

## Use of Chemical Profiling to Differentiate Geographic Growing Origin of Raw Pistachios

KIM A. ANDERSON<sup>\*,†,‡</sup> AND BRIAN W. SMITH<sup>†</sup>

Department of Environmental and Molecular Toxicology, and Food Safety and Environmental Stewardship Program, Oregon State University, Corvallis, Oregon 97331

The objective of this study was to demonstrate the feasibility of chemical profiling methods combined with multivariate methods to differentiate the geographical growing regions of pistachios (*Pistachia vera*). Elemental analysis (Ba, Be, Ca, Cu, Cr, K, Mg, Mn, Na, V, Fe, Co, Ni, Cu, Zn, Sr, Ti, Cd, and P) of pistachios samples was performed using inductively coupled plasma atomic emission spectrometry. Analysis of inorganic anions and organic acids (selenite, bromate, fumarate, malate, selenate, pyruvate, acetate, phosphate, and ascorbate) of pistachio samples was performed using capillary electrophoresis. Bulk carbon and nitrogen isotope ratios were performed using stable isotope MS. There were nearly 400 pistachio samples analyzed from the three major pistachio growing regions: Turkey, Iran, and California (United States). A computational evaluation of the trace element data sets was carried out using statistical pattern recognition methods including principal component analysis, canonical discriminant analysis, discriminant analysis, and neural network modeling. Several linear discriminant function models classified the data sets with 95% or higher accuracy. We report the development of a method combining elemental analysis and classification techniques that may be widely applied to the determination of the geographical origin of foods.

**KEYWORDS:** Geographic authenticity; canonical discriminant analysis; discriminant analysis; principal component analysis; metals; anions; organic acids; isotope ratios; *Pistachia vera*; geographic origin

### INTRODUCTION

Public health security and bioterrorism preparedness include protecting a nation's food supply. Establishing and maintaining knowledge about the origin of food are important components of securing the food supply. Authenticating food specimens to specific lots (shipments) or geographic regions would be part of ensuring a safe food supply and an important tool for forensic investigations or detention of foods that may pose a risk to public health. Chemical trace element compositional analysis of foods provides a scientific foundation to geolocate commodities (foods) on the basis of their chemical compositions.

Increasing demands on the agrifood industry from free trade, globalization, and changing technology only further the drive to determine the authenticity of foods (1). Financial incentives continue to drive retailers/resellers to misidentify geographic origin of commodities (2) and food products (3–7). Although recently publications in this area have begun to develop, geographic classification has focused primarily on processed foods, particularly wines (3, 8) and juices, and to a lesser extent drugs of abuse (5), cocoa (6, 7), and olive oil (9, 10). Most research literature on the geographic origin of commodities involves the analysis of vitamins or other organic molecules

(amino acids, triglycerides, volatile aromatic compounds, etc.). Some success (60–90% correct classification) has been reported using vitamin and/or amino acid assays to determine geographic origin (11–16). However, a shortcoming of using vitamins (or other organic compounds) is their susceptibility to degradation (including enzymatic changes) from the time of harvest through storage to the time of analysis. Storage conditions may be especially important for some vitamin assays; for example, vitamin E is light sensitive, and changes in vitamin E content during storage have been reported (17). It is important, therefore, if one wants to develop a technique that will ultimately be used to determine the geographic origin of unknown samples, that effects from storage conditions be minimized.

The determination of the geographic origin of wines has been an active area of research (11, 18–21). Wine geographic classification accuracy rates have ranged between 73 and 91% (11, 18–21), although a study by Day et al. (22) combined the analyses of <sup>2</sup>H NMR with multiple elemental and isotopic ratio determinations. Although this technique correctly classified with better than 99% accuracy, the approach requires the use of several instruments, including SNIF-nuclear magnetic resonance, elemental analyzer-isotope ratio mass spectrometry, flame atomic absorption spectrometry, electrothermal atomic absorption spectrometry, and inductively coupled plasma atomic emission spectrometry (ICPAES). In addition to the multiple analyses and instruments required used in Day's study (22),

\* To whom correspondence should be addressed. E-mail: Kim.anderson@orst.edu.

<sup>†</sup> Department of Environmental and Molecular Toxicology.

<sup>‡</sup> Food Safety and Environmental Stewardship Program.

sophisticated techniques were necessary for the determination of the five isotopic ratios used. Overall, many authenticity studies are surveyed in nature (less than 30 samples); therefore, conclusions concerning the effectiveness of these techniques should be made prudently.

Pistachio trees (*Pistacia vera*) are believed to have originated in Central Asia. They were brought to the Mediterranean Basin about 2000 years ago, and more recently, in the 1850s, they were introduced to California (United States). Most countries produce a couple of varieties while California produces only one variety (Kernan). Over 85% of the pistachios are grown in Iran (ca. 50%), the United States (California) (ca. 25%), and Turkey (ca. 10%). The world export market is dominated by Iran (86%), with the United States ranking second at 12%. Variation in quality, food safety (e.g., aflatoxins), import/export fees, legal implications, and financial concerns makes determining the country of origin for pistachios important, especially since the world pistachio export market is valued at over \$500 million U.S. dollars. Because each country's applied tariff rates on commodities vary dramatically from <1 to >40%, pistachio producers and traders are motivated to discover objective chemical techniques for determining the geographic origin of pistachios.

It is recognized that mineral and trace metal compositions of fruits and vegetables are a distorted reflection of the trace mineral compositions of the soil and environment in which the plant grows (23). The soil-plant system is highly specific for different elements, plant species, and environmental conditions. Under most conditions, a trace element present in the vegetable/fruit must have existed in the rooting zone of the plant, at least in a slightly soluble form. Trace elements must also pass through at least one cellular membrane in its movement from soil to plant. The selectivity of these processes of mineral bioaccumulation within the vegetable/fruit varies with different trace elements, with different plants, and with the unique environment in which the commodity is grown.

The purpose of this study was to determine the feasibility of differentiating raw pistachios grown in the three different major regions of the world, Iran, Turkey, and California, using chemical profiling together with multivariate analysis. In addition, feasibility of further separation of these subregions and effects of seasons were investigated through multivariate analysis. Here, we present data from three different sites and represent over a 6000 chemical parameter data set.

## MATERIALS AND METHODS

**Reagents.** The sources of chemicals and reference materials were as follows: concentrated nitric acid, trace metal analysis grade (J. T. Baker, St. Louis, MO); elemental stock standard solutions (J. T. Baker); certified reference materials (CRM), NIST 1575 Pine Needles, NIST Oyster Tissue 1566a, NIST Rice Flour 1568a, NIST 1577b Bovine Liver, NIST 8433 Corn Bran (National Institute of Standards and Technology, Gaithersburg, MD); and NRC TORT-2 Lobster Hepatopancreas (National Research Council Canada, Institute National Measurements Standards, Ottawa, Ontario, Canada).

**Apparatus.** The ICPAES was equipped and setup as follows: Varian model (Palo Alto, CA) Liberty 150 ICPAES; PMT, 650 V; nebulizer, 85 psi; auxiliary, 1.5 L/min; pump rate, 13 rpm; two integrations; 1.0 s scan integration time; acid flexible tubing, 0.030  $\mu\text{m}$  ID (internal diameter); wavelengths and background corrections have been previously presented (24, 25). A temperature controller/digester used was a Lab-line microprocessor digester block and controller. The capillary electrophoresis (CE) was equipped and setup as follows: Hewlett-Packard model (Palo Alto, CA) HP 3D CE; diode array detector, 50  $\mu\text{m}$  ID  $\times$  64.5 cm fused silica extended bubble light path capillary column; sample injection, 50 mbar; 2 s; applied voltage, -25 kV;

capillary temperature, 16  $^{\circ}\text{C}$ ; detection at 350 nm and reference at 225 nm. The analytical method time was 7 min.

Nitrogen ( $\delta^{15}\text{N}$ ) and carbon ( $\delta^{13}\text{C}$ ) stable isotopes were measured on a stable isotope mass spectrometer (MS) (Finnigan MAT-251). Isotopic data use the standard isotopic  $\delta$  notation ( $\delta$ ), in per mil (‰) relative to the Pee Dee Belemnite (PDB) scale. Calibration to PDB was done through the NBS-19 and NBS-20 standards of the National Institute of Standards and Technology (MD). External precision estimates of  $\delta^{15}\text{N}$  and  $\delta^{13}\text{C}$ , based on replicate analysis of acetanilide and oxalic acid standards, were  $\pm 0.12\%$  and  $\pm 0.11\%$ , respectively.

**Sampling, Preparation, and Analysis.** Samples were collected on-site in Turkey and Iran and shipped directly to our laboratory. Chain-of-custody was maintained for all samples. California samples were provided by the California Pistachio Commission. Specific subregions/cities, varieties, and season information were known for all samples analyzed. Each pistachio sample was analyzed as the whole nut (no shell). Samples were analyzed on a dry weight bases. For elemental analysis, pistachio samples were digested. A ca. 1.0 g sample was taken, representing one nut, and the sample was digested with 3.0 mL of nitric acid (trace metal grade) in a 10 mL graduated Kimax culture tube on a programmed heating block. The samples were allowed to react for ca. 4–8 h in a hood at ambient temperature. Then, the samples were digested using a programmable heating block. The samples were ramped to 140  $^{\circ}\text{C}$  over an hour and then maintained at 140  $^{\circ}\text{C}$  for 3–4 h. Digestion was confirmed complete when no nitrous oxide gases were evolved. The samples were diluted with type 1 water (18 Mohm cm) and mixed thoroughly using a vortexer. The samples were filtered through a 0.45  $\mu\text{m}$  filter prior to analysis. Analysis was by ICPAES. For anions (inorganic and organic acids), pistachio samples were extracted. The background electrolyte used was 2,6-pyridinedicarboxylic acid (PDC), and 0.5 mM hexadecyltrimethylammonium bromide was used with 5 mM PDC at a pH of 5.6. All samples were filtered prior to analysis through a 0.22  $\mu\text{m}$  filter. Samples for isotope analysis were dried overnight at 60  $^{\circ}\text{C}$ , ground to a fine powder, and loaded in capsules for MS analysis.

The chemical analytical technique is well-suited to analysis of modest to small samples. A minimum of 500 mg can be used; 1 g was used in this study. Dilution factors are minimized here; only a factor of 10 as compared to typical digestions that involve dilution factors of 50 or more is used. This small dilution factor permits determination of additional elements that would otherwise be below instrument detection limits. In addition, as a pollution prevention mechanism, this technique uses fewer reagents and in small volumes; thus, this technique reduces waste.

**Quality Control.** Each analytical batch contained a minimum of 25% quality control samples, including check standards, duplicates, spikes, and CRMs. Each individual analyte was quantitated based on a calibration curves consisting of 3–4 standard levels each, with correlation coefficients of  $>0.98$ . During the course of the study, over 50 CRM samples were analyzed; CRMs were dominantly plant matrices where available. Typical percent standard deviation (%SD) was  $<10\%$ , although analytes close to method detection limits had higher %SDs. Spike recoveries and check standards were typically within  $\pm 10\%$  of their true value.

## STATISTICAL AND MULTIVARIATE ANALYSIS

Univariate and multivariate data exploration and analysis methods were used to investigate the feasibility of characterizing pistachio growing origin based on the measured profile of trace element concentrations found in the pistachios. The methods employed include principal component analysis (PCA), canonical discriminant analysis (CDA), linear discriminant function analysis, and neural network modeling. The computational analysis was similar to that done to determine the feasibility of geographic classification of potatoes (2) and coffee (26). Initial trace element screening included all of the elements analyzed. Elements not showing significant concentrations were not used in the subsequent data analysis. The ultimate elements included Ca, Cu, Fe, K, Mg, Mn, Na, P, Sr, V, and Zn. Univariate

statistics were also computed for the data grouped by growing season for a given location. Seasonal comparisons could only be made within the location groups for California and Iran since these locations had two seasons of data while Turkey had data only for the 2001 growing season. The univariate statistics mean, SD, and sample number were computed for each element by location. These data were displayed visually with box plots. Additional information included in the box plots was upper-quartile, lower-quartile, standard error, maximum, and minimum.

**Multivariate Methods.** *Standardization.* To ensure that different measurement scales did not affect the multivariate analysis, the data were standardized by subtracting sample means and then dividing the resulting difference by the corresponding SDs. The software used to perform the multivariate analysis was SAS version 8e for Windows.

*PCA.* PCA generates principal components (PCs) that are linear combinations of the original variables. The first PC summarizes the maximum possible variation that can be projected onto one dimension; the second PC captures the second most and so on. The PCs are orthogonal in the original space of variables, and the number of PCs can equal the number of the original variables. However, it is sometimes the case that a large percentage of the total variation can be explained by the first few PCs, effectively reducing the number of variables needed to describe variation between individual samples. In this case, plotting the samples with respect to two or three PCs facilitates two- or three-dimensional views of how individual samples differ from one another (in the variation sense). For a geographic classification task, it is desirable to have group differences explicitly manifest with a low-dimensional view. However, this is not always the case since this method measures variation in the elemental concentrations in the samples but does not take into account group (geographic origin) membership. To get the best possible view of group clustering, we used CDA.

*CDA.* CDA is a dimension reduction technique related to PCA, but unlike PCA, predefined groups are included in the calculations. CDA generates canonical variables, which are linear combinations of the original variables that describe the variation between prespecified classes in a manner analogous to the way in which PCA summarizes the variation between individual samples. CDA can effectively reduce the number of variables and provide optimum low-dimensional views of the data, which display the maximum possible variation between different groups and the minimum possible variation within the same group. The number of possible canonical variables is the minimum of the number of classification groups minus one and the number of independent variables. CDA has previously been applied to data for the purpose of geographical classification of potatoes (2), coffee (26), and wine (22).

**Classification Models.** *Discriminant Function Analysis.* Discriminant function analysis here refers to a group of pattern recognition classification methods that use known data to determine a discriminant function, which can then be used to classify unknown samples into predetermined classes. Two types of discriminant functions were used for this study: a linear discriminant function and a quadratic discriminant function. Details on how each of these work can be found in the description of the DISCRIM procedure in the SAS technical manual (27). To estimate classification accuracy, a cross-validation error rate was calculated. A discriminant function is constructed, leaving out one sample as an "unknown". The sample is then classified. This process is repeated for every

sample, and the cross-validation classification accuracy estimate is obtained by taking the percentage of samples classified correctly.

## RESULTS AND DISCUSSION

**Chemical Analysis.** Whereas other geographic authenticity approaches require the use of several instruments, an important attribute of this study's approach is that all of the elemental chemical data can be determined with the use of a single analytical instrument, ICPAES. In this approach, the data are used directly from the ICPAES into the computational models and require no prior mathematical or interpretive data analyses as is not often the case with other geographic authenticity approaches. In this study, 18 elements were determined, unlike chromatography techniques, elemental spectroscopy data analysis requires little analyst time and only moderate expertise. In contrast, the organic acid and inorganic anion data from the capillary electrophoresis technique did not provide data that was usable for geographic profiling because the variations in organic acid and inorganic anion concentrations within each region were much larger than any differences seen between geographic regions.

Of the 15 elements tested, 11 were routinely above detection limit; see **Table 1**. Beryllium, barium, titanium, and zirconium were typically near or below detection limits. Strontium had the largest concentration difference within the geographic regions and samples tested. All Iranian pistachio samples, all regions and all varieties, had high strontium concentrations relative to the other geographic samples analyzed. Generally, Iranian pistachios had strontium concentrations  $>20 \mu\text{g/g}$ . One set of samples from Turkey also had high strontium, but the other Turkey samples and all of the California samples had strontium concentrations near or below the detection limit. Iran and Turkey are two of the top four producers of strontium, which is mined as Celestine ( $\text{SrSO}_4$ ). Although there are anthropogenic sources of strontium, it might be reasonable to expect strontium uptake in pistachios grown in regions of strontium production and export.

Plant interactions between strontium and calcium are complex. Although strontium may compete with calcium, strontium usually cannot replace calcium in biochemical functions. Calcium-to-strontium ratios have been proposed by some authors for better understanding of source and uptake of cations (28). The calcium/strontium ratios in the samples tested, however, did not provide any additional discriminating power beyond simply comparing strontium concentrations.

Unlike previous studies (2, 26), the plant macroelements, calcium, potassium, magnesium, and phosphorus, had some differences based on geographic growing area, although singly these data have limited applications. The Californian samples generally had calcium concentrations a factor of 2–3 less than Iranian or Turkish pistachios. Within the 2001 season, for the geographic regions and samples tested ( $n = 270$ ), calcium ranged about a factor of 3.5 between geographic regions. Potassium was typically lowest in the Californian samples, ranging from a factor of 0.3 to 2 lower than Iranian or Turkish samples. Potassium was generally highest in Iranian pistachios, although some subregions and varieties had lower potassium concentrations (discussed in more detail below). Magnesium was typically lowest in Californian samples from a factor of 0.5 to 2 less than Iranian or Turkish pistachios. Magnesium was generally highest in Iranian pistachios, for the 2001 season. Phosphorus ranged a factor of 2 among all geographic regions tested. From a three-dimensional plot of potassium, magnesium,

**Table 1.** Mean Concentrations ( $\mu\text{g/g}$ ) and SDs Dry Weight for 11 of the 15 Elements Determined in Pistachios

country/ subregion	variety	<i>n</i>	Ca avg (SD)	Cu avg (SD)	Fe avg (SD)	K avg (SD)	Mg avg (SD)	Mn avg (SD)	Na avg (SD)	P avg (SD)	Sr avg (SD)	V avg (SD)	Zn avg (SD)
Iran 2001													
central	Kaleh ghochi	20	1752 (663)	9.4 (1.7)	31.3 (7.6)	15891 (2615)	1984 (295)	10.7 (3.7)	137.0 (64.6)	8164 (1763)	23.4 (8.9)	7.2 (1.3)	24.2 (6.4)
central	Fandoghi	20	2890 (746)	7.1 (1.8)	40.8 (10.2)	9164 (1465)	1801 (222)	15.4 (4.3)	82.5 (31.6)	7838 (1332)	22.4 (8.6)	6.2 (1.1)	36.9 (12.4)
north	Fandoghi	20	1079 (298)	13.1 (2.4)	42.2 (13.7)	17549 (3087)	1476 (177)	10.5 (2.3)	248.3 (89.1)	5672 (694)	25.4 (5.8)	20.8 (1.5)	26.7 (6.5)
north	Fandoghi	20	3794 (953)	7.5 (2.7)	47.7 (13.9)	10174 (821)	1596 (161)	14.2 (5.1)	18.5 (6.6)	9467 (1865)	49.5 (14.5)	4.9 (0.7)	36.3 (16.4)
north	Fandoghi	20	2313 (1149)	11.2 (2.9)	38.9 (8.2)	10323 (1303)	2053 (177)	15.6 (4.2)	64.5 (31.6)	9681 (1430)	35.0 (12.8)	9.2 (0.9)	35.4 (8.8)
north	Kaleh ghochi	20	2097 (640)	8.5 (1.7)	33.0 (6.5)	10270 (1780)	1416 (139)	9.4 (2.5)	79.5 (28.1)	6033 (1054)	27.4 (6.7)	6.1 (0.8)	23.0 (4.4)
south-central	Kaleh ghochi	20	1547 (577)	8.3 (1.7)	35.2 (12.8)	8591 (1748)	1384 (226)	10.6 (4.7)	13.1 (7.0)	7292 (1592)	15.5 (3.7)	15.9 (2.6)	22.8 (8.5)
Iran 2000													
central	Fandoghi	20	1367 (526)	8.3 (5.2)	24.7 (10.7)	9801 (6413)	1697 (837)	11.3 (5.4)	27.7 (31.6)	8063 (1797)	31.6 (21.6)	9.7 (1.9)	33.0 (11.1)
central	Fandoghi	20	1282 (632)	6.0 (1.8)	28.8 (12.2)	7286 (3936)	1667 (455)	11.2 (5.4)	27.2 (19.6)	6622 (1583)	29.6 (10.4)	8.5 (2.0)	27.0 (11.7)
central	Kaleh ghochi	20	1087 (456)	9.1 (2.0)	25.1 (4.6)	15212 (5089)	1063 (163)	6.7 (1.9)	33.6 (5.8)	5534 (742)	24.0 (6.0)	15.0 (1.4)	16.2 (3.3)
south-central	Kaleh ghochi	20	3090 (620)	6.9 (2.4)	47.6 (6.2)	9863 (7332)	1674 (169)	17.5 (3.8)	26.5 (9.9)	8323 (1157)	25.3 (5.5)	6.0 (0.7)	36.9 (13.0)
south-central	Kaleh ghochi	20	2349 (709)	6.8 (1.4)	41.2 (5.9)	10981 (1284)	1527 (187)	14.6 (3.0)	46.6 (21.7)	6959 (1604)	27.7 (9.2)	6.2 (0.9)	24.3 (7.5)
Turkey 2001													
east	Siirt	18	2934 (810)	9.9 (1.4)	32.7 (10.3)	10623 (1248)	1688 (211)	8.6 (2.3)	17.3 (2.3)	6946 (1345)	6.7 (5.2)	6.8 (2.5)	17.1 (4.1)
central	Siirt	20	1453 (519)	12.0 (3.1)	29.2 (9.9)	10339 (1348)	1402 (154)	8.4 (2.0)	14.7 (2.2)	6847 (844)	1.4 (1.0)	6.3 (0.9)	20.9 (3.6)
central	Siirt	20	2514 (660)	8.5 (1.8)	24.0 (7.6)	10340 (1418)	1604 (252)	10.2 (3.8)	20.7 (3.3)	5148 (822)	24.5 (11.6)	6.6 (1.1)	18.8 (6.7)
central	Keten gomlegi	21	2899 (594)	9.0 (1.9)	36.5 (11.1)	9020 (1848)	1448 (178)	13.1 (2.7)	11.9 (3.3)	7174 (1141)	<1 (na)	4.3 (1.0)	26.9 (7.1)
United States													
CA 2000	Kernan	20	1123 (566)	16.2 (7.7)	54.7 (15.1)	9950 (4143)	1567 (634)	11.6 (4.7)	18.4 (9.5)	7401 (2045)	5.7 (1.0)	5.8 (1.9)	27.3 (10.6)
CA 2001	Kernan	30	1064 (393)	12.0 (4.9)	37.0 (10.3)	7896 (2033)	1049 (256)	11.8 (4.1)	15.0 (5.0)	6654 (1944)	2.0 (1.7)	4.1 (1.5)	23.6 (7.4)

and strontium, one can see that geographic origin begins to separate in comparisons; see **Figure 1**. However, individually, none of these elements alone appear to have discriminating power for the geographic regions tested. For examples, see a selection of box plots in **Figure 2**.

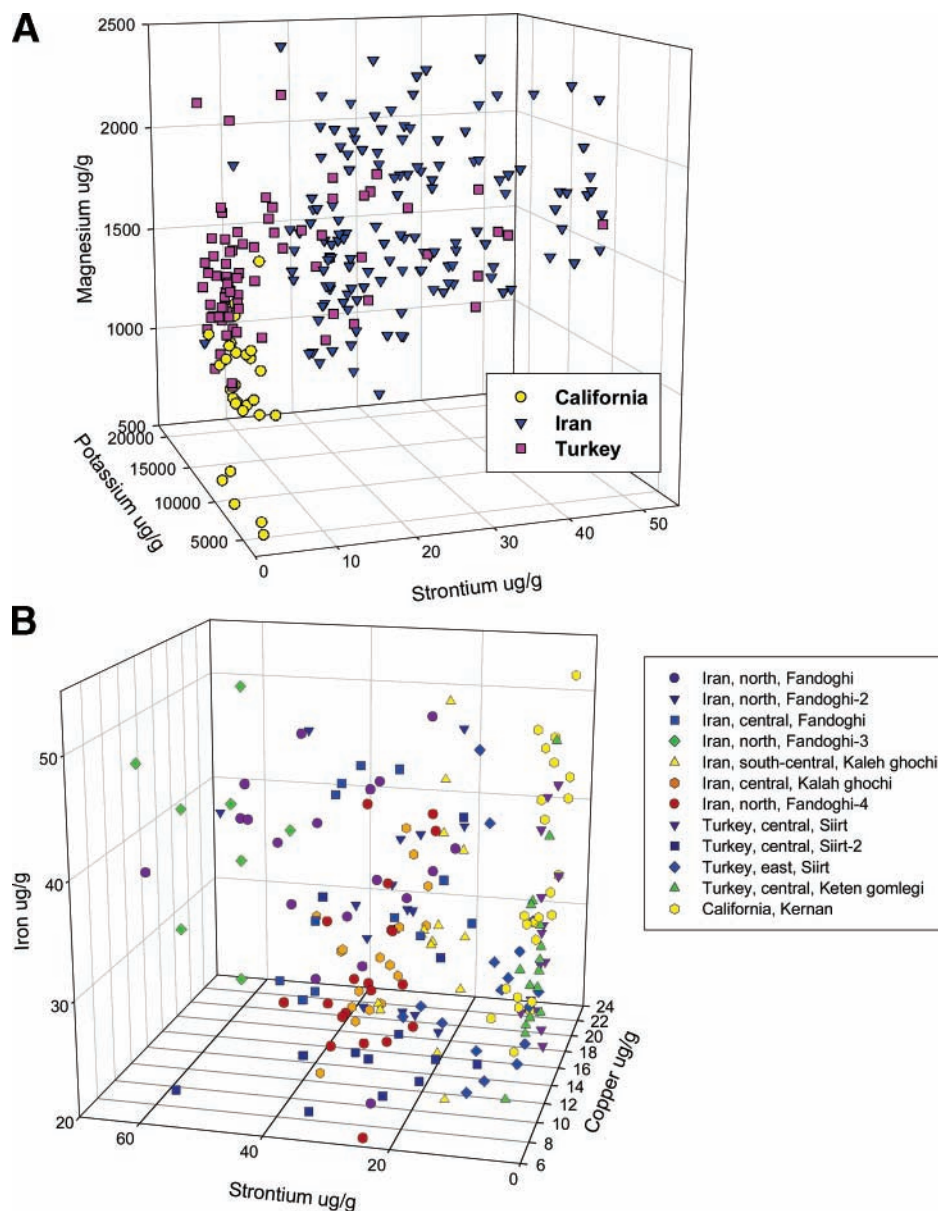
The plant microelements copper, iron, manganese, vanadium, and zinc also have some discriminating power with the geographic regions tested, but again, one cannot determine origins with these elements singly. More sophisticated computational analysis indicates that these data have value increasing modeling success, discussed below. Although individually no element is diagnostic of origin, **Figure 1** illustrates by combining elements that there is better discrimination among some geographic regions. For example, **Table 1** illustrates that copper ranged from 7 to 13  $\mu\text{g/g}$  in the 2001 pistachios tested ( $n = 270$ ); over two seasons ( $n = 371$ ), copper ranged from 6 to 13  $\mu\text{g/g}$ . Both the lowest and the highest copper concentrations occurred in Iranian pistachios in the same variety but in different subregions in Iran. Iron ranged from 24 to 48  $\mu\text{g/g}$ , a factor of 2 difference between geographic regions for the 2001 season. Manganese ranged from 9 to 15  $\mu\text{g/g}$ , a factor of only 1.5 difference between geographic regions. Vanadium for the 2001 season pistachios ranged from 4 to 21  $\mu\text{g/g}$ . The highest vanadium concentrations were in Iranian pistachios, and the lowest vanadium concentrations were in Turkish and Californian pistachios. Zinc concentrations for the 2001 samples ranged from 17 to 37  $\mu\text{g/g}$ . Generally, the highest zinc concentrations were found in Iranian pistachios while the lowest concentrations were found in pistachios from Turkey. For example, from a three-dimensional plot of strontium, iron, and copper, one can see that origins are beginning to separate (**Figure 1B**). With more dimensions and modeling, better separations are possible.

A selection of box plots appears in **Figure 2**. An analysis of variance was carried out, and in all cases, the location group means were found to be different. Some interesting differences by groups were discerned visually by looking at the box plots. However, again, the distributions do overlap and no clear-cut rule for group classification can be determined from this analysis alone.

Another important result of the element concentration distribution is that no one region is responsible for all of the high

or low concentrations. For example, Iranian pistachios had the highest average calcium, potassium, magnesium, strontium, vanadium, and zinc and Californian pistachios had the highest average copper. Turkey had the lowest iron, manganese, and zinc, and California had the lowest calcium, potassium, magnesium, and strontium. Overall, with so many differences, the use of computational modeling with elemental concentrations was a powerful tool.

Seasonal variability was guardedly investigated here. A select group of box plots comparing the distributions of each element by season (for a given region) are shown in **Figure 2A**. Analysis of variance was carried out, and seasonal group means were found to be different. In general, the trace metals were lower in 2000 than 2001, although additional data would be needed to evaluate seasonal effects on metal uptake. Although the same geographic regions (Iran and California) and many of the same subregions (Iran-central and south-central) (see **Table 1**) were sampled in both the 2000 and the 2001 seasons, the exact same farms/trees were not systematically resampled; therefore, geographic differences may still contribute to differences seen between seasons. Strontium, the most discriminating element, was similar between seasons. Iranian pistachios in 2000 were  $>25 \mu\text{g/g}$  and in 2001 were generally  $>25 \mu\text{g/g}$ , while both Californian seasons (2000 and 2001) were  $<6 \mu\text{g/g}$ . Calcium in Iranian samples for both seasons was  $\geq 1100 \mu\text{g/g}$ , and Californian pistachios were  $\leq 1100 \mu\text{g/g}$ . Copper values in Iranian pistachios in 2000 and 2001 were all  $\leq 9 \mu\text{g/g}$ , while in both Californian seasons, the copper values were  $\geq 12 \mu\text{g/g}$ . Other elements, such as zinc, vanadium, and magnesium, although less dramatic, were somewhat different between the seasons. Overall, the 2001 Californian metal concentrations were lower than the 2000 metal concentrations; see the box plot in **Figure 2A** for an example typical of the data trends. In contrast, the 2001 Iranian metal concentrations were higher than the 2000 metal concentrations. These seasonal trends for California and Iran were consistent for all metals tested. In a previous study with over 2000 potato samples collected over several seasons, only small variations (2) between seasons were noted. Without sufficient seasonal data to demonstrate otherwise, the generation of databases for each season is likely necessary to ensure good predictability of any model used routinely.

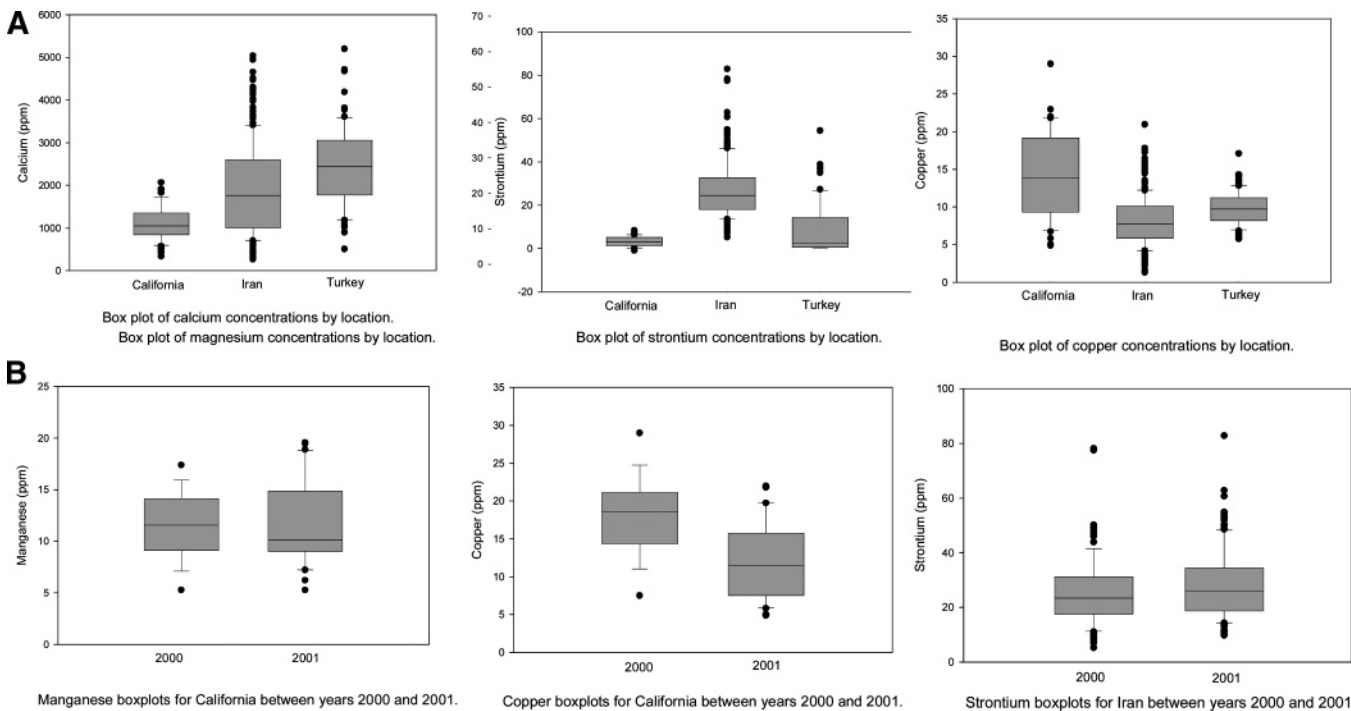


**Figure 1.** (A) Concentration of strontium, potassium, and magnesium vs geographic growing origin. All varieties and two growing seasons are shown ( $n = 371$ ). Three-dimensional trace element profile of regional origins of pistachios. (B) Concentration of strontium, copper, and iron for the 2001 season. Subregions and varieties are shown.

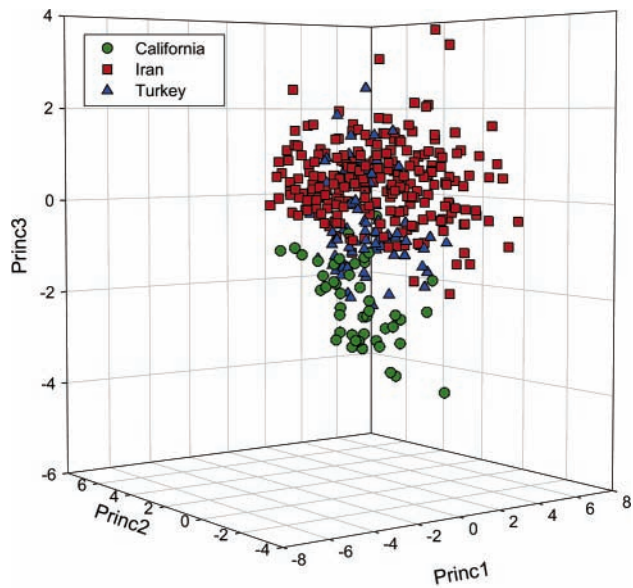
Variety differences were difficult to interpret, although within the same region and subregion, pistachios of different varieties were also grown in different orchards so geographic differences in subregions still exist to some extent (**Figure 1B**). The Fandoghi variety was collected from three farms in the north region of Iran in 2001 and two farms in central Iran in 2000. Keeping seasons separate, the 2000 Fandoghi were all quite similar for calcium, copper, iron, potassium, magnesium, sodium, phosphorus, strontium, vanadium, and zinc. The Fandoghi from the 2001 season, in northern Iran, had a larger variation among all elements for this variety. The Koleh ghochi variety for the 2000 season was collected at two south-central Iranian orchards; about half of the elements tested were similar while the remainder showed small variations.

**Isotope Ratios.** Isotope ratios have been used as another chemical profiling method to determine geographic origin of biota (29) or biota-derived products (e.g., crude oil) (30). Chemical, physical, and biological processes can have significant isotope fractionations. The principal of employing stable carbon isotope methods is that the distributions of isotopes in biological

matter are a function of photosynthetic fixation, temperature, plant type (e.g., C3 v C4 plants) (31) and/or the environment (e.g., including latitude). For example, the  $^{13}\text{C}/^{12}\text{C}$  ratios are different depending on the geography and climate. Plants have enzyme(s) that discriminate against the rarer  $^{13}\text{C}$  isotopes. Plants in humid environments, for instance, take in more  $\text{CO}_2$ , and they develop a lower ratio of  $^{13}\text{C}$  to  $^{12}\text{C}$  than plants in drier environments. Processes affecting nitrogen isotopic composition include N-fixation, assimilation (e.g., uptake of ammonium, nitrate, etc.), mineralization, nitrification, volatilization, sorption/desorption, and denitrification. Across a broad range of climate and ecosystem types, soil and plant  $\delta^{15}\text{N}$  values systematically have been reported to decrease with increasing mean annual precipitation and decreasing mean annual temperature (32). Globally, plant  $\delta^{15}\text{N}$  values are more negative than soils, suggesting a systematic change in the source of plant available N (organic/ $\text{NH}_4^+$  vs  $\text{NO}_3^-$ ) with climate (32). Spatial variability in foliar  $\delta^{15}\text{N}$  has been observed within forested catchments (33). A compilation of data for nonfixing trees showed a 3–15‰ range in  $\delta^{15}\text{N}$  values among the same species relative



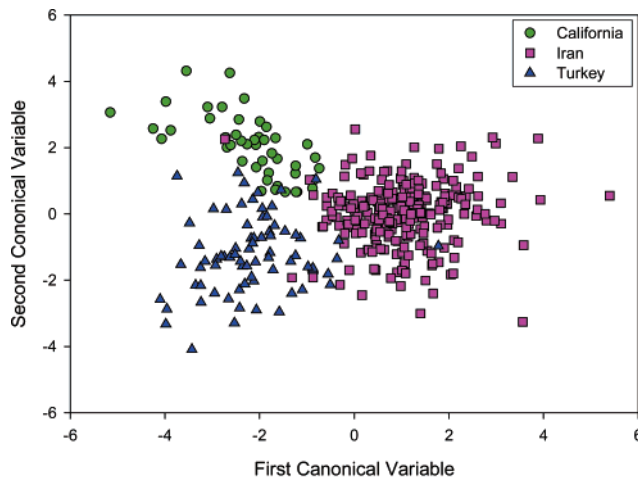
**Figure 2.** (A) Box plots of element metals in pistachios from different growing regions. (B) Box plots of elements in pistachios from different growing regions and in the 2000 and 2001 seasons.



Plot of location using principle components.

**Figure 3.** Score plots of the first three PCs for trace elements in pistachios from different pistachios from different growing regions.

to small geographic areas (33). The large range in  $\delta^{15}\text{N}$  reflects spatial variability in the relative amounts and bioavailability of atmospheric vs various soil sources of N (34). Carbon and nitrogen isotopes were determined and are reported as  $\delta^{13}\text{C}$ , measured as  $\text{CO}_2$  and  $\delta^{15}\text{N}$  measured at  $\text{N}_2$ . The total carbon/nitrogen ratios were also measured, and they are different for the three regions tested ( $n = 71$ ). **Figure 6A** shows the C/N ratio vs  $\delta^{15}\text{N}$ , and there appears to be a fair amount of separation between the geographic regions. The  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  were evaluated, and there is some separation based on geographic region (**Figure 6B**). Further mathematical manipulation or modeling was not needed with this isotope data set as the geographic origin groups separate simply by plotting the

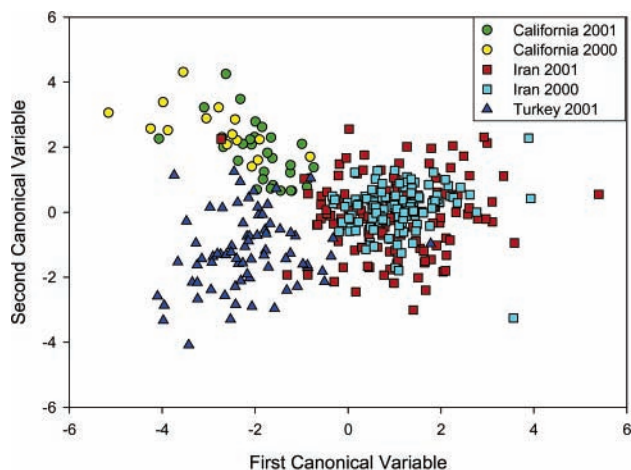


Plot of location with respect to the first two canonical variables

**Figure 4.** Score plots of the first two canonical variables used to discriminate trace elements in pistachios from different growing regions.

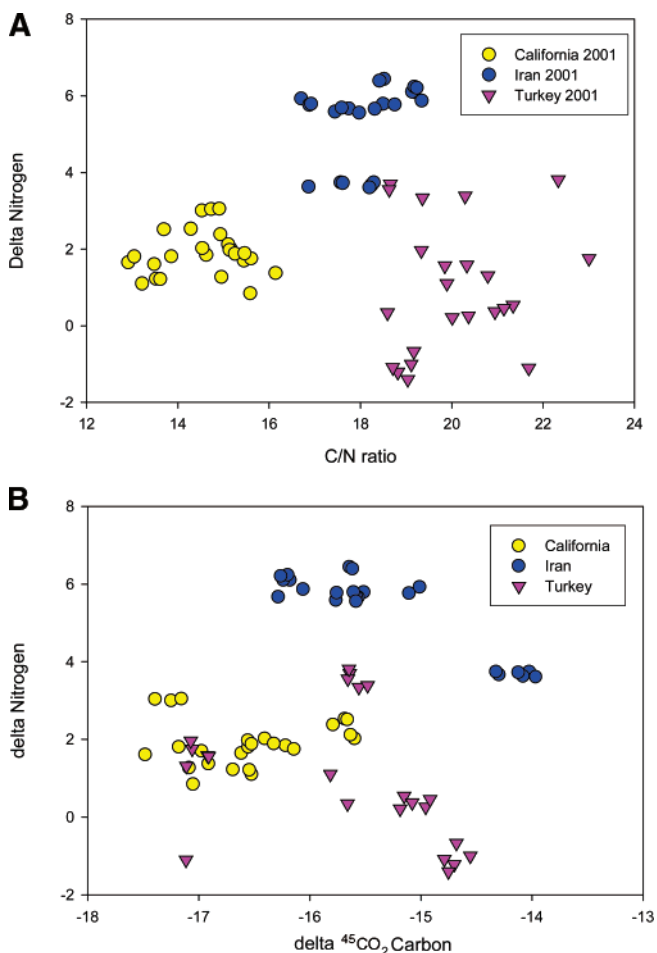
chemical data. The grouping of the five Turkish samples with smaller ( $-17$ )  $\delta^{13}\text{C}$  values as compared to other Turkish samples was from the same variety and from the same region (e.g., Turkey, central, Siirt). The grouping of five Iranian samples with a larger value  $\delta^{13}\text{C}$  ( $-14$ ) as compared with the other Iranian samples was also from a single variety and region (e.g., Iran, north, Fandoghi). Samples from all groups were rerun, and they strongly duplicated within the groups shown, including the five sample groups discussed above. The  $\delta^{13}\text{C}$  are apparently highly selective to the growing regions and conditions. This also illustrates that had these samples not been included in the database the chemical graphical separation determined could have been misleading.

**Multivariate Analysis.** For initial data exploration, PCA was applied to the trace metal data. A total of 372 pistachios samples representing three geographic regions with 20 samples from each were analyzed. A total of 270 samples were collected for the



Plot of location using the first to canonical variables.

**Figure 5.** Score plots of the first two canonical variables used to discriminate trace elements in pistachios from different growing regions.



**Figure 6.** (A) Plot of location using the carbon nitrogen ratios and  $\delta$  nitrogen for pistachios in 2001. A subset of the Iranian samples ( $n = 23$ ) from four different locations, a subset of Turkish samples ( $n = 23$ ) from four different locations, and a subset of Californian samples ( $n = 25$ ) for the 2001 season are shown. (B) Plot of location using the  $\delta$  nitrogen and  $\delta$  carbon ratios for pistachios in 2001. The subsets of the three geographic regions ( $n = 71$ ) and the samples analyzed are the same as in panel A.

2001 season from the three regions; see **Table 1**. The 2000 and 2001 element data were used for the computational analysis described herein. To adjust for different scales of measurement

between trace elements, the data were normalized by subtracting the elemental means from each entry, and then, each resulting difference was divided by the corresponding SD. Thus, each trace element had an adjusted mean of zero and an adjusted SD of one.

**PCA.** PCA has been successfully applied to geographical classification applications of various foods including processed orange juice (35), wine (20, 22), honey (36), and cocoa (6, 7). Sample scores with respect to the first three PCs are plotted in **Figure 3**. Visual separation by growing region is not necessarily expected since PCs are measures of total sample variation and do not explicitly take into account variation between groups (locations) of interest. Some visual separation of samples from California and a combination of the samples from Turkey and Iran is observed however along the third PC. The first PC accounted for about 42% of the total variation. The second and third PCs accounted for about 17 and 14% of the total variation, respectively. The remaining seven PCs accounted for remaining 27% of the total sample variation. For PC three, the most important metals were determined to be Sr, Fe, and Cu. For the second PC, the most important metals were K, Na, and Cu, and for the third PC, the most important metals were Mg, Mn, and P.

**CDA.** CDA was applied to the pistachio data set using the CANDISC procedure in the SAS software package. Because the number of groups was three, the total number of possible canonical variables was two. **Figures 4 and 5** show scatter plots of the pistachio data using these two canonical variables. There was good separation of the three regions using CDA; see **Figure 4**. The three most important metals for the first canonical variable were Sr, Cu, and Na. The three most important metals for the second canonical variable were Ca, Fe, and Cu. **Figure 5** shows additional seasonal data, using the first two canonical variables. This is the best possible two-dimensional view contrasting the three locations. Note that although different seasons are color-coded for California and Iran, this view does not seek to display separation of season.

**Pattern Recognition Methods.** Pattern recognition methods refer to methods that produce classification models based on the analysis of known sample data organized into predefined groups. Samples of unknown group membership can then be input into the model and assigned a probability of belonging to one of the predefined groups. Examples of these include the methods of linear and quadratic discriminant functions, non-parametric discriminant functions, and neural networks, to name a few. The methods have been discussed in earlier publications (2, 26).

To get some sense of how well the prediction model will work on actual data, cross-validation was used. For cross-validation, the models are trained using all of the data minus one sample. This one sample is then presented to the model for classification. This process is repeated for each sample, and then, the number of correctly classified samples is reported. Cross-validation results for this data set appear in **Table 2**. Another approach, perhaps yielding a more realistic assessment, can be accomplished by separating the data into training and test sets where the test set is larger than one sample. A reasonably large, say 25%, subset of the data is randomly selected for test or validation set. A predictive model is developed using only the remaining data (called the training set). The test set is then presented to the model for classification.

First, all element data (both seasons) were used to develop a linear discriminant function using the DISCRIM procedure from the SAS statistical software package. Other related methods such

**Table 2.** Cross-Validation Results and Percentage Correctly Classified for Elemental Data Set

geographic location	linear discriminant function		
	all data cross-validation results (%)	2001 data (trained on 2001 data) test set is 2000 data (%)	all data with 25% test set randomly generated (%)
California	97.83	65.50	100
Iran	95.22	84.87	93.55
Turkey	88.0	not applicable (no 2000 data)	88.89

as quadratic discrimination functions and nonparametric discrimination functions were also tried, but the linear discriminant function worked consistently better as measured by cross-validation and training/test set strategies. Neural network software was also applied yielding results similar to those obtained using linear discriminant function analysis. Cross-validation and test/training set results (percent classified correctly) for the linear discriminant function analysis are presented in **Table 2**.

The “all data” results in **Table 2** demonstrate that a linear discriminant function model generalizes well to the “so-called” unknown (test sets) data from the made set of locations and seasons. From the all data results, we would be guardedly optimistic that a true unknown sample from one of these regions and seasons would in fact be classified correctly by such a model. The correctly classified cross-validations were >88% for all regions, while Californian pistachios were classified correctly with nearly 98% success. Errors in classification for Californian samples were most often confused with Turkish samples. The validation success was even better using a 25% test set; utilizing this approach, success rates were >89%, with Californian pistachios 100% successfully classified. The “2001 data” results provide a kind of reality check; models developed using the available 2001 data showed modest predictive ability when applied to the 2000 data. This may be due to true seasonal differences, or perhaps a broader range of geographic locations was represented in the 2001 samples as compared to the 2000 samples. There were perhaps other unknown factors (e.g., variety) as well. A continuing challenge for a study of this nature is to ensure that sample data used to develop the predictive models do in fact adequately represent both underlying variability of the entire population and seasonal variability.

Our preliminary data exploration does suggest some seasonal differences in the pistachios data; however, to further explore the consequences of this, the data were separated into training and test sets based on season. The training set consisted of all 2001 season data, and the test set consisted of all the 2000 season data. The results appear in **Table 2**. Only 65 and 85% of the Californian and Iranian samples, respectively, were correctly classified, indicating a possible seasonal effect.

In conclusion, multiple approaches provided converging lines of evidence that geolocating pistachios based on chemical profiles was possible. Multiple approaches also pointed to similar elements having strong discriminating power. Several linear discriminant function models classified the data sets with 95% or higher accuracy. Within the boundaries of this study, chemical profiling combined with pattern recognition methods to determine geographic growing origins of pistachios appears feasible. Work is in progress to further substantiate the effects of isotope ratios and seasonal and varietal influences on pistachio modeling success and generalization.

## ACKNOWLEDGMENT

We thank Dr. Bekir Erol, Harran University, and Dr. Ludwig for providing authentic samples and Robert Grove for assistance with the graphical displays.

## LITERATURE CITED

- (1) U.S. Department of Commerce Bureau of Economic Analysis, Regional Division, 1992.
- (2) Anderson, K. A.; Magnuson, B. A.; Tschirgi, M. L.; Smith, B. Determining the geographic origin of potatoes with trace metal analysis using statistical and neural network classifiers. *J. Agric. Food Chem.* **1999**, *47*, 1568–1575.
- (3) Danzer, K.; Carcia, D. D. C.; Thiel, G.; Reichembacher, M. Classification of wine samples according to origin and grape varieties on the basis of inorganic and organic trace analyses. *Am. Lab.* **1999**, *31* (20), 29–34.
- (4) Desage, M.; Guilluy, R.; Brazier, J.; Chaudron, H.; Girard, J.; Cherpin, H.; Jumeau, J. Gas chromatography with mass spectrometry or isotope-ratio mass spectrometry in studying the geographical origin of heroin. *Anal. Chem. Acta* **1991**, *247*, 249–254.
- (5) Flurur, C.; Wolnik, K. Chemical profiling of pharmaceuticals by capillary electrophoresis in the determination of drug origin. *J. Chromatogr. A* **1994**, *674*, 153–163.
- (6) Hernández, C.; Rutledge, D. Characterization of coca masses: low resolution pulse NMR study of the effect of geographical origin and roasting on fluidification. *Food Chem.* **1994**, *49*, 83–93.
- (7) Hernández, C. V.; Rutledge, D. N. Multivariate statistical analysis of gas chromatograms to differentiate cocoa masses by geographical origin and roasting conditions. *Analyst* **1994**, *119*, 1171–1176.
- (8) Cozzolineo, D.; Smyth, H. E.; Gishen, M. Feasibility study on the use of visible and near-infrared spectroscopy together with chemometrics to discriminate between commercial white wine of different varietal origins. *J. Agric. Food Chem.* **2003**, *51* (26), 7703.
- (9) Roca, M.; Grandul-Rojas, B.; Gallardo-Guerrero, L. Food, Feed Science, and Nutrition—Pigment Parameters Determining Spanish Virgin Olive Oil Authenticity. *J. American Oil Chemists' Society* **2003**, *80* (12), 1237–1241.
- (10) Tapp, H. S.; Defernes, 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* (21), 6110.
- (11) Aires-De-Sousa, J. Verifying wine origin: a neural network approach. *Am. J. Enol. Vitic.* **1996**, *47*, 410–414.
- (12) Ferland, G.; Sadowski, K. Vitamin K<sub>1</sub> (phyloquinone) content of green vegetables: effects of plant maturation and geographical growth location. *J. Agric. Food Chem.* **1992**, *40*, 1874–1877.
- (13) Hulshof, P.; Xu, C.; van de Bovenkamp, P.; Muhilal; West, C. Application of a validated method for determination of provitamin A carotenoids in Indonesian foods of different maturity and origin. *J. Agric. Food Chem.* **1997**, *45*, 1174–1179.
- (14) Parcerisa, J.; Boatella, J.; Codony, R.; Farràn, A.; Garcia, J.; López, A.; Rafecas, M.; Romero, A. Influence of variety and geographical origin on the lipid fraction of hazelnuts (*Corylus avellana* L.) from Spain: I. Fatty acid composition. *Food Chem.* **1993**, *48*, 411–414.
- (15) Parcerisa, J.; Rafecas, M.; Castellote, A.; Condony, R.; Farràn, A.; Garcia, J.; López, A.; Romero, A.; Boatella, J. Influence of variety and geographical origin on the lipid fraction of hazelnuts (*Corylus avellana* L.) from Spain: II. Triglyceride composition. *Food Chem.* **1994**, *50*, 245–249.
- (16) Parcerisa, J.; Rafecas, M.; Castellote, A.; Condony, R.; Farràn, A.; Garcia, J.; Gonzalez, C.; López, A.; Romero, A.; Boatella, J. Influence of variety and geographical origin on the lipid fraction of hazelnuts (*Corylus avellana* L.) from Spain: III. Oil stability, tocopherol content and some mineral contents (Mn, Fe, Cu). *Food Chem.* **1995**, *53*, 71–74.



- (17) Lavedrine, F.; Ravel, A.; Poupard, A.; Alary, J. Effect of geographic origin, variety and storage on tocopherol concentrations in walnuts by HPLC. *Food Chem.* **1997**, *58*, 135–140.
- (18) Armanino, C.; Fornia, M.; Castino, M.; Piracci, A.; Ubigli, M. Chemical investigation of four red wines from a single cultivar grown in the Piedmont region. *Analyst* **1990**, *115*, 907–910.
- (19) Etievant, P.; Schlich, P.; Bouvier, J. C.; Symonds, P.; Bertrand, A. Varietal and geographic classification of French red wines in terms of elements, amino acids and aromatic alcohols. *J. Sci. Food Agric.* **1988**, *45*, 25–51.
- (20) LaTorre, M. J.; García-Jares, C.; Médina, B.; Herrero, C. Pattern recognition analysis applied to classification of wines from Galicia (Northwest Spain) with certified brand of origin. *J. Agric. Food Chem.* **1994**, *42*, 1451–1455.
- (21) Vanderschee, H. A.; Bouwknegt, J. P.; Tas, A. C.; Maarse, H.; Sarneel, M. M. The authentication of sherry wines using pattern-recognition—An inter laboratory study. *Z. Lebens.-Unters. Forsch.* **1989**, *188*, 324–329.
- (22) Day, M.; Zhang, B.; Martin, G. Determination of the geographical origin of wine using joint analysis of elemental and isotopic composition. II. Differentiation of the principal production zones in France for the 1990 vintage. *J. Sci. Food Agric.* **1995**, *67*, 113–123.
- (23) Esehie, H. Distribution of chemical constituents in the plant parts of six tropical-origin forage grasses at early anthesis. *J. Sci. Food Agric.* **1992**, *58*, 435–438.
- (24) Anderson, K. A. Micro-digestion and ICP-AES analysis for the determination of macro and micro elements in plant tissues. *At. Spectrosc.* **1996**, *1/2*, 30–33.
- (25) Anderson, K. A. *Analytical Techniques for Inorganic Contaminants*; AOAC International: Gaithersburg, MD, 1999.
- (26) Anderson, K. A.; Smith, B. W. Chemical profiling to differentiate geographic growing origins of coffee. *J. Agric. Food Chem.* **2002**, *50* (7), 2068–2075.
- (27) *SAS Systems for Windows*, Release 6.11; SAS Institute, Inc.: Cary, NC, 2003.
- (28) Kabata-Pendias, A.; Pendias, H. *Trace Elements in Soils and Plants*; CRC Press: Boca Raton, FL, 1992.
- (29) Guiseppe, N.; Monetti, A.; Reniero, F. *Monitoring Authenticity and Regional Origin of Wines by Natural Stable Isotopes Ratio Analysis*; ACS Symposium Series No. 661; American Chemical Society: Washington, DC, 1997; pp 113–132.
- (30) Mansuy, L.; Philp, R. P.; Allen, J. Source identification of oil spills based on isotopic composition of individual components in weathered oil samples. *Environ. Sci. Technol.* **1997**, *31*, 3417–3425.
- (31) Whilte, J. W.; Winters, K.; Martin, P.; Rossman, A. Stable carbon isotope ratio analysis of honey: validation of internal standards procedure for worldwide application. *J. AOAC Int.* **1998**, *81* (3), 610–618.
- (32) Amundson, R.; Austin, A. T.; Schuur, E. A. G.; Yoo, K.; Matzek, V.; Kendall, C.; Uebersax, A.; Brenner, D.; Baisden, W. T. Global patterns of the isotopic composition of soil and plant nitrogen. *Global Biogeochem. Cycles* **2003**, *17* (1), 1031.
- (33) Garten, C. T., Jr. Variation in foliar 15N abundance and the availability of soil nitrogen on Walker Branch watershed. *Ecology* **1993**, *74*, 2098–2113.
- (34) Kendall, C., McDonnell, J. J., Eds. *Tracing Nitrogen Sources and Cycling in Catchments*; Elsevier Science B. V.: Amsterdam, 1988; pp 519–576.
- (35) Nikdel, S.; Nagy, S.; Attaway, J. Trace metals: Defining geographical origin and detecting adulteration of orange juice. *Sci. Toxicol.* **1988**, *596*, 81–105.
- (36) Sanz, S.; Perez, C.; Herrera, A.; Sanz, M.; Juan, T. Application of a statistical approach to the classification of honey by geographic origin. *J. Sci. Food Agric.* **1995**, *69*, 135–140.

---

Received for review July 2, 2004. Revised manuscript received October 30, 2004. Accepted November 2, 2004. We thank the California Pistachio Commission for partial funding of this research.

JF048907U