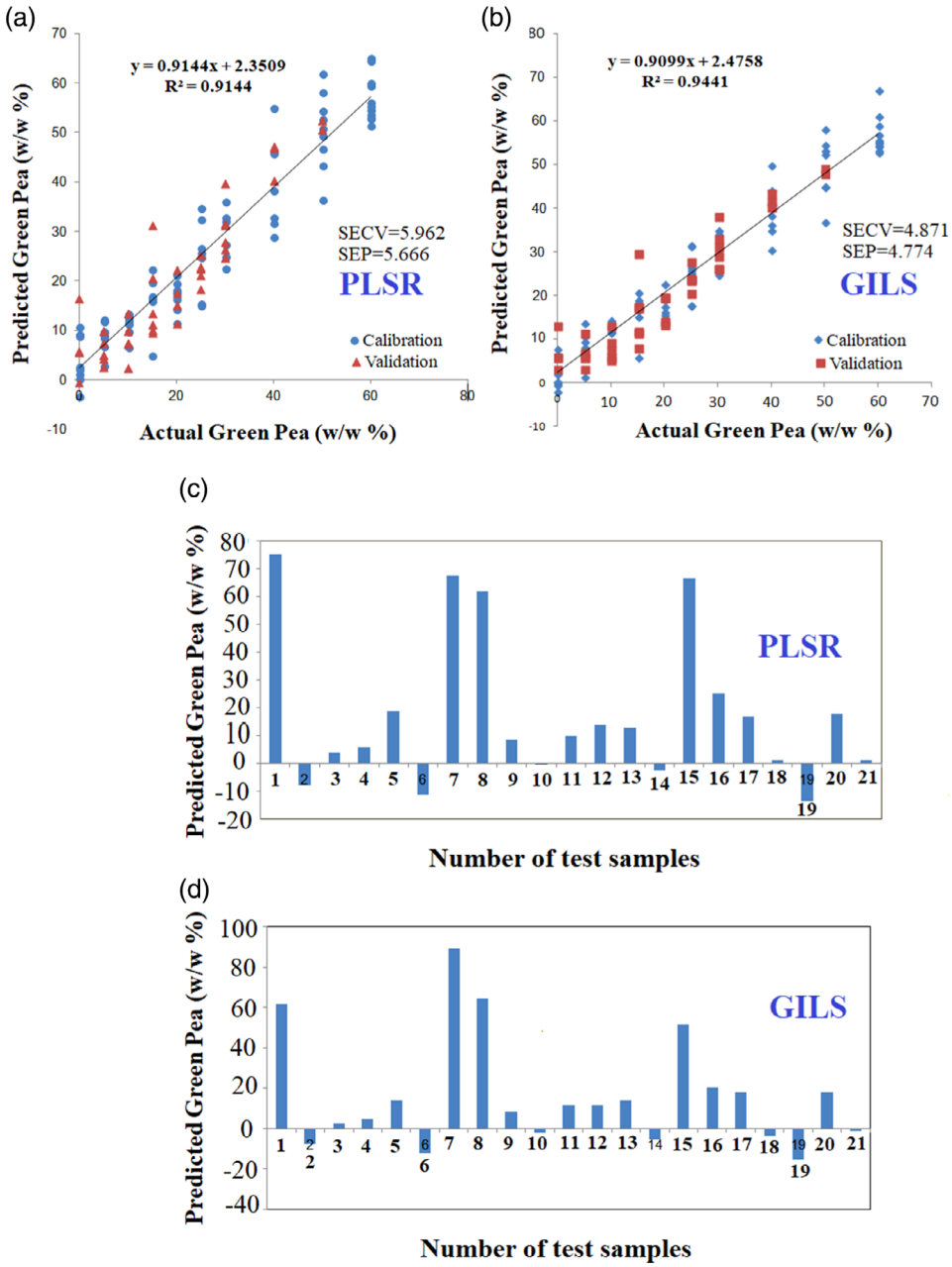

Figure 4. (a) Actual versus predicted correlation graph of PLSR analysis. (b) Actual versus predicted correlation graph of GILS analysis. (c) Bar graph of predicted green pea percentage (w/w) in test samples by PLSR. (d) Bar graph of predicted green pea percentage (w/w) in test samples by GILS.

according to the cluster that appears on the HCA dendrogram. The results are presented as Raman HCA identification in Table 2. While 11 samples were identified as being adulterated, ten samples were identified as pistachio nuts. Five samples that were known to be pure pistachio nuts (numbered 3,10,14,18, and 21) were determined to be pistachio nuts using the HCA methodology.

Multivariate calibration results

Two different multivariate calibration models were built to determine the green pea content of adulterated pistachio nut samples with regard to the samples given in Table 1–108 adulterated (12*9 = 108) and 12 pure pistachio nut samples. Among these 120 samples, 80 were used in the calibration set and the remaining 40 were reserved for the independent validation set. Multivariate calibration models were generated with PLSR and GILS by using leave-one-out cross validation for both the green pea and the pistachio nut content of the samples. Here, the PLSR model was constructed with six principal components (PCs). However, only the green pea calibration plots were displayed, as the mixtures were formed from binary components resulting in highly correlated concentrations of the components. This means that PLSR would give the same calibration model regardless of the component being modeled. Inverse least squares (ILS) is a multivariate calibration method based on the inverse Beer's law, in which the concentration of an analyte is modeled as a function of absorbance.[33] The ILS calibration method has capability to overcome the spectral problems such as collinearities and noise. GILS being a modified version of the original inverse least squares method, in which a small set of wavelengths is selected from a full

spectral data matrix and evolved to an optimum solution using a genetic algorithm.[33]

In GILS, the models are formed with randomly selected set of Raman intensities from the range of given spectra and therefore dependencies of the models to each other are somewhat waived. In addition to the independent validation set, the stability of the calibration models was tested by using a separate prediction test, to demonstrate the performance of the models with regard to real samples. In the current research, the GILS algorithm was set to run 100 times with 50 iterations and 30 genes for each component of the mixture. The success (performance) of the models were evaluated using standard error of cross-validation (SECV) and standard error of prediction (SEP) and also R^2 values. Statistical indices SECV and SEP are defined as follows:

$$\mathrm{SECV} = \sqrt{\frac{\sum_{i=1}^{m} (c_i - \hat{c}_i)^2}{m-2}} \quad (1)$$

$$\mathrm{SEP} = \sqrt{\frac{\sum_{i=1}^{m} (c_i - \hat{c}_i)^2}{m}} \quad (2)$$

where c_i is the reference and $\hat{c}_i$ is the predicted concentration values of *i*th sample, *m* is the number of samples. Degree of freedom is $m - 2$ because, when a linear model is assumed, there are only two parameters to be extracted, which are the slope of the actual versus reference concentration plot and the intercept.[34]

The resulting SECV and SEP values are presented on the actual green pea versus predicted green pea graphs in Fig. 4 for both the PLSR (Fig. 4(a)) and the GILS (Fig. 4(b)) models. The regression R^2 values were determined as 0.91 and 0.94 for the PLSR and GILS models, respectively, and less than 5% SECV and SEP values were obtained by using the GILS model. These are quite good model errors as the dynamic range of the model covers the values from 0% to 60% green pea content. In addition, 12 different pistachio nut samples were used to prepare the same composition of adulterated samples, and this also brings an additional variability because each of the pistachio samples is somewhat different from the others, as can be seen in the PCA score plot.

The bar graphs for the predicted green pea content of the commercial test samples given Table 2 are presented in Fig. 4(c) and Fig. 4(d), for the PLSR and GILS models, respectively. Twenty-one samples (Table 2) were evaluated using PLSR and GILS models. While 11 samples were identified as being adulterated, ten samples were identified as pistachio nuts. Five samples, which were known to be pure pistachio nuts (numbered 3,10,14,18, and 21), were determined to be pistachio nuts using PLSR and GILS models. The PLSR and GILS results for 21 samples (Table 2) were quite compatible those obtained by HCA identification. In other words, similar identification results were obtained by HCA, PLSR, and GILS models (Table 2).

The findings from the present study showed that the slightly better prediction results for the green pea content in pistachio nuts were obtained by using the GILS algorithm when the R^2, SECV and SEP values of the PLSR and the GILS models were compared. Previous studies reported that the GILS algorithm yields a better prediction of adulterants in complex food matrixes.[33,35] The results obtained in the present research are quite compatible with those of previous reports. As can be seen from the bar plots, 16 of the industrial test samples had various amounts of green-pea adulteration, ranging from 10% to 80% by mass, whereas the results with regard to the five samples that were known to be pure pistachio nuts gave reasonably low green pea content of less than 5% for all. These results are in agreement with the SECV and SEP model values. Twenty-one samples (Table 2) were evaluated using PLSR and GILS models. While 11 samples were identified as being adulterated, ten samples were identified as pistachio nuts. Five samples, which were known to be pure pistachio nuts (numbered 3,10,14,18, and 21) were determined to be pistachio nuts using the HCA methodology.

DISCUSSION AND CONCLUSIONS

The findings from this study show that Raman spectroscopy combined with chemometrics classification and clustering methods such as PCA and HCA have a strong ability to identify green-pea adulteration in pistachio nuts. The methodology developed can be successfully applied for discriminating authentic pistachio nuts from adulterated ones, and green pea samples. In particular, HCA analysis provides a deeper understanding of the relationship between clusters, as Ward's algorithm is applied in the case of HCA analysis. Ward's algorithm is quite effective in that it can handle a large amount of information, and its use is not limited to classification problems.[30] For this reason, Ward's algorithm has been applied successfully in many previous studies in which vibrational spectroscopy combined with chemometrics methods were implemented as a solution for important adulteration problems.[36–38] A perfect classification pattern was obtained using the HCA and PCA methods when the green-pea adulteration was implemented in the range of 20–60% (w/w). Adulterated pistachio nut, green pea, and pure pistachio nut samples were correctly discriminated in detailed hierarchical cluster analysis based on their fingerprinting Raman spectra.

While the PCA and the HCA classification methods were demonstrated to accurately separate the classes and clusters of the pure pistachio nut and green pea samples, in addition to the adulterated samples with 20% and higher adulterant contents, the multivariate calibration methods (PLSR and GILS) showed the existence of successful models which can distinguish green-pea adulteration levels as low as the 5% (w/w) range. The R^2 value of the PLSR model was 0.91, while the GILS model produced a slightly higher R^2 of 0.94. In addition, the standard error of cross validation (SECV) for the calibration set and the standard error of prediction (SEP) for the independent validation set results were 5.96 and 5.67 from the PLSR model and 4.87 and 4.77 from the GILS model. The developed methodology was tested using various test samples, and favorable results were obtained. Compatible results were obtained by all the HCA, PLSR, and GILS chemometrics.

Several studies have involved the determination of green-pea adulteration in pistachio nuts. Cavus *et al.* (2018) developed a methodology for the detection of green-pea adulteration in pistachio nuts by using liquid chromatography-mass spectrometry (LC–MS) combined with chemometrics. This methodology was successful, but it was not low cost, rapid, applicable in the field and easy to operate. Expensive and toxic chemicals were also needed for this approach. Infrared spectroscopy combined with chemometrics has been shown to have a high potential for rapid, strong, effective, and reliable detection of adulterants in complex food matrices. In addition, IR techniques (MIR, NIR, Raman, hyper-

spectral imaging) are sensitive, green, non-destructive and comparatively low cost.[11] In the current study, the perfect classification pattern was observed for adulterated samples with 20% (w/w) or higher adulterant contents. Adulteration levels as low as the 5% (w/w) range were detected by the multivariate calibration methods (PLSR and GILS). These findings may help other researchers to design methods for the detection of suspicious materials in terms of food safety. The method also opened a research gate for the online determination of food adulteration using deep learning approaches. The Raman spectroscopy technique developed in this research is cost-effective, rapid, easy to operate, and non-destructive. It can be thought of as a 'green analytical technique' because no solvents and reagents were used during the study. The findings from this study contribute in several ways to our understanding of the effectiveness of Raman spectroscopy combined with chemometrics multivariate data analysis, with regard to adulteration problems in food. Raman spectroscopy, using hand-held or portable options, can be used as an effective tool for at-line or on-line analysis of adulteration in fields such as agriculture and food production.[10] The method developed in the present study could be easily adapted for the quality control of pistachio nuts in the field (in an industrial setting) because a portable Raman spectrometer was used.

ACKNOWLEDGEMENTS

This project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, under grant no. (D1441-14-135). The authors, therefore, acknowledge with thanks, DSR technical and financial support.

COMPLIANCE WITH ETHICS REQUIREMENTS

This research does not relate to any studies involving human or animal subjects.

REFERENCES

1 Johnson R, *Food Fraud and 'Economically Motivated Adulteration' of Food and Food Ingredients*. Congressional Research Service, Washington, DC, pp. 1–40 (2014).
2 Pustjens AM, Weesepoel Y and Van Ruth SM, Food fraud and authenticity: Emerging issues and future trends, in *Innovation and Future Trends in Food Manufacturing and Supply Chain Technologies*, ed. by Leadley CE. Woodhead Publishing, Cambridge, pp. 3–17 (2015).
3 Moore JC, Spink J and Lipp M, Development and application of a database of food ingredient fraud and economically motivated adulteration from 1980 to 2010. *J Food Sci* **77**:R118–R126 (2012).
4 Hong E, Lee SY, Jeong JY, Park JM, Kim BH, Kwon K *et al*., Modern analytical methods for the detection of food fraud and adulteration by food category. *J Sci Food Agric* **97**:3877–3896 (2017).
5 Şimşek M and Gülsoy E, A review on organic pistachio growth and development opportunities in Turkey. *IJAWS* **4**:89–93 (2018).
6 INC. Nuts and Dried Fruit Statistical Year Book 2017/2018. INC Int. Nut&DriedFruit. 2018.
7 Dreher ML, Pistachio nuts: composition and potential health benefits. *Nutr Rev* **70**:234–240 (2012).
8 Sezer B, Apaydin H, Bilge G and Boyaci IH, Detection of *Pistacia vera* adulteration by using laser induced breakdown spectroscopy. *J Sci Food Agric* **99**:2236–2242 (2019).
9 Cavus F, Us MF and Guzelsoy N, Assesing pistachio nut (*Pistacia vera* L.) adulteration with green pea (*Pisum sativum* L.) by untargeted liquid chromatography- (quadrupole-time of flight)-mass spectrometry method and chemometrics. *J Food Feed Sci* **19**:34–41 (2018).
10 Yang D and Ying Y, Applications of Raman spectroscopy in agricultural products and food analysis: a review. *Appl Spectrosc Rev* **46**:539–560 (2011).
11 Lohumi S, Lee S, Lee H and Cho BK, A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. *Trends Food Sci Technol* **46**:85–98 (2015).
12 Yaseen T, Sun DW and Cheng JH, Raman imaging for food quality and safety evaluation: fundamentals and applications. *Trends Food Sci Technol* **62**:177–189 (2017).
13 Rigano F, Stead S, Mangraviti D, Jandova R, Petit D, Marino N *et al*., Use of an 'intelligent knife' (iknife), based on the rapid evaporative ionization mass spectrometry technology, for authenticity assessment of pistachio samples. *Food Anal Methods* **12**:558–568 (2019).
14 Boukid F, Abbattangelo S, Carini E, Marseglia A, Caligiani A and Vittadini E, Geographical origin discrimination of pistachio (*Pistacia vera* L.) through combined analysis of physical and chemical features. *Eur Food Res Technol* **245**:143–150 (2019).
15 Czamara K, Majzner K, Pacia MZ, Kochan K, Kaczor A and Baranska M, Raman spectroscopy of lipids: a review. *J Raman Spectrosc* **46**:4–20 (2015).
16 López-Díez EC, Bianchi G and Goodacre R, Rapid quantitative assessment of the adulteration of virgin olive oils with hazelnut oils using Raman spectroscopy and chemometrics. *J Agric Food Chem* **51**: 6145–6150 (2003).
17 Zimmermann B, Bağcıoğlu M, Sandt C and Kohler A, Vibrational microspectroscopy enables chemical characterization of single pollen grains as well as comparative analysis of plant species based on pollen ultrastructure. *Planta* **242**:1237–1250 (2015).
18 Ellis DI, Brewster VL, Dunn WB, Allwood JW, Golovanov AP and Goodacre R, Fingerprinting food: current technologies for the detection of food adulteration and contamination. *Chem Soc Rev* **41**: 5706–5727 (2012).
19 Turkish Food Composition Data Base (2014). Available: http://www.turkomp.gov.tr/database?type=foods [18 April 2019].
20 Cebi N, Durak MZ, Toker OS, Sagdic O and Arici M, An evaluation of Fourier transforms infrared spectroscopy method for the classification and discrimination of bovine, porcine and fish gelatins. *Food Chem* **190**:1109–1115 (2016).
21 Czekus B, Pećinar I, Petrović I, Paunović N, Savić S, Jovanović Z *et al*., Raman and Fourier transform infrared spectroscopy application to the Puno and Titicaca cvs. Of quinoa seed microstructure and perisperm characterization. *J Cereal Sci* **87**:25–30 (2019).
22 Almeida MR, Alves RS, Nascimbem LBLR, Stephani R, Poppi RJ and De OLFC, Determination of amylose content in starch using Raman spectroscopy and multivariate calibration analysis. *Anal Bioanal Chem* **397**:2693–2701 (2010).
23 Ratnayake WS, Hoover R, Shahidi F, Perera C and Jane J, Composition, molecular structure, and physicochemical properties of starches from four field pea (*Pisum sativum* L.) cultivars. *Food Chem* **74**: 189–202 (2001).
24 Synytsya A, Copikova J, Matejka P and Machovic V, Fourier transform Raman and infrared spectroscopy of pectins. *Carbohydr Polym* **54**: 97–106 (2003).
25 Camp CH and Cicerone MT, Chemically sensitive bioimaging with coherent Raman scattering. *Nat Photonics* **9**:295–305 (2015).
26 Shanmugam K, Raman fingerprints in detection of breast cancer. *J Biosens Biomarkers Diagn* **1**:1–11 (2016).
27 Singh SK, Jha SK, Chaudhary A, Yadava RDS and Rai SB, Quality control of herbal medicines by using spectroscopic techniques and multivariate statistical analysis. *Pharm Biol* **48**:134–141 (2010).
28 Galli R, Tamosaityte S, Koch M, Sitoci-Ficici KH, Later R, Uckermann O *et al*., Raman-based imaging uncovers the effects of alginate hydrogel implants in spinal cord injury. *Adv Microsc Tech IV; Neurophotonics II* **9536**:95360Y (2015).
29 Lu F, Li S, Le J, Chen G, Cao Y, Qi Y *et al*., A new method for testing synthetic drugs adulterated in herbal medicines based on infrared spectroscopy. *Anal Chim Acta* **589**:200–207 (2007).
30 Hartel RW, Clark JP, Rodriguez-Lazaro D and Topping D eds, *Fourier Transform Infrared Spectroscopy in Food Microbiology*. Springer, New York (2012).
31 Murtagh F and Legendre P, Ward's hierarchical clustering method: clustering criterion and agglomerative algorithm. *J Classif* **31**: 274–295 (2011).

32 Tulukcu E, Cebi N and Sagdic O, Chemical fingerprinting of seeds of some salvia species in Turkey by using GC-MS and FTIR. *Foods* **8**: 1–12 (2019).
33 Özdemir D and Öztürk B, Genetic multivariate calibration methods for near infrared (NIR) spectroscopic determination of complex mixtures. *Turkish J Chem* **28**:497–514 (2004).
34 Başar B and Özdemir D, Determination of honey adulteration with beet sugar and corn syrup using infrared spectroscopy and genetic-algorithm-based multivariate calibration. *J Sci Food Agric* **98**:5616–5624 (2018).
35 Öztürk B, Yalçin A and Özdemir D, Determination of olive oil adulteration with vegetable oils by near infrared spectroscopy coupled with multivariate calibration. *J Near Infrared Spectrosc* **18**:191–201 (2010).
36 Cebi N, Yilmaz MT and Sagdic O, A rapid ATR-FTIR spectroscopic method for detection of sibutramine adulteration in tea and coffee based on hierarchical cluster and principal component analyses. *Food Chem* **229**:517–526 (2017).
37 Wang LH, Chen JX and Wang CC, Rapid quantitative analysis of suspected fragrance allergens in between commercial essential oils and using attenuated total reflectance-infrared (ATR-IR) spectroscopy. *J Essent Oil Res* **26**:185–196 (2014).
38 Muik B, Lendl B, Molina-Díaz A, Ortega-Calderón D and Ayora-Cañada MJ, Discrimination of olives according to fruit quality using Fourier transform Raman spectroscopy and pattern recognition techniques. *J Agric Food Chem* **52**:6055–6060 (2004).

