4506 J. Agric. Food Chem. 2006, 54, 4506–4516

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Stable Isotope and Trace Element Profiling Combined with Classification Models To Differentiate Geographic Growing Origin for Three Fruits: Effects of Subregion and Variety

ANGELA L. PEREZ,[†,‡] BRIAN W. SMITH,[†] AND KIM A. ANDERSON*,[†,‡]

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

Classifications of geographic growing origin of three fresh fruits combining elemental profiles with various modeling approaches were determined. Elemental analysis (Ca, Cd, Cr, Cu, Fe, K, Mg, Mn, Na, Ni, P, V, and Zn) of strawberry, blueberry, and pear samples was performed using inductively coupled plasma argon atomic emission spectrometer. Bulk stable carbon and nitrogen isotope analyses in pear were performed using mass spectrometry as an alternative fingerprinting technique. Each fruit, strawberry (*Fragaria* × *ananassa*), blueberry (*Vaccinium caesariense*/*corymbosum*), and pear (*Pyrus communis*), was analyzed from two growing regions: Oregon vs Mexico, Chile, and Argentina, respectively. Principal component analysis and canonical discriminant analysis were used for data visualization. The data were modeled using linear discriminant function, quadratic discriminant function, neural network, genetic neural network, and hierarchical tree models with successful classification ranging from 70 to 100% depending on commodity and model. Effects of Oregon subregional and variety classification were investigated with similar success rates.

KEYWORDS: Geographic authenticity; principal component analysis; canonical discriminant analysis; elemental analysis; stable isotope; strawberry (*Fragaria* × *ananassa*); blueberry (*Vaccinium caesariense*/*corymbosum*; pear (*Pyrus communis*); geographic origin; food labeling; linear discriminant function; quadratic discriminant function; hierarchal tree; neural network; genetic neural network; modeling

INTRODUCTION

Globalization has shifted the world market for fresh fruit, making availability year round commonplace. Concerns surrounding disparate agricultural practices, such as a lack of food safety standards and protection of the market share, have led commerce officials to prioritize the dissemination of methods to determine the geographic origin of commodities. For example, on May 23, 2000, the Food and Drug Administration announced that a major strawberry production company in Mexico recalled almost 13000 pounds of fresh strawberries, including the variety Fresh Delight used in this study, due to *Salmonella* contamination (*1*). Food traceability studies are important for three primary reasons: to improve supply management, to facilitate traceability for food safety and quality, and to differentiate and market foods with subtle quality attributes (*2*). Knowledge of geographic growing region is not only paramount in upholding accountability in the food production industries but is also important to consumers. In February 2001, the Consumer Right-to-Know Act (S. 280) was passed, requiring country of origin labels on perishable agricultural commodities. This act came about largely from public concern about potentially harmful substances in consumables (*3*), and polls show that a majority of consumers prefer country of origin labels (*3*, *4*). On January 27, 2004, President Bush signed Public Law 108-199, which delays until September 30, 2006, the implementation of mandatory country of origin labeling for all covered commodities except wild and farm-raised fish and shellfish (*5*).

Previous attempts have been made to elucidate the country of origin of edible commodities but, until recently, have been limited to processed foods. The geolocation of juices (*6*), drugs of abuse (*7*, *8*), cocoa (*8*, *9*), olive oil (*10*), nuts and coffee (*11*, *12*), and wine (*13*, *14*) has had moderate success. These techniques often require the use of multiple instruments, which can become laborious and increase costs. Vitamin or amino acid assays have proven successful in geolocating some commodities but are expensive due to sensitivity to degradation and, therefore, are not always conducive to broad implementation. Alternatively, we employ a chemical profiling method that is efficient, nonreagent intensive, and has reliable accuracy to differentiate the country of origin and geographic growing regions of several commodities. Although geographic origin analysis of potatoes, coffee, and pistachios has been performed successfully (*15*−*17*), geographical analysis of fresh strawberry, blueberry, and pear has not been previously investigated using this method. Finally, geographic subregional and varietal effects on geo-

* To whom correspondence should be addressed. Tel: 541-737-8501. Fax: 541-737-0497. E-mail: Kim.Anderson@orst.edu.
[†] Department of Environmental and Molecular Toxicology.
[‡] Food Safety and Environmental Stewardship Program.

10.1021/jf0600455 CCC: $33.50
Published on Web 06/07/2006

graphic classification prediction of strawberry and blueberry have not been previously published.

Multielement profiling is based on several environmental and geologic factors such as soil type, rainfall, and temperature of a growing region and provides a scientific underpinning to determine the geographic origin of a commodity. Soils contain variable concentrations of major, minor, and trace elements. The availability of these for plant uptake is dependent on the soil system. Plant elements become available for uptake in soil solution by several processes, including soil mineral weathering, decomposition of organic matter, ion exchange processes, application of soil amendments, and deposition (*18*). Plant element sequestration depends on the chemical form in soil solution. However, plants have evolved several mechanisms for adequate uptake, including soil acidification by release of hydrogen ion at the rhizosphere, anion uptake, modification of soil moisture content, organic compound exudation from roots, and root respiration (*18*).

Isotope ratios have been used as another chemical profiling method to determine geographic origin of biota (*19*) or biota-derived products (e.g., crude oil) (*20*). Recently, Kelly et al. published a review of the application of multielement and multiisotope analysis for tracing the geographical origin of food (*21*). Chemical, physical, and biological processes, such as photosynthetic fixation, can result in significant fractionation of heavy to light stable isotopes in biological matter. Plants have enzyme(s) that select against the less abundant and heavier ^{13}C isotopes relative to the ^{12}C isotopes. Other factors involved in stable carbon isotope fractionation are temperature, plant type (e.g., C3 v C4 plants) (*22*), and the environment (*23*). For example, the $^{13}C/^{12}C$ ratios in plants may differ depending on geography, latitude, location, and climate. Plants in humid environments take in more CO_2 and develop a lower ratio of ^{13}C to ^{12}C than plants in arid environments. Processes affecting nitrogen isotopic composition include N-fixation, assimilation (e.g., uptake of ammonium, nitrate, etc.), mineralization, nitrification, volatilization, sorption/desorption, and denitrification. Soil and plant $\delta^{15}N‰$ values consistently have been reported to decrease with increasing mean annual precipitation and decreasing mean annual temperature across a range of climate and ecosystem types (*24*). Globally, plant $\delta^{15}N‰$ values are more negative than soils, suggesting a systematic change in the source of plant-available N (organic/NH_4^+ vs NO_3^-) with climate (*24*). A compilation of data for nonfixing trees showed a 3−15‰ range in $\delta^{15}N‰$ values among the same species relative to small geographic areas (*25*). The large range in $\delta^{15}N‰$ reflects spatial variability in the relative amounts and bioavailability of atmospheric N vs various soil sources of N (*26*).

The hypothesis of this study is that strawberry, blueberry, and pear can be geographically classified using multielement chemical profiling and bulk stable isotope ratio techniques. The first objective of this study was to classify between two geographical growing regions Oregon vs Mexico, Chile, and Argentina for strawberry, blueberry, and pear, respectively. The second objective was to determine the classification effects of Oregon blueberry and strawberry varieties, some of which were grown on the same field only a few feet apart. The effects of subregional differences within Oregon were also evaluated.

MATERIALS AND METHODS

Reagents. Concentrated nitric acid, trace metal grade (Fisher Optima, Pittsburgh, PA); elemental stock standard solutions (Alfa Aesar Specpure, Ward Hill, MA); and 18 MΩ cm water (Barnstead, Dubuque, IA) were used. The inductively coupled plasma argon atomic emission spectrometer (ICP-AES) was used to analyze digested samples. Employed were the following parameters: model, Liberty 150 ICP-AES (Mulgrave, Victoria, Australia); V-groove nebulizer, 85 psi; Varian SPS5 autosampler system; scan integration time, 1 s (all elements); acid flexible tubing, 0.030 mm i.d. (internal diameter); replicates, three (all elements); scan window (first order), 0.120 nm; photomultiplier tube voltage, 650 V; plasma flow, 15 L/min; auxiliary flow, 1.50 L/min; sample uptake delay, 13 s; pump rate, 15 rpm; instrument stabilization delay, 13 s; and rinse time, 60 s. The wavelengths selected were as follows: Ca, 214.434; Cd, 422.673; Cr, 267.716; Cu, 324.754; Fe, 259.94; K, 285.213; Mg, 257.61; Mn, 231.604; Na, 213.618; Ni, 769.896; P, 589; V, 294.402; and Zn, 213.856.

Bulk Stable Isotope Analysis. Nitrogen ($\delta^{15}N‰$) and carbon ($\delta^{13}C‰$) bulk stable isotopes and bulk C/N ratios were measured and calculated on a stable isotope mass spectrometer (MS) (Finnigan MAT-251, ThermoFinnigan, Waltham, MA). Isotopic data used the standard isotopic δ notation (δ), in per mil (‰) relative to the Pee Dee Belemnite (PDB) scale for carbon isotopes and relative to air (^{15}N) for nitrogen. By convention, the following equation for δ was used for carbon (and an analogous equation for nitrogen):

$$(\delta)\ ^{13}C\ ‰ = \{[(^{13}C/^{12}C_{sample}) - (^{13}C/^{12}C_{std})]/(^{13}C/^{12}C_{std})\} \times 1000$$

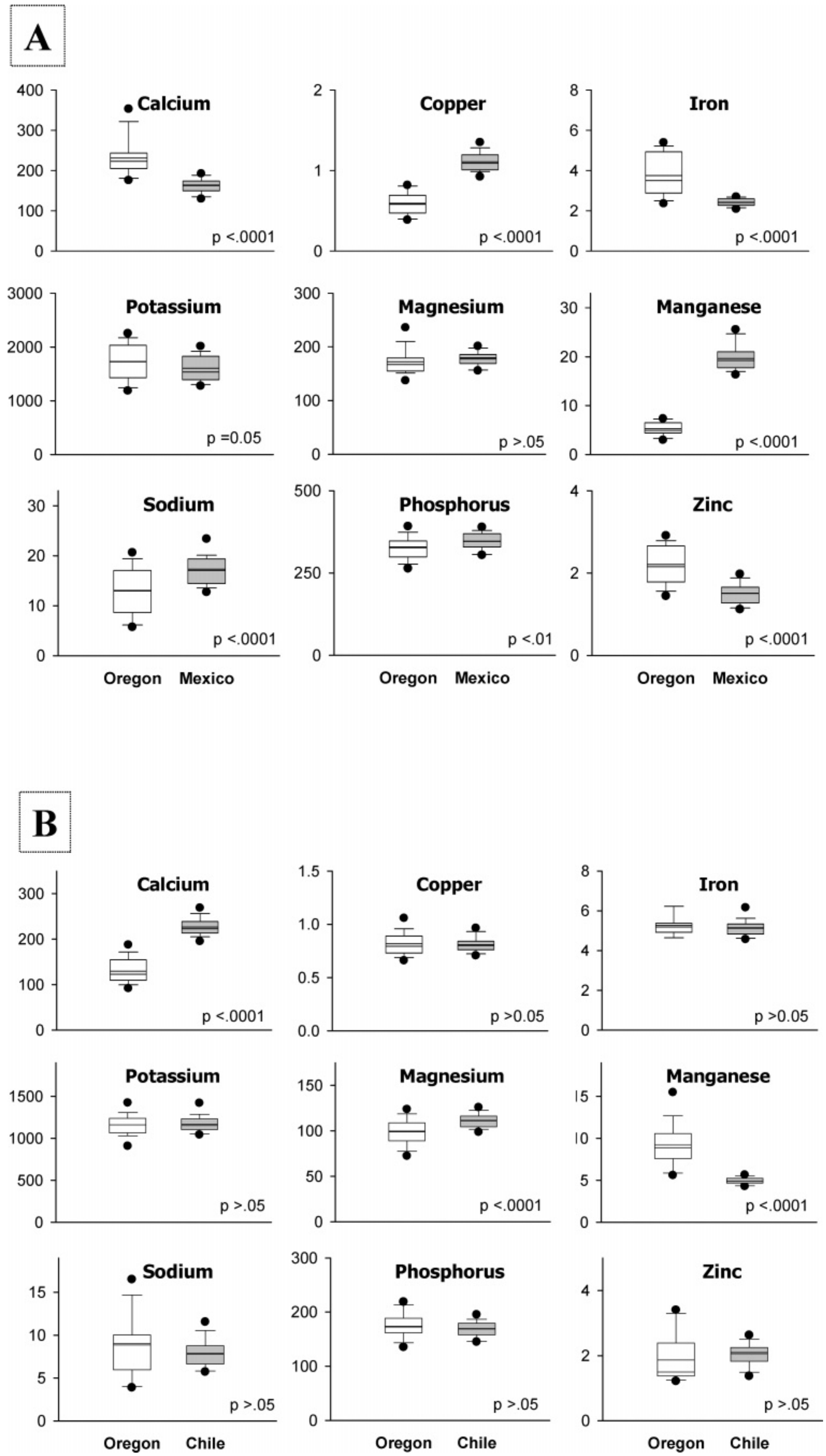
The enrichment of heavy isotopes relative to the standard gives positive values and enrichment of light isotopes relative to the standard gave negative values. Calibration to PDB was done through the NBS-19 and NBS-20 standards of the National Institute of Standards and Technology (NIST, Gaithersburg, MD).

Oregon Field Sampling. Oregon samples were collected in summer 2002 from field locations spanning the state (∼350 miles in length), including Hood River, Portland, Salem, Brownsville, Corbett, Corvallis, and Central Point, depending on commodity. At each Oregon farm, approximately 8 L of blueberry (*Vaccinium caesariense*/*corymbosum*), 8 L of strawberry (*Fragaria* × *ananassa*), and >12 pears (*Pyrus communis*) were collected by hand and labeled according to farm location (subregion) and variety. All Oregon samples were hand-picked at each individual field location, except for one pear collection site. Pears collected from the site labeled Portland were purchased at a local organic food market where they were labeled as having been grown in the Portland area. Individual field replicates were analyzed separately and represent randomized field collection (i.e., picked from multiple blueberry bushes, strawberry rows, or pear trees). Only the most common varieties in the fresh market, both nationally and internationally, were analyzed. The international samples were collected from Oregon grocery stores that offered produce labels indicating geographic origin. We intentionally collected fresh market samples when they would be out of "season" for Oregon and therefore more likely from South America/Mexico. On the basis of the differences in availability of these fruits, we made the assumption that these internationally labeled samples were authentic. No international sublocations were specified.

Sample Preparation and Analysis. All samples were rinsed under a stream of tap water, followed by a 3-fold rinse with 18 MΩ cm water, and blotted dry with paper towels. Each sample was homogenized using a Robot Coupe industrial BLIXER RS1 BX6 (Ridgeland, MS) and liquid nitrogen, until the homogenate resembled a fine powder. All samples were stored in individually HNO_3-cleaned glass jars at −20 °C until further analysis. Samples were processed according to a method previously described (*15*, *17*). Analysis of total elements within the digestate was performed using an ICP-AES. This ICP-AES multielement method required little sample (1 g), and low solvent use, resulting in decreased reagent cost, less generated waste, decreased disposal cost, and fewer hazards to the analyst.

Isotope Analysis. Pear samples were analyzed as the whole pear from freeze fracture homogenization. Homogenates were freeze-dried. Samples were loaded in capsules for MS analysis. The chemical analytical technique was well-suited to analysis of modest-to-small samples; a minimum of 2.0 ± 0.5 mg was used.

Quality Control and Statistical Analysis. Certified reference materials (CRMs) were included in each multielement analytical batch: NIST 1515 apple leaf, NIST 1573a tomato leaf (NIST). CRMs,


A
Calcium
p <.0001
Copper
p <.0001
Iron
p <.0001
Potassium
p =0.05
Magnesium
p >.05
Manganese
p <.0001
Sodium
p <.0001
Phosphorus
p <.01
Zinc
p <.0001
Oregon Mexico
B
Calcium
p <.0001
Copper
p >0.05
Iron
p >0.05
Potassium
p >.05
Magnesium
p <.0001
Manganese
p <.0001
Sodium
p >.05
Phosphorus
p >.05
Zinc
p >.05
Oregon Chile

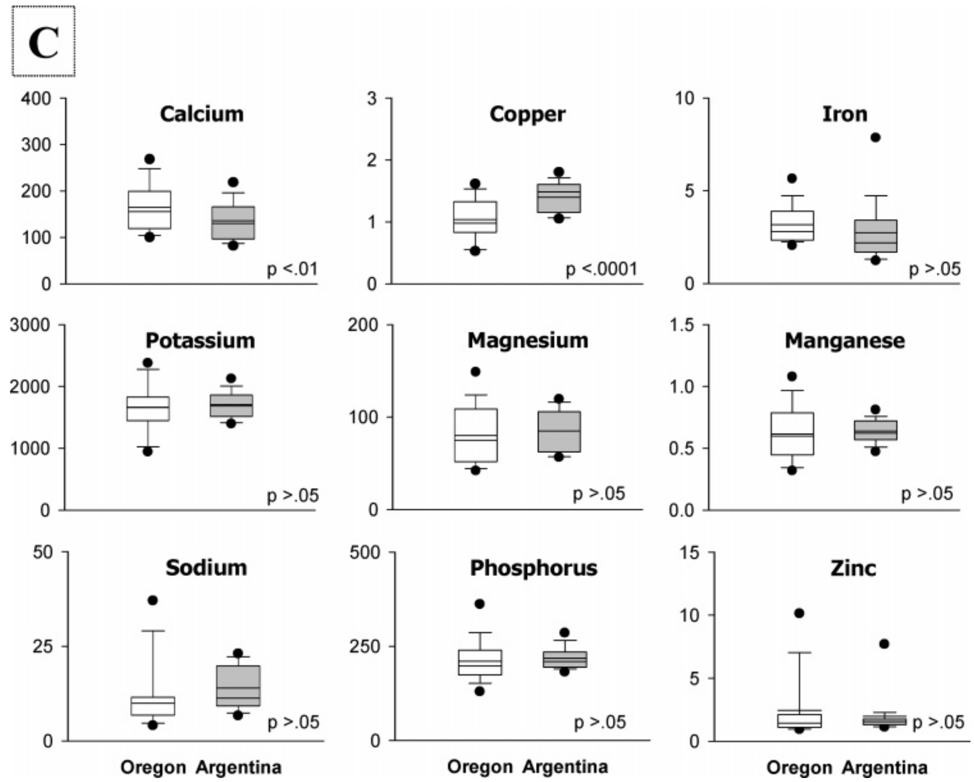


Figure 1. Element concentrations (mg/kg) of Oregon and Mexican strawberries (**A**), Oregon and Chilean blueberries (**B**), and Oregon and Argentine pears (**C**) (**A**: Oregon, $n = 40$; Mexico, $n = 42$; Iron Oregon, $n = 20$; and Iron Mexico, $n = 21$. **B**: Oregon, $n = 32$; Chile, $n = 37$; Iron Oregon, $n = 16$; and Iron Chile, $n = 37$. **C**: Oregon, $n = 40$; Argentina, $n = 40$). Oregon commodities are indicated by the white boxes, and international commodities are indicated by the gray boxes. Significant separation was determined using a two sample *t*-test. The boundary of the box indicates the 25th and 75th (top and bottom) percentiles. The lines within the box mark the mean and the median. The whiskers above and below the box indicate the 90th and 10th percentiles. The 5th and 95th percentiles are displayed with the star symbol.

check standards, and blanks accounted for at least 25% of each analytical batch. A minimum of three standards were used per calibration curve with R^2 values > 0.99. Detection limits were calculated as three standard deviations based on seven blanks. Average recoveries for each element were as follows: Ca, 108%; Cu, 120%; Fe, 98%; K, 96%; Mg, 125%; Mn, 106%; Na, 99%; P, 120%; and Zn, 116%. Check standard recoveries averaged 101%.

Each isotope sample was analyzed in triplicate. NIST 8542 sucrose ANU-sucrose and NIST 8548 IAEA-N2-ammonium sulfate samples were analyzed with each batch. External precision estimates of δ^{15}N‰ and δ^{13}C‰, based on replicate analysis of acetanilide and oxalic acid standards, were ±0.12 and ±0.11‰, respectively.

Several statistical analysis methods were applied to the data. Multiple comparisons analysis of variance (ANOVA) were used in subregional and variety analysis by Sigma Stat for Windows, Version 2.0 (Systat, Point Richmond, CA). Graphical presentations and *t*-tests comparing geographical location used SigmaPlot 2003 for Windows, Version 8.0 (Systat). Significance was determined using a two-sampled *t*-test in Sigma Plot. Canonical discriminant analysis (CDA), linear discriminant function, and quadratic discriminant function analyses were applied utilizing SAS version 9.0 (SAS Institute Inc., Cary, NC), and neural network and genetic neural network analysis were applied using NeuroShell Classifer (Ward Systems Group, Inc, V2.2, Frederick, MD). Hierarchal tree model and principal component analyses (PCAs) were applied using S-Plus (Lucent Technologies, Inc.). The modeling approach was previously described (*16*). The tree models were based on a minimum deviation = 0.01. The models were tested with up to three different approaches, resubstitution, cross-validation (i.e., leave-out-one), and test set, previously described in ref *15*. From each geographic group, five samples were randomly selected (from 40) to form a test set of 10 samples (two geographic regions) and removed from the training set. The remaining samples (nominally 35 from each group, a total of 70) were used as the training set for the classification models. Once trained, each model was then used to classify the 10 "unknown" samples in the test set. Variety testing was preformed only when we had two varieties at the same site so as not to have the confounding variation created by different geographic sites. Training and test sets were created that were variety specific; test sets were typically $n = 8$. For example, a training set was created without a specific variety and the test set contained only the variety withheld from the training set. Classification performances for each classification model for each commodity are discussed below.

RESULTS AND DISCUSSION

Regional Element Profiling. Nine elements were consistently above detection limits: calcium (Ca), copper (Cu), iron (Fe), magnesium (Mg), manganese (Mn), potassium (K), sodium (Na), phosphorus (P), and zinc (Zn). Cadmium (Cd), chromium (Cr), vanadium (V), and nickel (Ni) were often near or below detection limits. Box plots, **Figure 1**, are shown for each fruit; the boundary of the box closest to zero indicates the 25th percentile, the solid lines within the box mark the mean and median, and the boundary of the box farthest from zero indicates the 75th percentile. Whiskers above and below the box indicate the 90th and 10th percentiles, while symbols represent the 5th and 95th percentiles. Significant differences are defined at the 95% confidence level. Simple elemental distribution plots show clustering by geographic origin for Oregon and Mexican strawberries and Oregon and Chilean blueberries but not for Oregon and Argentine pears (**Figure 2**).

The data were further analyzed to explore the feasibility of classifying fruit samples according to geographic origin. Initially, this is investigated through statistical visualization methods. PCA measures variation in the elemental concentrations in the samples but does not take into account group (geographic origin)

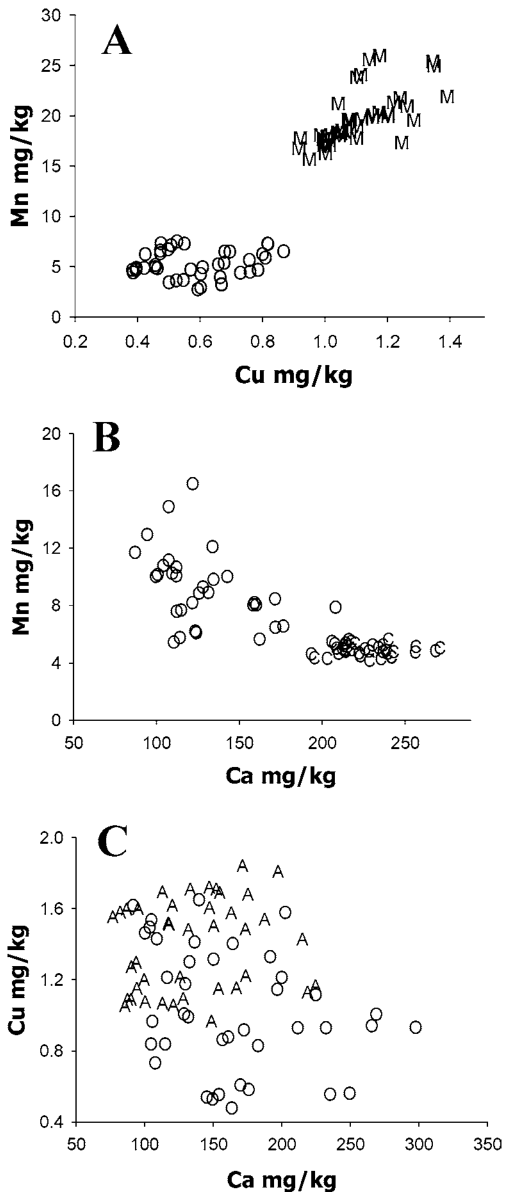


Figure 2. Concentrations of copper and manganese in Oregon and Mexican strawberries (mg/kg) (**A**); concentrations of calcium and manganese in Oregon and Chilean blueberries (mg/kg) (**B**); and concentrations of calcium and copper in Oregon and Argentina pears (mg/kg) (**C**) are shown.

membership; however, it is sometimes the case that a large percentage of the total variation can be explained by the first few principle components. This effectively reduces the number of variables needed to describe variation among samples. PCAs of geographic origin group memberships are well-manifested for strawberry and blueberry but not for pear (**Figure 3**). Because PCA does not take into account group membership, to get the best possible view of group cluster, we used CDA. Strawberry and blueberry separate well, although, pears overall are still poorly separated by geographic origin group (**Figure 4**). Modeling the data further explored the feasibility of classifying fruit samples according to geographic origin; linear discriminate function, quadratic discriminant function, neural network, genetic neural network, and hierarchal tree modeling methods were employed and are discussed below for each commodity.

Regional Strawberry Analysis (Oregon vs Mexico). The general element concentration variability in Oregon and Mexican strawberries is shown in **Figure 1A**. Strawberry concentrations of Ca, Cu, Fe, Mn, Na, and Zn showed significant separation

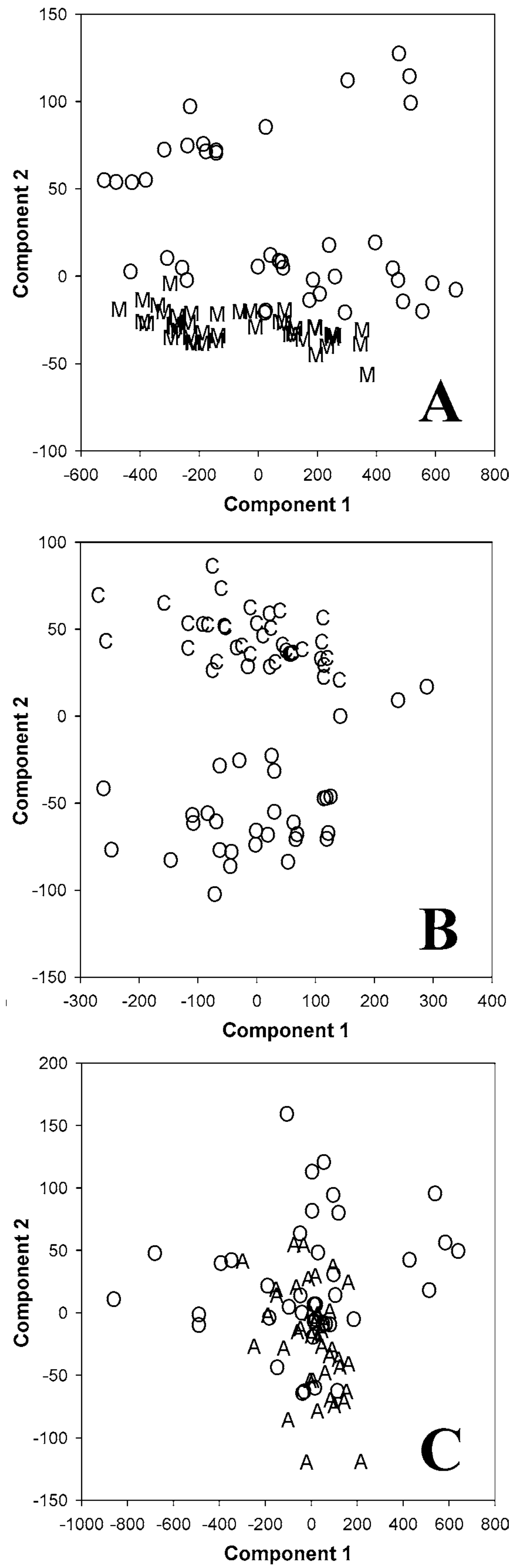


Figure 3. Principal component 1 vs principal component 2 for the chemical profile of elements in Oregon and Mexican strawberries ($n = 80$) (**A**), Oregon and Chilean blueberries ($n = 68$) (**B**), and Oregon and Argentine pears ($n = 80$) (**C**).

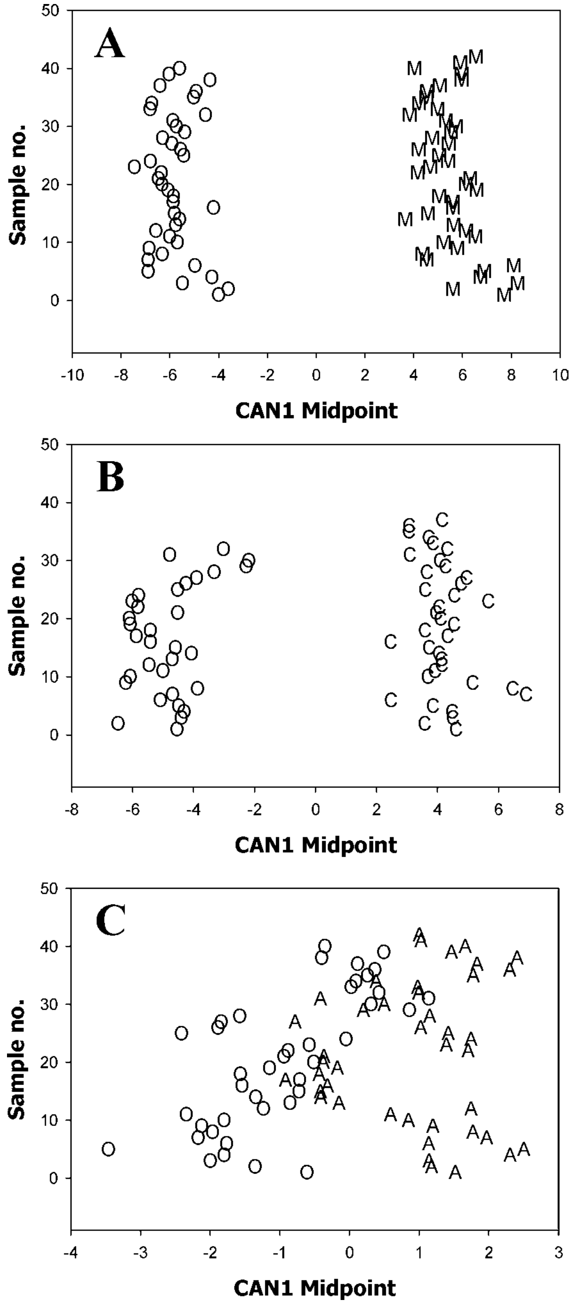


Figure 4. CDA frequency chart using the canonical variable. All 10 available dimensions are utilized in this simplified visual representation of the separation between Oregon and Mexican strawberries ($n = 80$) (**A**), Oregon and Chilean blueberries ($n = 68$) (**B**), and Oregon and Argentine pears ($n = 80$) (**C**).

($p < 0.0001$), as did P and K concentrations ($p < 0.01$ and 0.05, respectively, 78 d.f.). No significant difference for Mg concentration was observed between Oregon and Mexican strawberries. Combinations of Ca, Mn, K, Cu, Fe, or Zn could be used to visually depict geographic origin group clustering; for example, see **Figure 2A**. On average, Mexican strawberries contained 380% the concentration of Mn and 190% the Cu concentration of Oregon strawberries, while Oregon strawberries contained more Ca (29%), Fe (35%), and Zn (32%) than Mexican strawberries. Potential sources of increased Mn and Cu in Mexican strawberry might include atmospheric deposition, irrigation water, fertilizers, and soil amendments, such as biosolids (*27*). PCA generates principle components that are linear combinations of the original variables. The first principle component describes the maximum possible variation that can be projected onto one dimension. PCA on strawberry showed that the first three components accounted for 99.7% of the total variability (95, 98, and 99.7%, respectively). PCA and CDA showed strong visual clustering with Mexican and Oregon strawberries (**Figures 3A** and **4A**).

Strawberry Modeling. The results of linear discriminate function, quadratic discriminant function, neural network, genetic neural network, and hierarchal tree modeling methods are shown in **Table 1**. Multiple approaches to evaluate each model included resubstitution, cross-validation, and test sets. The linear discriminant function, the quadratic discriminant function, the neural network, and the genetic neural network models all had a 100% success classification rate for strawberries. The trace elements P, Cu, Zn, and Mg (**Figure 5A**) were found to have the most relative importance to the genetic neural network model. The hierarchal tree model also had 100% success rates. Cu concentrations less than 0.87 mg/kg were classified as Mexican (two terminal nodes) (**Figure 6**).

Regional Blueberry Analysis (Oregon vs Chile). Figure 1B depicts variable element concentrations in blueberry samples from Oregon and Chile. Ca, Mg, and Mn were strongly separated ($p \ll 0.0001$, 66 d.f.) using a two-sample *t*-test, while Cu, Fe, P, K, Na, and Zn concentrations showed no significant differences between regions ($p > 0.05$). Combinations of Ca, Mg, K, Cu, Na, Fe, or P could be used to visually depict geographic origin group clustering; for example, see **Figure 2B**. In general, Chilean blueberry had 50% the concentration of Mn and 180% the Ca concentration of Oregon blueberries. One U.S. Department of Agriculture blueberry collection site in Corvallis, Oregon, was excluded ($n = 8$) due to the historical land use at the agricultural experiment station and because no retail commodities are grown for human consumption at this site. Interestingly, blueberries from this experimental site had elevated levels of Cu and Mn (152 and 678%, respectively) relative to the average concentrations at the remaining Oregon sites. Although high bush blueberries are not fertilizer intensive, they grow readily in acidic, moist soils. This optimum growing condition renders them susceptible to increased nonnutritive metal uptake. It is suggested that blueberries could be used as a bioindicator species to assess metal contamination in soils (*28*, *29*). The U.S. Department of Agriculture is also exploring the use of coal ash and biosolid compost as a soil amendment in large scale blueberry production operations globally (*30*). High metal concentrations in blueberry marked this Corvallis site as statistically independent from all others from typical agronomical practices. These data points were removed from further statistical analysis.

Using PCA, 99.9% of the total variability could be explained by the first three principal components (75, 96, and 99.9% respectively). Strong visual regional clustering was observed for Chilean and Oregon blueberries using PCA (**Figure 3B**) and CDA (**Figure 4B**).

Blueberry Modeling. The results of linear discriminant function, quadratic discriminant function, neural network, genetic neural network, and hierarchal tree modeling methods are shown in **Table 1**. The linear discriminant function, the quadratic discriminant function, the neural network, and the genetic neural network models all had a 100% success classification rate for blueberry. The trace elements Cu, Mg, and Zn, **Figure 5B**, were found to have the most relative importance to the genetic neural network model. The hierarchal tree model had 100% success rates. Mn concentrations <6.65 mg/kg were classified as Chilean (two terminal nodes) (**Figure 6**).

Table 1. Sample Number, Linear Discriminant Function, Quadratic Discriminant Function, Neural Network, Genetic Neural Network, and Hierarchal Tree Model Classification Performance Analysis Results for Regional Geographical Origin Prediction of Blueberry, Strawberry, and Pear Samples Based on Total Recoverable Element Concentration Profiling

		%											
		linear discriminant function			quadratic discriminant function			neural network		genetic neural network		hierarchal tree	
fruit (all varieties)	region	resubstitution	cross-validation	test set ($n = 10$)	resubstitution	cross-validation	test set ($n = 10$)	resubstitution	test set ($n = 10$)	resubstitution	test set ($n = 10$)	resubstitution	test set ($n = 10$)
strawberry	Mexico ($n = 40$)	100	100	100	100	100	100	100	100	100	100	100	100
	OR ($n = 40$)	100	100	100	100	100	100	100	100	100	100	100	100
blueberry	Chile ($n = 36$)	100	100	100	100	100	100	100	100	100	100	100	100
	OR ($n = 40$)	100	100	100	100	100	100	100	100	100	100	100	100
pear	Argentina ($n = 40$)	74	75	60	100	100	100	100	100	100	100	93	90
	OR ($n = 40$)	75	70	80	88	85	100	94	80	100	100	93	90

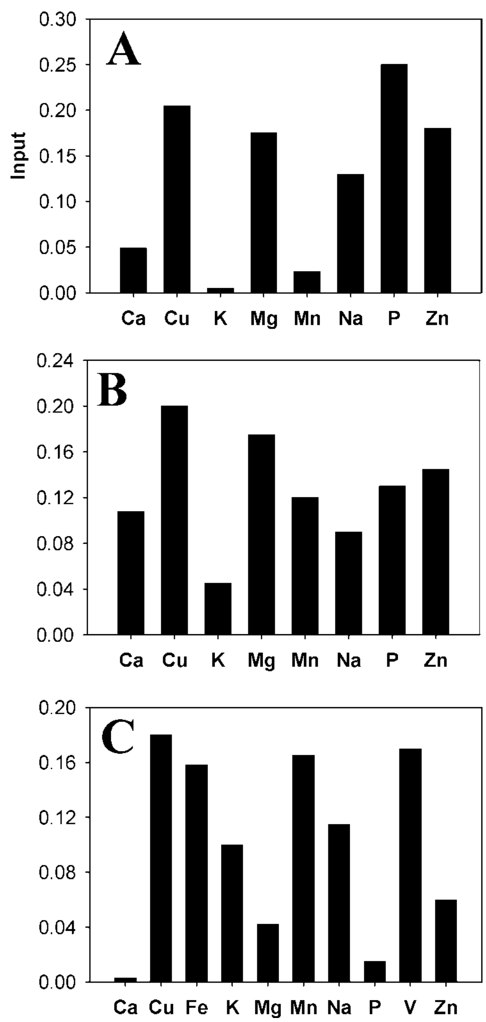


Figure 5. Genetic neural network model, relative importance of inputs, used to classify Oregon and Mexican strawberries (**A**), Oregon and Chilean blueberries (**B**), and Oregon and Argentine pears (**C**).

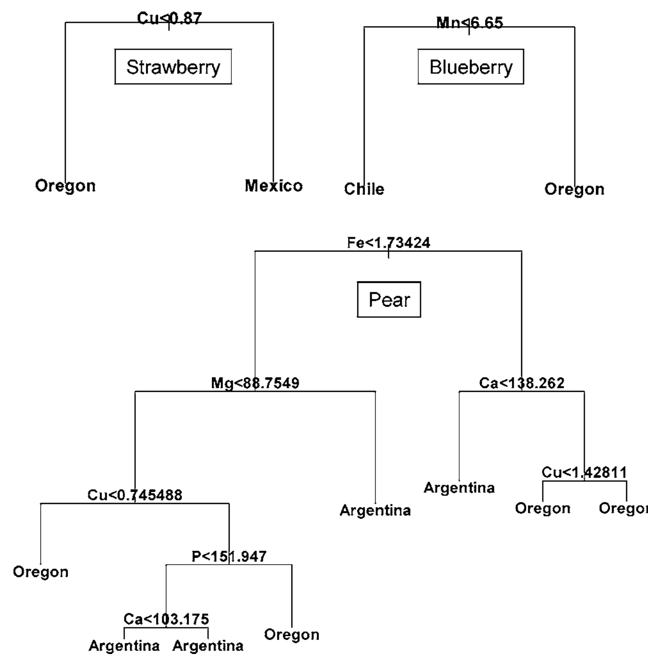


Figure 6. Hierarchal tree models for classification of Oregon and Mexican strawberries, Oregon and Chilean blueberries, and Oregon and Argentine pears are shown. A tree-based model results in a simplified hierarchical tree of decision rules useful for geographic origin classification. Resubstitution rules result in 100, 100, and 93% correct classification rates, respectively, for the data sets.

Regional Pear Analysis (Oregon vs Argentina). Element concentration variation in Oregon and Argentine pears can be observed in **Figure 1C**. Two-sample t-tests suggest that Cu concentration showed significant separation ($p < 0.0001$, 78 d.f.) as did Ca ($p < 0.01$), while all other element concentrations were not significantly different between Oregon and Argentine pears ($p > 0.05$). Combinations of trace elements could not be found that provided good visually depicted geographic origin group clustering (**Figure 2C**). PCA results showed that the first three components explained 99.9% of the variability (93, 98, and 99.9%, respectively); however, no visual clustering of Oregon and Argentine pears was observed (**Figure 3C**). The CDA frequency chart for Oregon and Argentina pears also shows a great deal of overlap (**Figure 4C**).

Pear Modeling. The results of linear discriminate function, quadratic discriminant function, neural network, genetic neural network, and hierarchal tree modeling methods. Multiple approaches to evaluate each model included resubstitution, cross-validation, and test sets as shown in **Table 1**. Overall, the linear

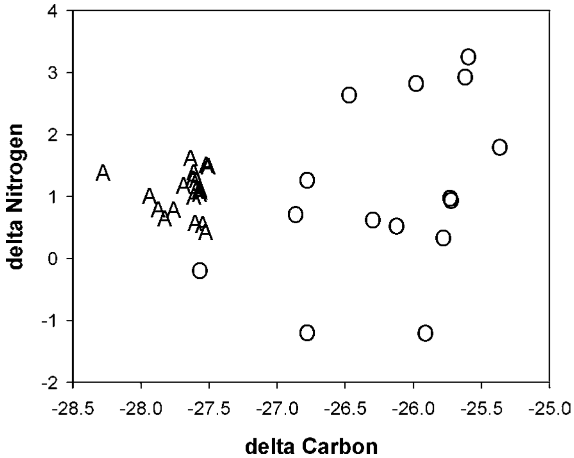


Figure 7. Argentina (A) and Oregon (O) pear isotope ratios (Oregon, $n = 16$; Argentina, $n = 20$).

discriminant function model did not perform very well on the pear data set; this modeling analysis had only a 60−80% success rate. The other modeling methods were more successful. The quadratic discriminant function had an 85−100% success rate, and the neural network had an 80−95% success rate. The best model for the pear data set was the genetic neural network models, which had a 100% success rate. Genetic algorithms seek to solve optimization problems using the methods of evolution, explicitly survival of the fittest. In a typical optimization problem, there are a number of variables that control the process and a formula or algorithm that combines the variables to fully model the process. The problem is then to find the values of the variables that optimize the model in some way. Other traditional methods tend to break down when the problem is not so "well-behaved", but genetic algorithms are designed to perform on data that is not so "well-behaved", which may account for its success with pear. The microtrace elements Cu, Mn, and V, **Figure 5C**, were found to have the most relative importance to the genetic neural network model. Copper, manganese, and vanadium are trace elements that are unlikely to be directly amended into soil; therefore, although the pears were difficult to model, this model may be especially robust to environmental/agronomic changes. Additional research is warranted to test this hypothesis. The hierarchal tree model used for regional classification prediction of Oregon and Argentine pear is shown in **Figure 6** and is significantly more complex than for the other fruits tested**.** The tree model requires eight terminal nodes to meet the classification criteria and then has a classification success rate of 93%.

Bulk stable isotope ratios in pear samples were used to address the lack of initial modeling success between Oregon and Argentine samples. We have used bulk stable isotope ratios previously to successfully investigate geographic origin (*16*, *31*). Bulk stable isotope ratios, $\delta^{13}C/\delta^{15}N$, depict visual separation between Oregon and Argentine pears, as shown in **Figure 7**. Oregon pear had significantly less enrichment of lighter ^{12}C than Argentine pear ($p < 0.0001$, 42 d.f.). No significant differences in $\delta^{15}N$ were observed between Oregon and Argentine pears ($p > 0.05$). The addition of the bulk stable isotope ratio data to the models would most likely increase the modeling success rate for pear.

Variety and Subregional Analysis. One caveat of using profiles of elemental concentrations based on country-to-country data is the possibility of misclassification due to varietal effects. It is difficult to get good variety data, because typically each variety is grown in a different location, so there are inherent geographic differences leading to large confounding variables. In this study, we were able to collect two varieties of strawberry and two varieties of blueberry from adjacent plants (same soil, same environment, and same agronomy practices), therefore providing an excellent opportunity to evaluate the variety effect without many of the typical confounding variables. Effects of variety on geographic origin analysis of strawberry and blueberry have not been previously published. **Figure 8** shows some of the element variety and subregional differences, suggesting differing variety element uptake for strawberry, blueberry, and pear.

Table 2. Linear Discriminant Function, Quadratic Discriminant Function, Neural Network, Genetic Neural Network, and Hierarchal Tree Model Classification Performance Analysis Results for Variety Effects on Geographical Origin Prediction of Strawberry and Blueberry Samples Based on Total Recoverable Element Concentration Profiling

		%				
		discriminant function		network		
fruit	variety[a]	linear	quadratic	neural	genetic neural	hierarchal tree
strawberry	Hood	100	100	100	100	88
	Totem	100	100	100	100	100
blueberry	Bluecrop	100	100	100	100	100
	Jersey	100	100	100	100	63

[a] Varieties selected for test set modeling were those field sites where two varieties were collected. Field sites where only a single variety was available (Mt. Angel, Puget summer and Hood; Brownsville, Totem) were not modeled individually.

Oregon Strawberry Variety Effects. Although there are large differences in Cu and Mn concentrations between Mexican and Oregon strawberry, there are also some Cu and Mn variety differences between Oregon strawberries (**Figure 8A**) (multiple comparisons ANOVA). Significant Na concentration differences were seen between Totem and Hood cultivars from the Corvallis field location ($p < 0.01$). Fe concentrations were significantly different between Hood and Puget summer cultivars at the Mt. Angel field site ($p < 0.01$). Although there are variety differences within Oregon strawberries grown in the same field, these differences are relatively small as compared to the overall elemental profile differences with Mexican strawberries and, most importantly, within the framework of this study, do not appear to adversely affect modeling success (**Table 2**). We tested the effects of variety on all of the models. At field sites where we had two varieties, we removed one variety from the model training set. The training set then contained some strawberries from the geographic site (i.e., representing environmental conditions, soil, agronomical practices, etc.) but would not contain the second variety, in this way isolating the variety effect. The test set would then be composed of a single variety, as always, withheld from the training set. The linear discriminant function, quadratic discriminant function, neural network, and genetic neural network models all had 100% success rates (**Table 2**). The hierarchal tree model had 88−100% success rates. This suggests that within this strawberry data set that variety differences do not adversely affect geographic origin modeling using profile elemental concentrations.

Oregon Strawberry Subregional Effects. The strawberry cultivar Totem had significantly higher mean Zn concentrations at the Brownsville site as compared to the Corvallis site only 22 miles away ($p < 0.01$). Significant mean concentration differences among the Hood cultivar between Corvallis and Mt. Angel field locations were also seen for Cu, K, and Zn ($p < 0.0001$). These subregional differences are not surprising,

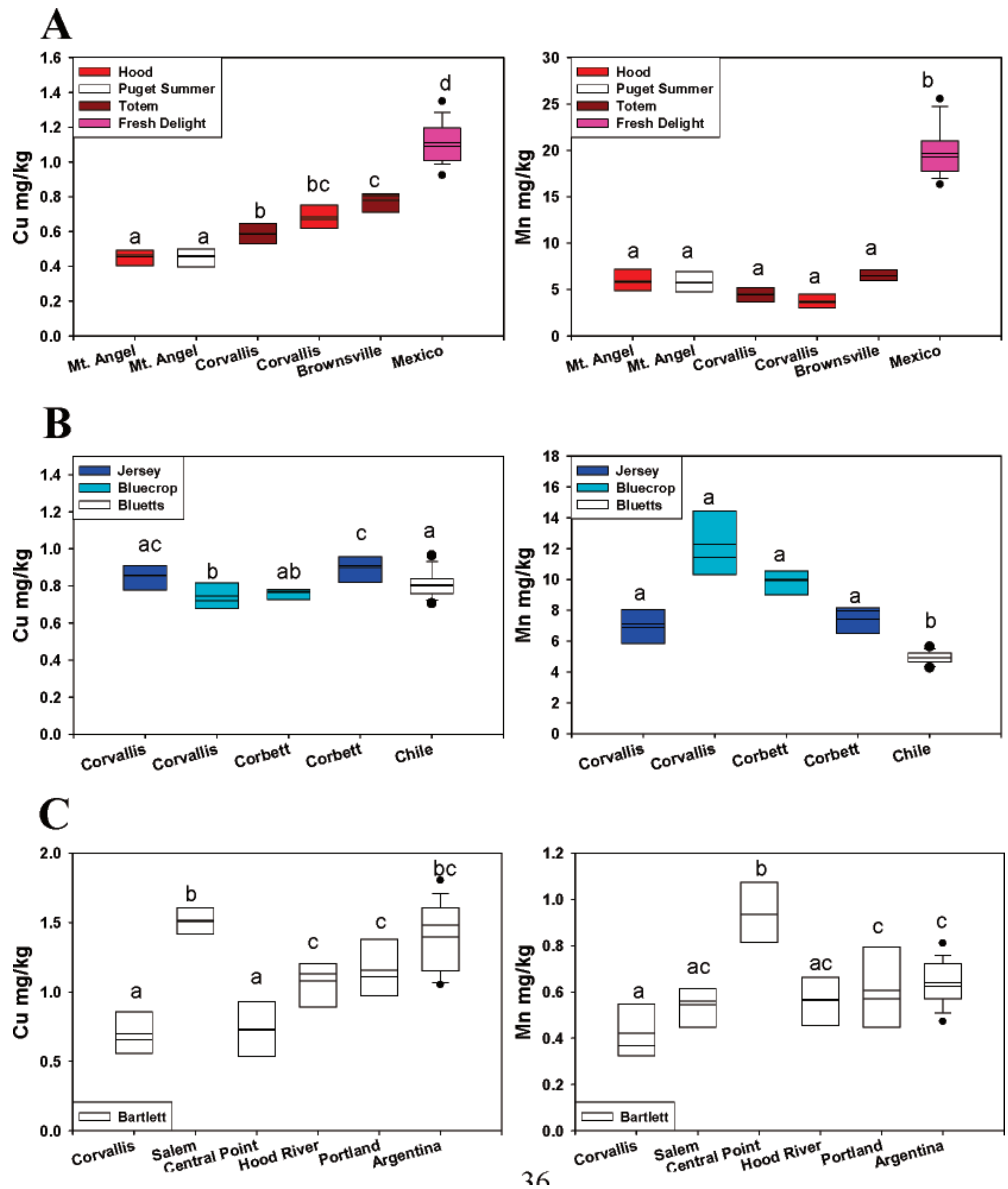


Figure 8. Strawberry (**A**), blueberry (**B**), and pear (**C**) copper and manganese concentrations vs subregion and variety are shown. Statistical differences were determined using multiple comparisons ANOVA. Letters denote statistical differences at the 95% confidence level. The boundary of the box indicates the 25th and 75th (top and bottom) percentiles. The solid lines in the box mark the median and mean. The 5th and 95th percentiles are displayed with the star symbol.

considering the diversity of Oregon soils. Coetzee et al. found that, in South Africa, the combination of elements characterizing wines from a particular region was different within each region (*32*). However, like the variety data, within this strawberry data set, subregional differences are relatively small and do not adversely affect geographic origin modeling combined with profile elemental concentrations; for example, hierarchal tree model test set success rates were >88% (Brownsville, 88%; Corvallis, 94%; and Mt. Angel, 100%).

Oregon Blueberry Variety Effects. A significant difference in element concentrations among blueberries, Jersey variety and Bluecrop cultivar, suggests that there are discernible differences between varieties/cultivars of blueberries picked from the same field location, as is the case with Cu shown in **Figure 8B**. Mean element concentrations of Cu and Zn were significantly different between Jersey and Bluecrop blueberries at the Corvallis field location ($p < 0.0001$). Jersey and Bluecrop blueberries also showed significant differences between mean Ca, Cu, and Mg picked from the Corbett field site ($p < 0.005$). Variety test sets were created as described above for strawberry. The linear discriminant function, quadratic discriminant function, neural network, and genetic neural network models all had 100% success rates (**Table 2**). The hierarchal tree model had 63–100% success rates. This suggests that within this blueberry data set that variety/cultivar differences do not adversely affect most geographic origin modeling using profile elemental concentrations. The hierarchal tree model, however, did not perform as well overall within the framework of this study, suggesting that blueberry variety/cultivar may adversely affect some models. The hierarchal tree model is a less complicated model, synthesizing the decision rules to single elements; this probably makes it inherently less robust than the discriminant analyses and neural network modeling methods, which take into account all elements. The tree model appears to be more susceptible to inadequate or changing population variation, due to the simplification nature of the model, at least within the framework of this study.

Oregon Blueberry Subregional Effects. The Bluecrop cultivar showed significant differences between the Corvallis and the Corbett field sites for mean Ca, Mg, Mn, K, and Zn concentrations ($p < 0.05$). The Jersey variety showed a significant difference between the Corvallis and the Corbett sites only for mean Ca and Mg concentrations ($p < 0.04$). Similar success rates were achieved on subregional test sets (>80%). Models could also be created with a high degree of success based on subregional geographic origins. Hierarchal tree model test set success rates were >82% (Corvallis, 82%; Corbett, 88%).

Oregon Pear Subregional Analysis. Differences in metal concentrations among Bartlett pear samples from Oregon subregions can be seen in **Figure 8C** and are small relative to strawberries and blueberries. This may be a consequence of having analyzed only one variety of pear. Another explanation is the differences in growing conditions and element assimilation in pear vs blueberry and strawberry. Tree fruits undergo a more significant element translocation distance, and reproductive sinks are directly related to the age of the tree and climate of the growing site. Despite these results, significant differences were observed between sites. The most dramatic differences were found with Cu concentrations at Salem ($p < 0.001$) and with Mn concentrations at Central Point ($p < 0.001$), with respect to the Portland site (the next site closest in concentration for both metals). Hierarchal tree model test set success rates were ≥50% (Central Point, 63%; Corvallis, 75%; Salem, 50%; Portland, 50%; and Hood River, 100%). Only one variety of pear was included in the study, so variety analysis was not performed.

Pear Bulk Stable Isotope Ratios. Because the modeling of element profiles for pear was less successful, we investigated bulk stable isotope ratios as a means to gain further discriminating chemical data. Isotopic analysis of Oregon sublocations showed significant separation among δ^{15}N ratios, ranging from -2 to $+4$ δ^{15}N. Most Oregon sublocations δ^{15}N ratios were significantly different from one another. Central Point showed strong significant differences from all other sublocations ($p < 0.01$), while Hood River was significantly different from Portland and Salem ($p < 0.05$). Portland and Salem were not statistically different from one another ($p > 0.05$). Positive δN ratios indicate a selective enrichment of heavy ^{15}N as compared to ^{14}N. Central Point Bartlett pear samples accumulated the heavier ^{15}N isotope as compared to ^{14}N followed by Salem, Portland, and Hood River, respectively. This could be, in part, due to the latitudinal differences of the field sites; δ^{15}N has been associated with latitude (*33*).

Another potential caveat of the pear data set was potentially revealed when subregional geographic origin CDA plots were generated. Oregon pears show visual clustering differences from Argentine pears, with one notable exception: the pear labeled as Portland. These were the only samples from Oregon not hand collected. One possible explanation for this overlap could be that the pears were mislabeled as being grown in Oregon. As with all authenticity studies, authenticating samples is critical to the study, as well as developing a database that contains all of the potential variations in the population to be studied.

The genetic neural network model performed the best of all modeling methods. Interestingly, some elements from the genetic neural network model were consistently found to be important to the model input, specifically Cu, Mn, Mg, and Zn. It may be possible to create further simplification of the method by analyzing and modeling only these elements and as needed adding bulk stable isotopes. Creating a fingerprint or unique chemical signature using trace element and stable isotope ratio chemical profiling may serve as a cost-effective approach toward determining the geographic growing region of a food commodity. The identification of distinct chemical−signature effects on geographic origin from sublocation and variety/cultivar of fresh fruit has not previously been described. The ease and efficiency of trace element analysis make it an optimal choice for geographic regional and subregional determination of blueberry, strawberry, and pear. Within the framework of this study, it appears that the geographic origin of strawberry, blueberry, and pear may be feasible through their chemical profile. Statistical analyses revealed groupings between the two major geographic regions for each commodity studied. The progression of this type of profiling study includes the addition of other geographic regions, seasonal variation (including agronomical changes), and additional varieties from all locations (*31*). This information may ultimately increase food safety measures and command accountability in global food production.

ACKNOWLEDGMENT

We acknowledge W. E. Hillwalker, J. Basile, L. Quarles, and B. Rugh for assistance with the analyses and the Anderson lab group for technical guidance and support. Finally, we thank three anonymous reviewers for their helpful discussion of this manuscript.

LITERATURE CITED

(1) Department of Health and Human Services. *Recalls and Field Corrections: Foods Class I*; U.S. Public Health Service, U.S. Food and Drug Administration: Rockville, MD, 2000; Recall #F-460-0.

(2) Golan, E.; Krissoff, B.; Kuchler, F. Food traceability: One ingredient in a safe and efficient food supply. *Amber Waves* **2004**, *2*, 14−21.

(3) Fresh Trends 2002. *The Packer Magazine*; Shawnee Mission: KS, 2002.

(4) Umberger, F.; Calkins; Sitz. 2003 FAMPS Conference: Emerging Roles for Food Labels: Inform, Protect, Persuade, Washington, DC.

(5) Mandatory Country of Origin Labeling of Fish and Shellfish; Interim Final Rule. 7 CFR Part 60, 2004; pp 59708−59750.

(6) Ogrinc, N. K. I. J.; Spangenberg, J. E.; Kidric, J. The application of NMR and MS methods for detection of adulteration of wine, fruit juices and olive oil. *Anal. Bioanal. Chem.* **2003**, *376*.

(7) 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.

(8) Flurur, C.; Wolnik, K. Chemical profiling of pharmaceuticals by capillary electrophoresis in the determination of drug origin. *J. Chromatogr.* **1994**, *674*, 153−163.

(9) Hernández, C.; Rutledge, D. Characterization of cocoa masses: Low resolution pulse NMR study of the effect of geographical origin and roasting on fluidification. *Food Chem.* **1994**, *49*, 83−93.

(10) Tapp, H. S.; Deferes, M.; Kemsley, E. K. FTIR spectroscopy and multivariate analysis can distinguish the geographic origin of extra virgin olive oils. *J. Agric. Food Chem.* **2003**, *51*, 6110−6115.

(11) Krivan, V.; Barth, P.; Feria Morales, A. Multielement analysis of green coffee and its possible use for the determination of origin. *Mikrochim. Acta* **1993**, *110*, 217−236.

(12) Bicchi, C. P.; Binello, A. E.; Legovich, M. M.; Pellegrino, G. M.; Alfredo, C. Vanni characterization of roasted coffee by S-HSGC and HPLC-UV and principal component analysis. *J. Agric. Food Chem.* **1993**, *41*, 2324−2328.

(13) Cozzolineo, D.; Smythe, 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*, 7703.

(14) Danzer, K.; Carcia, D. D. C.; Thiel, G.; Reicembacher, M. Classification of wine samples according to origin and grape varieties on the basis of inorganic and organic trace analyses. *Am. Lab.* **1999**, *31*, 29−34.

(15) Anderson, K. A.; Smith B. W. Chemical profiling to differentiate geographic growing origin of coffee. *J. Agric. Food Chem.* **2002**, *50*, 2068−2075.

(16) Anderson, K. A.; Smith, B. W. Use of chemical profiling to differentiate geographic growing origin of raw pistachios. *J. Agric. Food Chem.* **2005**, *53*, 410−418.

(17) Anderson, K. A.; Magnuson, B. A.; Tschirgi, M. L.; Smith, B. W. Determining the geographic origin of potatoes with trace metal analysis using statistical and neural network classifiers. *J. Agric. Food Chem.* **1999**, *47*, 1568−1575.

(18) Cresser, M.; Killham, K.; Edwards, T. *Soil Chemistry and Its Applications*; Cambridge University Press: Cambridge, 1993.

(19) Guiseppe, N.; Monetti, A.; Reniero, F. Monitoring authenticity & regional origin of wines by natural stable isotopes ratio analysis. *ACS Symp. Ser.* **1997**, *661*, 113−132.

(20) Mansuy, L.; Philp, R. P.; Allen, J. Source identification of oil spills bases on isotopic composition of individual components in weathered oil samples. *Environ. Sci. Technol.* **1997**, *31*, 3417−3425.

(21) Kelly, S.; Heaton, K.; Hoogewerff, J. Tracing the geographical origin of food: The application of multi-element and multi-isotope analysis. *Food Sci. Technol.* **2005**, *16*, 555−567.

(22) Whilte, J. W.; Winters, K.; Martin, P.; Rossman, A. Stable carbon isotope ratio analysis of honey: Calidation of internal standards procedure for worldwide application. *J. AOAC Int.* **1998**, *81*, 610−618.

(23) Guy, R. D.; Holowachuk, D. L. Population differences in stable carbon isotope ratio of Pinus contorta Dougl.ex lour.: Relationship to environment, climate of origin, and growth potential. *Can. J. Bot.* **2001**, *79*, 274−283.

(24) 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*, 1031.

(25) Garten, C. T., Jr. Variation in foliar 15N abundance and the availability of soil nitrogen on Walker Branch watershed. *Ecology* **1993**, *74*, 2098−2113.

(26) Kendall, C. Isotope tracers in catchment hydrology. In *Tracing Nitrogen Sources and Cycling in Catchments*; Kendall, C. M. J. J., Eds.; Elsevier Science B. V.: Amsterdam, 1998; pp 519−576.

(27) U.S. Environmental Protection Agency. *Biosolids Technology Fact Sheet: Land Application of Biosolids*; Office of Water, U.S. EPA: Washington, DC, 2000; EPA 832-F-00-064.

(28) Buzman, B.; Lorek, E. Bioindicatory value of *V. angustifolium*-spp. *Pr. Nauk. Uniw. Slask. Katowicach* **1982**, *505*, 242−254.

(29) Sheppard, S. C. A field and literature survey, with interpretations, of elemental concentrations in blueberry (*V. angustifolium*). *Can. J. Bot.* **1991**, *69*, 63−77.

(30) Black, B. L.; Zimmerman, R. H. Mixtures of coal ash and compost as substrates for highbush blueberry. *J. Am. Soc. Hortic. Sci.* **2002**, *127*, 869−877.

(31) Anderson, K. A.; Smith, B. W. Effect of season and variety on the differentiation of geographic growing origin of pistachios by stable isotope profiling. *J. Agric. Food Chem.* **2006**, *54*, 1747−1752.

(32) Coetzee, P. P.; Steffens, F. E.; Eiselen, R. J.; Augustyn, O. P.; Balcaen, L.; Vanhaecke, F. Multi-element analysis of South African wines by ICP-MS and their classification according to geographical origin. *J. Agric. Food Chem.* **2005**, *53*, 5060−5066.

(33) Flanagan, L. B.; Ehleringer, J. R.; Pataki, D. E. *Stable Isotopes and Biosphere-Atmosphere Interactions: Processes and Biological Controls*; Elsevier Academic: San Diego, 2005.

Received for review January 5, 2006. Revised manuscript received May 4, 2006. Accepted May 8, 2006. We thank the Oregon Agriculture Research Foundation for partial funding of this research.

JF0600455