Available Online at ESci Journals
Journal of Food Chemistry and Nutrition
ISSN: 2307-4124 (Online), 2308-7943 (Print)
http://www.escijournals.net/JFCN

DETERMINING THE MOST EFFECTIVE COMBINATION OF CHEMICAL PARAMETERS FOR DIFFERENTIATING THE GEOGRAPHIC ORIGIN OF FOOD PRODUCTS: AN EXAMPLE USING COFFEE BEANS

[ab]Rebecca J. McLeod, [a]Mikaela Garland, [a]Robert V. Hale, [c]Shawn Steiman, [ad]Russell D. Frew

[a] *Department of Chemistry, University of Otago, PO Box 56, Dunedin, New Zealand.*
[b] *Oritain Global Ltd., Invermay Agricultural Centre, Private Bag 50034, Mosgiel, Dunedin, New Zealand.*
[c] *Coffea Consulting, 546A 12th Avenue, Honolulu, Hawaii, 96816, USA.*
[d] *FAO/IAEA Division of Nuclear Techniques in Food and Agriculture, Vienna, Austria.*

ABSTRACT

Numerous chemical measures have been explored as ways to discriminate the geographic origin of food products. The "best" set of measures to determine the provenance of a food product may vary depending on the spatial scale in question. Canonical analysis of principle coordinates was used to determine which chemical measure(s) provided the best ability to determine the provenance of green coffee beans sourced from nine international growing regions. Models were constructed at two spatial scales. The chemical analyses used were the stable isotope ratios of carbon ($\delta^{13}C$), nitrogen ($\delta^{15}N$) and hydrogen ($\delta^{2}H$), concentrations of fatty acids, and concentrations of major, minor and trace elements. Variations in elemental concentrations provided the best predictor of whether a sample was from Kona, Hawaii or not (classification success rate 100%). Variability in elemental concentrations was also the single best discriminator of growing region of origin; however, the highest classification success, 86%, was achieved by elemental concentration data combined with $\delta^{13}C$ and $\delta^{15}N$ or $\delta^{2}H$. The statistical framework allows comparisons at multiple spatial scales to assist with decisions regarding which chemical analyses may be appropriate for the development of proof of origin methods for specific food and beverage products in the future.

Keywords: provenance, geographic origin, canonical analysis of principal coordinates, multivariate statistics, coffee beans, *Coffea*, elemental analysis, stable isotopes, fatty acids, food.

INTRODUCTION

Recently, there has been an increase in the number of studies using chemical approaches to determine the geographic origin and authenticity of a range of food products (Carcea *et al*., 2009). An impetus for these studies has increased public awareness and concern about the prevalence of food fraud in the international marketplace (Carcea *et al*., 2009; Kelly *et al*., 2005). Research to date covers a wide range of commodity and specialty products including, but not limited to, wine (Baxter *et al*., 1997; Martin *et al*., 2012), meat (Camin *et al*., 2007), honey (Kropf *et al*., 2010; Schellenberg *et al*., 2010), and olive oil (Watling *et al*., 2010). Studies span a range of spatial scales (e.g. comparisons among sites within a country (Baroni *et al*., 2011; Bertrand *et al*., 2008) versus comparisons among countries (Camin *et al*., 2007; Maggi *et al*., 2011)), and the analytical approaches used for site differentiation include quantification of elemental concentrations (Heaton *et al*., 2008; Kelly *et al*., 2005; Watling *et al*., 2010), ratios of heavy and light isotopes (Kelly *et al*., 2005; Ogrinc *et al*., 2009; Rummel *et al*., 2010; Watling *et al*., 2010), concentrations of fatty acids (Bertrand *et al*., 2008; Martín *et al*., 2001; Villarreal *et al*., 2009), and quantification of rare earth elements (Joebstl *et al*., 2010), with varying degrees of success. The commercial application of this body of research is currently challenged by the wide range of analytical approaches tested to date (Ye, 2012), which may well leave the reader questioning which technique or suite of techniques is best suited to their particular food product

* Corresponding Author:
Email ID: rebecca.mcleod@otago.ac.nz

and issue. In looking for a commercial solution to determining the provenance of food, the chemical approach used will depend upon a number of factors including the food type, the analytical methods available (and their cost) and the particular origin question to be answered. One producer may want to distinguish their product from others produced in the same country, whilst a consumer may want to confirm the continent of origin. Another consideration is what level of confidence is required in the assignment of origin. In the current study multiple chemical approaches are compared in their ability to differentiate the origin of samples of green (unroasted) coffee beans. Canonical analysis of principle coordinates (CAP; Primer+PERMANOVA, U.K.), a multivariate statistical approach, is used to determine the best measure or combination of measures for discriminating sample origin. This approach is proposed as a quantitative means for determining the best-suited analytical approach for discriminating origin, and it also allows origin discrimination ability to be compared across a range of spatial scales. The approach is important as methods of food origin differentiation begin to be used in a commercial capacity.

International trade in coffee exceeds 8.2 million tonnes annually and in 2008 involved 77 coffee-producing nations, exporting to 167 countries (Food and Agriculture Organization of the United Nations, http://www.faostat.fao.org/site/609). Most coffees are priced according to the New York Board of Trade "C" price, however, specialty coffees can command higher prices due to their quality, reputation, and/or rarity. One such example is the coffee produced in the Kona region of Hawaii, USA, which is currently priced some 5 times greater than the New York "C" price (Gary Strawn, pers comm.). The Kona region has enjoyed relatively high worldwide prices for several decades, which encouraged attempts to counterfeit the product (Schoenholt, 2001). For several years in the mid-1990's, a merchant purchased cheaper coffee from Central America and resold it as the more expensive Kona coffee. This incident temporarily damaged the reputation of the Kona region and prompted the creation of new Hawaii State Laws to protect regional names (Schoenholt, 2001). Chemical approaches for testing the provenance of coffee beans may provide a means to test the authenticity of beans marketed as being from premium growing regions, and therefore go some way to protecting the supply chain of premium coffee brands.

Previous studies have investigated a range of chemical analyses as a means to distinguish the geographic origin of coffee beans, including fatty acid analyses (Bertrand *et al*., 2008; Villarreal *et al*., 2009), stable isotope ratios (Rodrigues *et al*., 2011b; Rodrigues *et al*., 2009; Weckerle *et al*., 2002), and elemental profiling (Rodrigues *et al*., 2011a; Watling *et al*., 2010). The basis for these approaches is that spatial differences in these chemical measurements for plants provide the foundations for a tool to determine the provenance of produce. Fatty acids have shown promise for discriminating coffee bean origin over small scales (e.g. plantations within one growing region/country, Bertrand *et al*., 2008; Villarreal *et al*., 2009), but to the best of our knowledge have not been tested at a global scale. Stable isotope ratios of coffee beans and their extracts have been shown to vary across large geographical scales. For example, Weckerle *et al*. (2002) measured $\delta^{13}C$, $\delta^{2}H$ and $\delta^{18}O$ of caffeine and found $\delta^{2}H$ and $\delta^{18}O$ to give high classification success rates (>90%) at the "continent" scale. However, the scope of the study was limited by a small overall sample size and only included samples from two continents. Rodrigues *et al*. (2009) similarly measured $\delta^{13}C$, $\delta^{15}N$ and $\delta^{18}O$ of green beans from over 20 geographic locations and found that distinct patterns emerged based on multivariate analyses of these isotopic ratios. In a following study, Rodrigues *et al*. (2011b) found statistical analysis of strontium and oxygen isotope ratios to accurately predict the provenance of coffee beans by country. Watling *et al*. (2010) were able to determine the continent and country of origin of coffee bean samples based on elemental profiles with high classification success rates (often >90%). Research in determining the provenance of coffee beans is now advancing towards using multivariate statistics to combine multiple chemical approaches. For example, Rodrigues *et al*. (2011a) found that samples of coffee beans could be grouped in multivariate space into growing regions from Hawaii using multiple isotopic ratios (of C, N, S, O and Sr) and trace elements. This approach is an advance on previous studies where chemical techniques have been individually tested as tools for predicting sample origin.

The current study builds on this body of work by using a multivariate framework (CAP) to select the analytical approach(es) best able to differentiate the geographic origin of coffee beans at multiple scales. The study included elemental profiling, stable isotope analysis

($\delta^{13}C$, $\delta^{15}N$ and $\delta^{2}H$), and fatty acid profiling. The first test determined which combination of methods was best able to predict the country of origin of samples of coffee beans from throughout the world. The second test determined which combination of methods was best able to distinguish between coffee beans from the growing region of Kona, Hawaii and beans collected from a range of coffee-producing nations. As all of these chemical signatures have been shown previously to vary geographically in plants, the hypothesis was that a combination of all of the above measures would provide the most accurate prediction of bean origin.

MATERIALS AND METHODS

Sample collection and preparation: Samples of green coffee beans from Hawaii, United States of America, were obtained direct from coffee growers (Hawaii n=16 from the growing region of Kona). Samples of green beans from Brazil (n=6), Colombia (n=6), Guatemala (n=6), Ethiopia (n=7), Tanzania (n=6), Uganda (n=6), Indonesia (n=6) and Papua New Guinea (n=6) were supplied by a coffee bean importer based in New Zealand. Beans were dried at 40°C until a constant mass was obtained, then ground using a commercial coffee grinder. For each sample ~5 grams of ground beans were set aside for trace element analysis prior to the sample being screened (Endecotts Test Sieves, England) and powder finer than 420 µm being retained for analyses of fatty acids and bulk stable isotopes.

Elemental analysis: Triplicate samples of 500 mg of ground coffee bean powder were placed in Teflon vessels with 10 mL quartz-distilled nitric acid ($qHNO_3$). Samples were digested using a microwave accelerated reaction system (MARS), which employed closed vessel acid digestions (CEM Corporation). The heating program began at 1600 W, ramping to 200°C, which was held for 15 minutes. Once the acid digestion was complete, the samples were quantitatively transferred to 50 mL polypropylene tubes and dried down for 8 hours at 105 °C. The resulting residues were redissolved in 10 mL 2 % $qHNO_3$. Trace element concentrations were quantified using ICP-MS (Agilent 7500cs, Agilent Technologies), at the Community Trust of Otago Centre for Trace Element Analysis, University of Otago, Dunedin, New Zealand. Calibration stability was ensured by using a single robust tune and digested samples were presented to the instrument as 2% HNO_3 solutions after 10× dilution. A cocktail of 6 elements were added as internal standards (Be, Ge, Rh, In, Tb, Bi). These elements were selected to span the mass range of interest and on the basis of their low natural concentrations in the samples. The monitoring of these internal standards over the entire mass range allows mass biases to be identified. The coffee samples contained very high concentrations of K, approaching 1 mg ml^{-1} in the reconstituted residue. Tests were conducted on solutions prepared to mimic the concentrations of the target elements as found in the coffee samples but with [K] ranging from zero to 1 mg ml^{-1}. These showed no significant interference by K on the elements of interest at these levels. The 10-fold dilution of the samples ensured the [K] was always well below this level and that the total dissolved solids concentrations were also low.

The data quality was determined using the certified corn meal reference material CRM-CM-A (High Purity Standards). The certified concentration of the reference materials is measured after total digestion with Hydrofluoric Acid (HF). For the purposes of this study, digestion using HF was not undertaken due to safety concerns associated with the use of HF. The use of only HNO3 for digestion of the corn meal reference material does not cause breakdown of the silicate matrices in the plant material and hence the results presented here should be considered the acid-extractable portion rather than the total concentration. Corn meal contains appreciable amounts of these silicates, which has resulted in low recovery of Fe (70%) and reduced precision of Al and Fe. All other elements were recovered in good agreement (± 10%) with the certified values. No insoluble residues were observed for the coffee dissolutions and the precisions of the Al and Fe results were acceptable (Table 1).

Isotope ratio determination: Samples for $\delta^{15}N$ and $\delta^{13}C$ analyses were prepared by weighing triplicate 0.8 ± 0.1 mg subsamples (<420 µm) into tin foil capsules. Nitrogen and carbon isotopes were assayed by combustion of the whole material to N_2 and CO_2 gas in a Carlo Erba NC2500 elemental analyser (CE Instruments, Milan) using helium carrier gas, enriched with oxygen. The gases were separated on a packed molecular sieve GC column and sent sequentially to the inlet of a Europa Scientific "20/20 Hydra" (Sercon, UK) isotope ratio mass spectrometer (IRMS) in continuous flow mode. Raw isotope ratios were normalized to the international scales using laboratory standard materials, assayed with the unknown samples, which have been calibrated against IAEA reference materials. The laboratory

standard used for nitrogen and carbon analysis was EDTA (Elemental Microanalysis Ltd, UK) with $\delta^{15}N$ = -0.82‰ and $\delta^{13}C$ = -38.59‰. The typical precision for analysis of control materials is ±0.3‰ for $\delta^{15}N$ and ±0.1‰ for $\delta^{13}C$.

Hydrogen isotope ratios were measured using steam equilibration with two water vapors of known $\delta^{2}H$ to enable correction for exchangeable hydrogen. The method used (Wassenaar, pers comm) is similar to that of Sauer *et al.* (2009) with the differences that a static volume of water vapor was used rather than a continuous flow in nitrogen carrier and drying of samples after equilibration was achieved by heating under vacuum instead of in a nitrogen gas purge. Hextuplicate samples of 0.6 ± 0.1 mg (<420 μm), were weighed into silver capsules and the capsules crimped loosely to allow exchange of water vapor. Three replicates of each sample were placed into autosampler trays with standards and controls, sealed into vacuum chambers and exposed to either of two water vapors, in 100-fold molar excess of the exchangeable hydrogen, at 120 °C for 2 hours. After equilibration, excess moisture was removed by drying under vacuum for 2-4 hours using a rotary vacuum pump and a liquid nitrogen trap as the samples were slowly cooled to room temperature. The $\delta^{2}H$ of the equilibration waters used were -274‰ and +60‰. Equilibrated and dried samples were converted to hydrogen gas by pyrolysis over glassy carbon at 1400°C in the Thermo (Bremen, Germany) TC/EA coupled to a Thermo Delta V IRMS in continuous flow mode. Raw values for $\delta^{2}H$ of total hydrogen were corrected using one point calibration against IAEA-CH7 primary standard material assayed with each batch of samples at 12-sample intervals. The triplicate samples for each steam treatment were averaged and paired to calculate $\delta^{2}H$ for the non-exchangeable hydrogen using the equations of Schimmelmann *et al*. (1999). Repeat measurements of a coffee control sample over 4 analysis batches yielded results of -32.0 ± 0.4‰.

Fatty acid determination: Lipid was extracted from ~5 mg subsamples of ground green beans by a modified Bligh and Dyer procedure (1959). The lipid extract was evaporated under nitrogen and reconstituted in 50 μL chloroform. All chloroform and hexane used were pesticide residue analysis grade and methanol was analytical reagent grade (BDH Chemicals, Poole, UK). The lipid fraction was treated with 0.1% BF_3 in methanol solution (Supelco, USA) and underwent acid-catalysed esterification at 70°C for 20 min. Fatty acid methyl esters (FAMEs) were extracted in hexane/water, with the hexane-containing layer removed and stored at 8°C.

Fatty acid composition was determined by gas chromatography (GC) on a 6850N Network GC System (Agilent Technologies, Santa Clara, CA, USA) equipped with a flame ionization detector. FAMEs were separated on an Equity-1 capillary column, 15m x 0.1 mm i.d., 0.1 μm film (Sigma-Aldrich/Supelco). The column oven temperature was held at 30°C for 3 minutes, ramped to 270°C at 15°C min^{-1}, then ramped to 290°C at 5°C min^{-1}. Fatty acid peaks were identified by retention time matching with commercially available FAME standards (NuCheck Prep, Elysian, Minnesota and Sigma, St. Louis, Missouri). Peak identities were then confirmed by mass spectrometry. Peak areas were calculated and reported as percentage of total fatty acids. Fatty acid notation describes the carbon chain length, followed by a colon and the number of double bonds. The position of the first double bond is specified from the methyl end (ω), with all subsequent bonds methylene-interrupted.

Statistical analysis: The statistical analyses were designed to answer questions at two geographical scales: 1) Which combination of chemical variables is best able to discriminate samples from Kona, Hawaii from those originating from other global growing regions? and 2) Which combination of chemical variables is best able to predict the country of origin of samples of green coffee beans? These two spatial scales were deemed appropriate as the only samples obtained directly from coffee growers were those from Hawaii, and Kona coffee has been a target for fraud in the past (Schoenholt, 2001).

Multivariate analyses were conducted using the software PRIMER+PERMANOVA (v6.1.12, Primer-E Ltd., Plymouth, UK). PERMANOVA is a permutation-based technique, and therefore does not assume data to be normally distributed, which is useful when dealing with large data sets of many variables, such as in the current study. CAP is a constrained ordination, and discriminates among a priori groups, similar to linear discriminant analysis. Whilst originally popular in the ecological sciences, PRIMER+PERMANOVA is now more widely used including the disciplines of physics (de Dios Rivera *et al*., 2012), geology/geochemistry (Keegan *et al*., 2012; Lyla *et al*., 2012), paleoanthropology (Harvati *et al*., 2011), metabolomics (Vidoudez and Pohnert, 2012), microbiology (Di Cagno *et al*., 2011) and medicine

(Smith *et al*., 2012) for example. Samples of coffee beans were categorized by their country of origin and as being either Kona, Hawaii or elsewhere ("World"). Data were not transformed with the exception of Ni measurements, which were square root transformed to achieve a more normal distribution of variances. Results were checked for co-variance and one of any pair that was highly correlated (>0.96) was excluded from the analysis to reduce redundancy in the model (Campbell *et al*., 2009), resulting in a subset of 13 elements being included in the analyses. Data were normalized to account for differences of scale, prior to resemblance matrices being constructed based on Euclidean distance measures. Initial investigations were based on multidimensional scaling (MDS) plots to look for patterns in the unconstrained data. One-way PERMANOVA tests were conducted on resemblance matrices to test for significant differences among geographic groups (random factor, 9999 permutations). Significant results were followed by pairwise PERMANOVA tests. An indication of samples grouping by origin was followed with CAP using combinations of 1, 2, 3 or 4 of the chemical variables (e.g. one combination of 2 chemical variables is fatty acids and δ^2H). The CAP procedure positioned axes through the multivariate data cloud that were the best at discriminating among pre-defined groups. The procedure included cross validation in a leave one out procedure to predict group membership (e.g. country of origin) and thus obtain overall classification success rates. The PERMANOVA software has been designed to avoid model over-fitting, by enabling the operator to select the size of the subset of principal component axes (*m*) (Anderson and Willis, 2003). For each model a value of *m* was selected that gave relatively low residual sum of squares and high classification success, whilst also considering the relative amounts of CAP axes and samples. The overall classification success rates of the different combinations of chemical variables were then compared. For the purposes of this comparison, chemical variables that were measured simultaneously were grouped (i.e. δ^{15}N and δ^{13}C, all elemental concentrations, all fatty acid compounds). CAP models were rebuilt with smaller sets of country groupings to achieve maximum classification success.

RESULTS AND DISCUSSION

Effectiveness of techniques in discriminating coffee beans from Kona, Hawaii: The mean values of elemental concentrations, stable isotope ratios and fatty acids for beans from Kona, Hawaii versus a global average are provided in Table 1. δ^2H ranged from -102.8 ± 3.0‰ (Papua New Guinea) to -27.9 ± 0.8‰ (Ethiopia). δ^{13}C and δ^{15}N were less variable, ranging 3.3‰ and 3.7‰ respectively. Six fatty acids were identified, with palmitic acid (16:0) being the most abundant, ranging in relative quantity from 43.8 ± 0.54% (Brazil) to 46.2 ± 0.29% (Uganda). A polyunsaturated C_{18} fatty acid (linoleic acid; 18:2(ω6?)) was the next most abundant, ranging from 31.0% (Brazil, Uganda, Kona) to 34.0 ± 0.97% (Guatemala). For multiple combinations of variables, samples of coffee beans from Kona tended to cluster in multivariate space in relation to samples from the other growing regions (e.g. Fig.1). Significant differences (PERMANOVA, $p \leq 0.005$) between sample origins at this spatial scale were detected using all combinations of chemical variables, with the exception of δ^2H (p(permutational) = 0.347). Constrained multivariate analyses using the CAP procedure generated a classification success rate (%) that enabled the comparison of multiple chemical approaches. 100% classification success was achieved using multiple combinations of chemical variables (Table 2). All of the combinations of chemical variables returning 100% classification success included elemental concentration data, making variability in concentrations of elements the most effective single discriminator of samples of beans at this spatial scale. Variability in δ^{13}C & δ^{15}N also provided a reasonable predictor of origin (85% classification success).

Effectiveness of techniques in determining country of origin: In unconstrained multivariate space, the samples tended to group by country of origin (Fig. 2), confirming the ability of these chemical measures to discriminate samples at this spatial scale. Using the variables that provided the best separation amongst countries (elemental concentrations, δ^{13}C & δ^{15}N), there were significant differences among samples from all countries (PERMANOVA df = 8, pseudo-F = 12.439, p(permutational) = 0.0001; pairwise PERMANOVA p(permutational) < 0.005, Brazil ≠ Uganda ≠ Guatemala ≠ Indonesia ≠ PNG ≠ Ethiopia ≠ Tanzania ≠ Colombia ≠ Hawaii). Using CAP, variability in elemental concentrations provided the highest classification success rate out of any of the approaches when used alone (80%), providing a success rate 20% higher than the next best indicator, δ^2H (60%).

Table 1. Mean values for each chemical measure for each region with associated error (1 SE). "World" is a global average of all samples except those from Kona, Hawaii.

	Brazil	Colombia	Ethiopia	Guatemala	Indonesia	Papua New Guinea	Tanzania	Uganda	Kona, Hawaii	World
Elements										
B (μg g^{-1})	5.07 ± 0.40	6.34 ± 0.36	7.48 ± 0.33	5.17 ± 0.15	4.70 ± 0.19	3.87 ± 0.18	2.72 ± 0.11	3.08 ± 0.25	8.95 ± 0.50	4.86 ± 0.24
Mg (mg g^{-1})	1.91 ± 0.02	1.89 ± 0.01	1.68 ± 0.02	1.88 ± 0.05	1.76 ± 0.03	1.97 ± 0.06	1.84 ± 0.05	1.81 ± 0.06	2.10 ± 0.04	1.84 ± 0.02
Al (μg g^{-1})	6.19 ± 1.54	2.04 ± 0.37	4.43 ± 1.51	1.74 ± 0.73	6.60 ± 0.92	4.90 ± 1.00	0.74 ± 0.34	1.71 ± 0.54	0.26 ± 0.15	3.56 ± 0.45
K (mg g^{-1})	17.5 ± 0.39	17.30 ± 0.29	16.9 ± 0.13	15.8 ± 0.35	16.7 ± 0.35	15.9 ± 0.47	17.5 ± 0.42	18.0 ± 0.43	16.6 ± 0.19	17.0 ± 0.16
Ca (mg g^{-1})	0.91 ± 0.04	1.00 ± 0.03	0.80 ± 0.07	0.98 ± 0.02	0.94 ± 0.02	0.96 ± 0.01	0.74 ± 0.03	0.75± 0.03	0.85 ± 0.02	0.88 ± 0.02
Mn (μg g^{-1})	26.1 ± 2.00	22.8 ± 2.57	12.8 ± 0.37	31.0 ± 1.02	23.7 ± 1.04	22.7 ± 0.78	24.9 ± 1.76	17.5 ± 1.68	16.7 ± 0.83	22.5 ± 0.91
Fe (μg g^{-1})	29.7 ± 1.06	27.9 ± 0.40	22.7 ± 0.78	26.3 ± 1.04	27.3 ± 0.56	28.4 ± 0.97	19.7 ± 0.42	23.4 ± 0.84	29.2 ± 1.12	25.6 ± 0.53
Ni (μg g^{-1})	0.25 ± 0.02	0.28 ± 0.07	0.23 ± 0.06	0.55 ± 0.02	0.21 ± 0.03	0.22 ± 0.02	0.67 ± 0.03	0.19 ± 0.06	0.69 ± 0.10	0.25 ± 0.03
Cu (μg g^{-1})	12.0 ± 0.32	11.9 ± 0.24	9.89 ± 0.29	10.2 ± 0.29	10.3 ± 0.25	11.7 ± 0.28	9.83 ± 0.27	10.4 ± 0.20	10.5 ± 0.38	10.8 ± 0.15
Zn (μg g^{-1})	4.13 ± 0.23	4.78 ± 0.17	3.40 ± 0.17	4.06 ± 0.37	4.55 ± 0.13	4.67 ± 0.14	2.84 ± 0.06	4.63 ± 0.67	5.52 ± 0.34	4.12 ± 0.14
Rb (μg g^{-1})	19.2 ± 1.55	32.9 ± 2.68	12.1 ± 1.72	25.1 ± 4.20	52.2 ± 1.92	33.0 ± 2.22	54.3 ± 1.72	16.2 ± 3.34	13.8 ± 1.90	30.2 ± 2.30
Sr (μg g^{-1})	3.09 ± 0.22	7.14 ± 0.78	1.61 ± 0.17	3.82 ± 0.35	4.98 ± 0.21	4.32 ± 0.21	4.24 ± 0.19	4.41 ± 0.18	2.79 ± 0.26	4.15 ± 0.24
Ba (μg g^{-1})	1.88 ± 0.16	5.69 ± 0.55	1.78 ± 0.27	3.61 ± 0.52	3.23 ± 0.32	1.56 ± 0.17	1.84 ± 0.20	3.35 ± 0.13	0.81 ± 0.11	2.85 ± 0.22
Stable isotopes (‰)										
$\delta^{13}C$	-27.43 ± 0.18	-28.82 ± 0.32	-26.52 ± 0.12	-29.10 ± 0.12	-27.12 ± 0.09	-29.45 ± 0.13	-26.14 ± 0.06	-27.99 ± 0.31	-26.27 ± 0.17	-27.80 ± 0.18
$\delta^{15}N$	3.74 ± 0.23	1.91 ± 0.54	4.61 ± 0.43	1.62 ± 0.38	2.94 ± 0.21	1.62 ± 0.16	3.84 ± 0.15	5.31 ± 0.26	1.72 ± 0.27	3.23 ± 0.22
$\delta^{2}H$	-74.18 ± 3.54	-68.30 ± 4.73	-27.90 ± 0.84	-64.12 ± 3.37	-77.72 ± 1.41	-102.77 ± 3.01	-55.52 ± 1.01	-32.60 ± 0.91	-56.46 ± 0.67	-62.17 ± 3.45
Fatty acids (relative abundance (%))										
16:0	43.8 ± 0.54	45.2 ± 0.39	44.8 ± 0.40	44.0 ± 0.74	44.8 ± 0.30	44.8 ± 0.25	44.5 ± 0.28	46.2 ± 0.29	45.9 ± 0.19	44.8 ± 0.17
18:2(ω6?)	31.0 ± 0.25	31.9 ± 0.36	31.1 ± 0.25	34.0 ± 0.97	31.7 ± 0.27	33.0 ± 0.50	31.9 ± 0.23	31.0 ± 0.36	31.0 ± 0.22	31.9 ± 0.21
18:1ω9c	9.65 ± 0.23	8.60 ± 0.29	9.82 ± 0.21	8.33 ± 0.27	8.68 ± 0.32	7.87 ± 0.12	9.05 ± 0.13	8.13 ± 0.26	8.74 ± 0.07	8.79 ± 0.12
18:1ω7c	1.16 ± 0.07	1.14 ± 0.12	1.17 ± 0.03	1.04 ± 0.09	1.00 ± 0.08	1.20 ± 0.08	1.09 ± 0.07	1.08 ± 0.08	1.24 ± 0.03	1.11 ± 0.03
18:0	9.56 ± 0.19	8.63 ± 0.13	8.97 ± 0.10	8.20 ± 0.22	9.02 ± 0.16	8.74 ± 0.09	9.28 ± 0.09	8.97 ± 0.22	9.04 ± 0.12	8.92 ± 0.08
20:0	4.87 ± 0.19	4.55 ± 0.24	4.11 ± 0.18	4.44 ± 0.24	4.76 ± 0.13	4.37 ± 0.13	4.19 ± 0.17	4.59 ± 0.20	4.11 ± 0.09	4.48 ± 0.07

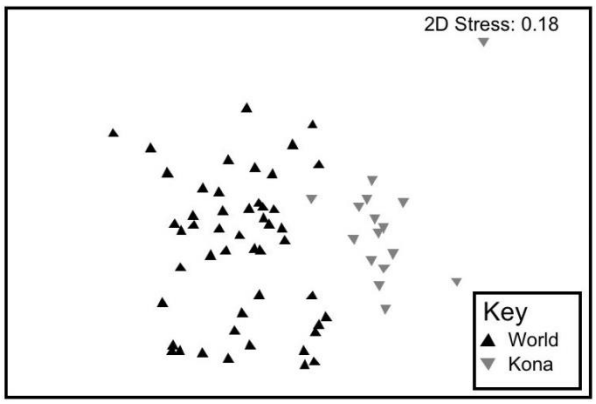

Figure 1. MDS plot based on the set of chemical variables that provided the best prediction of whether samples originated from Kona, Hawaii or other global locations (major, minor and trace elements).

Table 2. Overall classification success of the CAP procedure of samples originating from Kona or elsewhere, based on different combinations of chemical measures (1, 2, 3 and 4 measures used). The number of principal coordinate axes (*m*) used in each model is indicated. Combinations returning the highest rates of overall classification success are indicated in bold print. Note that $\delta^{13}C$ and $\delta^{15}N$ are measured simultaneously. TE trace, minor and major elements, FA fatty acids.

		# variables	*m*	Overall classification success
1.	TE	13	7	**100%**
	$\delta^{13}C$ & $\delta^{15}N$	2	2	85%
	$\delta^{2}H$	1	1	65%
	FA	6	5	75%
2.	TE + $\delta^{13}C$ & $\delta^{15}N$	15	6	**100%**
	TE + $\delta^{2}H$	14	10	**100%**
	TE + FA	19	6	**100%**
	$\delta^{13}C$ & $\delta^{15}N$ + $\delta^{2}H$	3	2	83%
	$\delta^{13}C$ & $\delta^{15}N$ + FA	8	7	92%
	$\delta^{2}H$ + FA	7	6	75%
3.	TE + $\delta^{13}C$ & $\delta^{15}N$ + $\delta^{2}H$	16	5	**100%**
	TE + $\delta^{13}C$ & $\delta^{15}N$ + FA	21	6	**100%**
	TE + $\delta^{2}H$ + FA	20	7	**100%**
	$\delta^{13}C$ & $\delta^{15}N$ + $\delta^{2}H$ + FA	9	7	92%
4.	TE + $\delta^{13}C$ & $\delta^{15}N$ + $\delta^{2}H$ + FA	22	7	**100%**

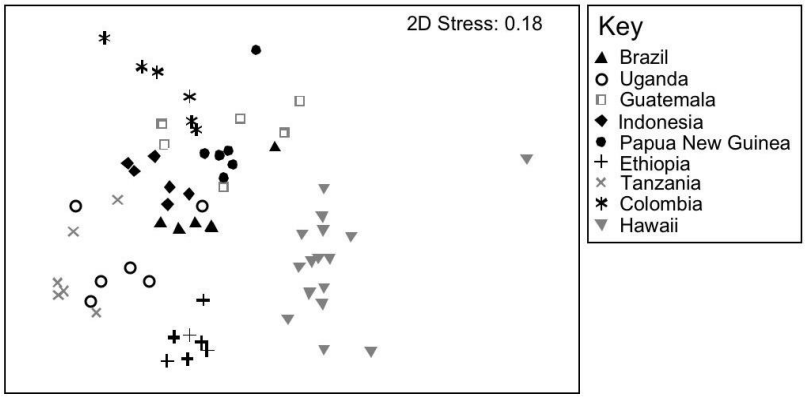

Figure 2. MDS plot based on the set of chemical variables that provided the best prediction of country of origin (major, minor and trace elements, $\delta^{13}C$ and $\delta^{15}N$).

Table 3. Overall country classification success of the CAP procedure based on different combinations of chemical measures (1, 2, 3 and 4 measures used). The number of principal coordinate axes (*m*) used in each model is indicated. Combinations returning the highest rates of overall classification success are indicated in bold print. Note that $\delta^{13}C$ and $\delta^{15}N$ are measured simultaneously. TE trace, minor and major elements, FA fatty acids.

		# variables	*m*	Overall classification success
1.	TE	13	4	**80%**
	$\delta^{13}C$ & $\delta^{15}N$	2	2	58%
	$\delta^{2}H$	1	1	60%
	FA	6	4	38%
2.	TE + $\delta^{13}C$ & $\delta^{15}N$	15	4	**86%**
	TE + $\delta^{2}H$	14	4	**86%**
	TE + FA	19	4	77%
	$\delta^{13}C$ & $\delta^{15}N$ + $\delta^{2}H$	3	3	74%
	$\delta^{13}C$ & $\delta^{15}N$ + FA	8	4	46%
	$\delta^{2}H$ + FA	7	4	68%
3.	TE + $\delta^{13}C$ & $\delta^{15}N$ + $\delta^{2}H$	16	4	**86%**
	TE + $\delta^{13}C$ & $\delta^{15}N$ + FA	21	4	78%
	TE + $\delta^{2}H$ + FA	20	4	82%
	$\delta^{13}C$ & $\delta^{15}N$ + $\delta^{2}H$ + FA	9	4	63%
4.	TE + $\delta^{13}C$ & $\delta^{15}N$ + $\delta^{2}H$ + FA	22	4	**83%**

The high classification success rate for elemental concentrations may be a function of the relatively high number of variables produced by this analysis (13) compared to e.g. $\delta^{2}H$ (1) (see Table 3).

Adding variability in either $\delta^{13}C$ & $\delta^{15}N$ or $\delta^{2}H$ to that of elemental concentrations improved the classification success rate to 86%. The combination of elemental concentrations, $\delta^{13}C$ & $\delta^{15}N$, and $\delta^{2}H$ did not improve the classification success any further. Using variability in elemental concentrations, $\delta^{13}C$ & $\delta^{15}N$, samples labeled as originating from Hawaii, Indonesia, Papua New Guinea and Colombia were all classified with 100% success (Fig. 3, Table 4). Using variability in elemental concentrations and $\delta^{2}H$, only samples from Indonesia and Colombia were all classified correctly with 100% success. Accordingly, the CAP model was rerun using variability in elemental concentrations, $\delta^{13}C$ & $\delta^{15}N$ (m = 4) with a reduced sample set (excluding samples from Hawaii, Indonesia, Papua New Guinea and Colombia) (Fig. 4a). This model achieved an overall classification success of 87%, providing 100% classification success for samples from Brazil and Guatemala, 83% classification success for samples from Uganda and Tanzania, and 71% classification success for samples from Ethiopia (Fig. 4b-c, see Table 4).

When using multivariate statistics to classify samples into pre-defined groups, consideration needs to be given to the tendency to over-fit models that are built using small data sets. Over-fitting occurs when a model is strongly influenced by each individual result. PERMANOVA+ for PRIMER software approaches this problem by allowing the operator to select the value of *m*, with consideration given to the sample size, the number of variables, the number of groups (in this case countries), and the resulting diagnostic statistics (Anderson and Willis, 2003). For example, the first model developed here to classify samples based on country of origin was based on 65 samples across 8 groups, so *m* was kept low (≤4) in order to avoid over-fitting.

One concern when using hydrogen stable isotopes in agricultural studies to predict provenance is the exchangeability of hydrogen atoms in organic materials such as plant matter. Up to 20% of the hydrogen atoms, being attached to the heteroatoms O, N or S and so being of acidic character, are able to exchange with atmospheric water vapor and/or water absorbed into the bulk material (Cormie *et al*., 1994; Schimmelmann *et al*., 1999; Wassenaar and Hobson, 2000). However, hydrogen attached to carbon is non-exchangeable and so preserves the signal recorded at the time and place of synthesis (Cormie *et al*., 1994).

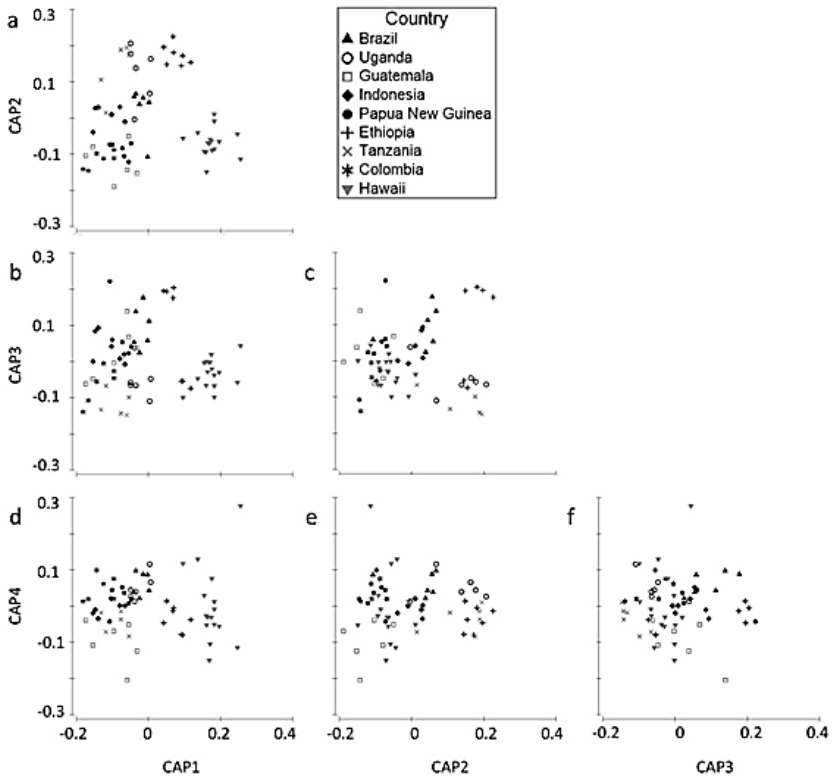

Figure 3. Pairs plot of the four CAP axes generated using Euclidean distance measures of major, minor and trace elements, and $\delta^{13}C$ and $\delta^{15}N$ data. The largest amount of discrimination among a priori defined groups is encompassed in CAP1, with subsequent axes decreasing in ability to discriminate groups: a) CAP1 vs. CAP2, b) CAP1 vs. CAP3, c) CAP2 vs. CAP3, d)CAP1 vs. CAP4, e) CAP2 vs. CAP4, and CAP3 vs. CAP4.

Table 4. Classification success rates of samples by country based on values of multiple major, minor and trace elements, $\delta^{13}C$ and $\delta^{15}N$. For both CAP analyses $m = 4$.

	Using all countries	Using subset of countries
Brazil	67%	100%
Uganda	67%	83%
Guatemala	67%	100%
Indonesia	100%	-
Papua New Guinea	100%	-
Ethiopia	71%	71%
Tanzania	83%	83%
Colombia	100%	-
Hawaii	100%	-

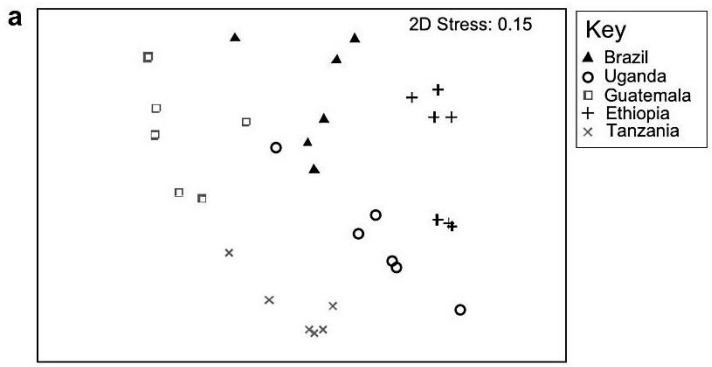

Figure 4.a. MDS plot CAP plots based on major, minor and trace elements, $\delta^{13}C$ and $\delta^{15}N$ for a subset of countries.

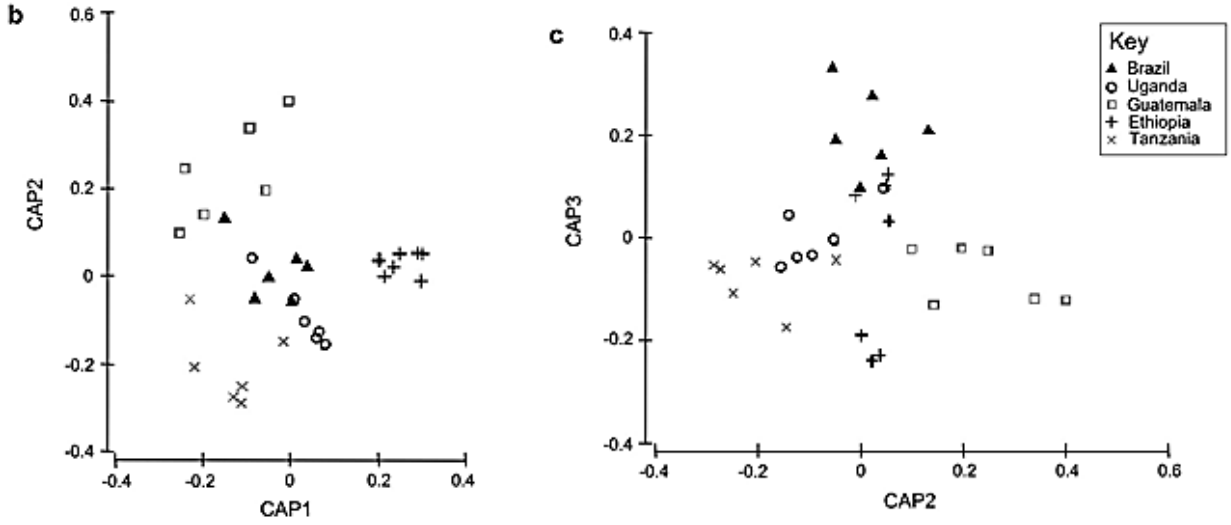

Figure 4. b and c. CAP plots based on major, minor and trace elements, $\delta^{13}C$ and $\delta^{15}N$ for a subset of countries.

As described above, the isotope ratio of atmospheric water vapor changes with location and season and so the same material measured in different laboratories or at other times may give a different result. Methods such as those used in the current study are currently being developed to eliminate the variation between laboratories (Qi and Coplen, 2011). Defining and measuring the non-exchangeable portion of the total hydrogen in organic materials is important to achieve inter-laboratory comparability and to access only hydrogen that was assimilated by the organism during its growth.

For the country of origin study, the assumption was made that origin labels provided by the coffee bean supplier were correct. It is of course possible that some of the samples may have been mislabeled somewhere along the supply chain, but given that none of these samples were from brands that command a premium price (such as Ethiopian Yirgacheffee or Jamaican Blue Mountain), the chance of mislabeling is considered to be low. However, samples of coffee beans were obtained direct from growers in the Kona region of Hawaii. Given that the authenticity of these samples is certain, it is appropriate to test whether they can be distinguished from samples taken from other regions in Hawaii and throughout the rest of our global growing regions.

For both approaches – discriminating Kona, Hawaii from other global regions and discriminating country of origin - one must take care when interpreting the results as potential temporal variability in chemical signatures has not been considered. Whilst fatty acid composition has shown promise in distinguishing geographic origin (Bertrand *et al*., 2008), it appears to be influenced by the temperature experienced during fruiting, and therefore the fatty acid composition of beans at a single location can vary among growing seasons (Villarreal *et al*., 2009). Similarly, temporal fluctuations in the isotope ratio of atmospheric carbon (Flanagan and Ehleringer, 1998) and seasonality in biological processes have been demonstrated to cause temporally variable $\delta^{13}C$ values in grassland and forest plants (Buchmann *et al*., 1997; Smedley *et al*., 1991). As the trace element composition of plants reflects that of the soil and parent rock (Watling *et al*., 2010) it is likely that this measure has low seasonal variability (Pilgrim *et al*., 2010), provided the agricultural method remains consistent (e.g. not shifting from organic to conventional agriculture) (Gunderson *et al*., 2000).

CONCLUSIONS

Variability in trace element concentrations allowed samples from Kona to be distinguished from those grown elsewhere in the world. A combination of variability in trace element concentrations and $\delta^{13}C$ and $\delta^{15}N$, was sufficient to correctly determine the country of origin of samples of green coffee beans with a high degree of confidence. Variations in trace element concentrations were the single best discriminator of origin at both of these spatial scales; however, the other measures also showed promise when combined with trace element values. The study demonstrates the usefulness of multivariate statistics in comparing the effectiveness of multiple chemical techniques for determining the provenance of produce across multiple scales. The most effective measures to use will depend on the product in question and the scale required, and should thus be approached on a case-by-case basis.

ACKNOWLEDGEMENTS

Our sincere thanks to the many Hawaiian coffee growers

who supplied samples for this study, to Loren Gautz for facilitating sample collection in Hawaii, to David Barr for assistance with trace element analysis, to Marti Anderson for guidance with multivariate statistics, and the Department of Chemistry at the University of Otago for supporting this research. Comments provided by two anonymous reviewers were much appreciated. Sample analyses were conducted under a summer studentship awarded to M. Garland by Oritain Global Ltd. R. McLeod was funded by a Foundation for Research, Science and Technology (FRST) Science and Technology Postdoctoral Fellowship (UOOX0814).

REFERENCES

Anderson, M.J., T.J. Willis. 2003. Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. Ecol. 84: 511-525.

Baroni, M.V., N.S. Podio, R.G. Badini, M. Inga, H.A. Ostera, M. Cagnoni, E. Gallegos, E. Gautier, P. Peral-García, J. Hoogewerff and D.A. Wunderlin. 2011. How much do soil and water contribute to the composition of meat? A case study: Meat from three areas of Argentina. J. Agric. Food Chem. 59: 11117-11128.

Baxter, M.J., H.M. Crews, M.J. Dennis, I. Goodall and D. Anderson. 1997. The determination of the authenticity of wine from its trace element compostion. Food Chem. 60: 443-450.

Bertrand, B., D. Villarreal, A. Laffargue, H. Posada, P. Lashermes and S. Dussert. 2008. Comparison of the effectiveness of fatty acids, chlorogenic acids, and elements for the chemometric discrimination of coffee (*Coffea arabica* L.) varieties and growing origins. J. Agric. Food Chem. 56: 2273-2280.

Bligh, E.G. and W.J. Dyer. 1959. A rapid method of total lipid extraction and purification. Can J. Biochem. Physiol. 37: 911-917.

Buchmann, N., J.M. Guehl, T.S. Barigah and J.R. Ehleringer. 1997. Interseasonal comparison of CO_2 concentrations, isotopic composition, and carbon dynamics in an Amazonian rainforest (French Guiana). Oecologia. 110: 120-131.

Camin, F., L. Bontempo, K. Heinrich, M. Horacek, S.D. Kelly, C.F.T. Schlicht, F.J. Monahan, J. Hoogewerff and A. Rossmann. 2007. Multi-element (H,C,N,S) stable isotope characteristics of lamb meat from different European regions. Anal. Bioanal Chem. 389: 309-320.

Campbell, G.P., J.M. Curran, G.M. Miskelly, S. Coulson, G.M. Yaxley, E.C. Grunsky and S.C. Cox. 2009. Compositional data analysis for elemental data in forensic science. Forensic Sci Int. 188: 81-90.

Carcea, M., P. Brereton, R. Hsu, S.D. Kelly, N. Marmiroli, F. Melini, C. Soukoulis and D. Wenping. 2009. Food authenticity assessment: ensuring compliance with food legislation and traceability requirements. Qual. Assur. Saf. Crop. 1: 93-100.

Cormie, A.B., H.P. Schwarcz and J. Gray. 1994. Determination of the hydrogen isotopic composition of bone collagen and correction for hydrogen exchange. Geochim. Cosmochim. Ac. 58: 365-375.

de Dios Rivera, J., G.M. Davies and W. Jahn. 2012. Flammability and the heat of combustion of natural fuels: A review. Combust. Sci. Technol. 184: 224-242.

Di Cagno, R., M. De Angelis, I. De Pasquale, M. Ndagijimana, P. Vernocchi, P. Ricciuti, F. Gagliardi, L. Laghi, C. Crecchio, M.E. Guerzoni, M. Gobbetti and R. Francavilla. 2011. Duodenal and faecal microbiota of celiac children: molecular, phenotype and metabolome characterization. BMC Microbiol. 11: 219-240.

Flanagan, L.B. and J.R. Ehleringer. 1998. Ecosystem–atmosphere CO_2 exchange: interpreting signals of change using stable isotope ratios. Trends in Ecology and Evolution. 13: 10-14.

Gunderson, V., I.E. Bechmann, A. Behrens and S. Stürup. 2000. Comparative investigation of concentrations of major and trace elements in organic and conventional Danish agricultural crops. 1. onions (*Allium cepa* Hysam) and peas (*Pisum sativum* Ping Pong). J. Agric. Food Chem. 48: 6094-6102.

Harvati, K., C. Stringer, R. Grün, M. Aubert, P. Allsworth-Jones and C.A. Folorunso. 2011. The Later Stone Age calvaria from Iwo Eleru, Nigeria: morphology and chronology. PLoS ONE. 6: 1-8.

Heaton, K., S.D. Kelly, J. Hoogewerff and M. Woolfe. 2008. Verifying the geographical origin of beef: The application of multi-element isotope and trace element analysis. Food Chem. 107: 506-515.

Joebstl, D., D. Bandoniene, T. Meisel and S. Chatzistathis. 2010. Identification of the geographical origin of pumpkin seed oil by the use of rare earth elements and discriminant analysis. Food Chem. 123: 1303-1309.

Keegan, E., M. Wallenius, K. Mayer, Z. Varga and G. Rasmussen. 2012. Attribution of uranium ore concentrates using elemental and anionic data. Appl. Geochem. 27: 1600-1609.

Kelly, S.D., K. Heaton and J. Hoogewerff. 2005. Tracing the geographical origin of food: The application of multi-element and multi-isotope analysis. Trends Food Sci. Tech. 16: 555-567.

Kropf, U., M. Korošec, J. Bertoncelj, N. Ogrinc, M. Nečemer, P. Kump and T. Golob. 2010. Determination of the geographical origin of Slovenian black locust, lime and chestnut honey. Food Chem. 121: 839-846.

Lyla, S., S. Manokaran and A. Khan. 2012. Petroleum hydrocarbon distribution in continental shelf region of southeast coast of India. Int. J. Sediment Res. 27: 73-83.

Maggi, L., M. Carmona, S.D. Kelly, N. Marigheto and G.L. Alonso. 2011. Geographical origin differentiation of saffron spice (*Crocus sativus* L. stigmas) – Preliminary investigation using chemical and multi-element (H, C, N) stable isotope analysis. Food Chem. 128: 543-548.

Martin, A.E., R.J. Watling and G.S. Lee. 2012. The multi-element determination and regional discrimination of Australian wines. Food Chem. 133: 1081-1089.

Martín, M.J., F. Pablos, A.G. González, M.S. Valdenebro and M. León-Camacho. 2001. Fatty acid profiles as discriminant parameters for coffee varieties differentiation. Talanta. 54: 291-297.

Ogrinc, N., K. Bat, I.J. Košir, T. Golob and R. Kokkinofta. 2009. Characterization of commercial Slovenian and Cypriot fruit juices using stable isotopes. J. Agric. Food Chem. 57: 6764-6769.

Pilgrim, T.S., R.J. Watling and K. Grice. 2010. Application of trace element and stable isotope signatures to determine the provenance of tea (*Camellia sinensis*) samples. Food Chem. 118: 921-926.

Qi, H. and T.B. Coplen. 2011. Investigation of preparation techniques for δ^2H analysis of keratin materials and a proposed analytical protocol. Rapid Commun. Mass Sp. 25: 2209-2222.

Rodrigues, C., M. Brunner, S. Steiman, G.J. Bowen, J.M. Noqueira, L. Gautz, T. Prohaska and C. Máguas. 2011a. Isotopes as tracers of the Hawaiian coffee producing regions. J. Agric. Food Chem. 59: 10239-10246.

Rodrigues, C., C. Máguas and T. Prohaska. 2011b. Strontium and oxygen isotope fingerprinting of green coffee beans and its potential to proof authenticity of coffee. Eur. Food Res. Technol. 232: 361-373.

Rodrigues, C.I., R. Maia, M. Miranda, M. Ribeirinho, J.M.F. Nogueira and C. Máguas. 2009. Stable isotope analysis for green coffee bean: A possible method for geographic origin discrimination. J. Food Comp. Anal. 22: 463-471.

Rummel, S., S. Hoelzl, P. Horn, A. Rossmann and C. Schlicht. 2010. The combination of stable isotope abundance ratios of H, C, N and S with $^{87}Sr/^{86}Sr$ for geographical origin assignment of orange juices. Food Chem. 118: 890-900.

Sauer, P.E., A. Schimmelmann, A.L. Sessions and K. Topalov. 2009. Simplified batch equilibration for D/H determination of non-exchangeable hydrogen in solid organic material. Rapid Commun. Mass Sp. 23: 949-956.

Schellenberg, A., S. Chmielus, C. Schlicht, F. Camin, M. Perini, L. Bontempo, K. Heinrich, S.D. Kelly, A. Rossmann, F. Thomas, E. Jaminb and M. Horacek. 2010. Multielement stable isotope ratios (H, C, N, S) of honey from different European regions. Food Chem. 121: 770-777.

Schimmelmann, A., M.D. Lewan and R.P. Wintsch. 1999. D/H isotope ratios of kerogen, bitumen, oil, and water in hydrous pyrolysis of source rocks containing kerogen types I, II, IIS, and III. Geochim. Cosmochim. Ac. 63: 3751-3766.

Schoenholt, D.N. 2001. We was robbed! The Kona Kai scandal and its aftermath. Tea and Coffee Trade Journal. 173: 85-96.

Smedley, M.P., T.E. Dawson, J.P. Comstock, L.A. Donovan, D.E. Sherrill, C.S. Cook and J.R. Ehleringer. 1991. Seasonal carbon isotope discrimination in a grassland community. Oecologia. 85: 314-320.

Smith, B.C., T. McAndrew, Z. Chen, A. Harari, D.M. Barris, S. Viswanathan, A.C. Rodriguez, P. Castle, R. Herrero, M. Schiffman and R.D. Burk. 2012. The cervical microbiome over 7 years and a comparison of methodologies for its characterization. PLoS ONE. 7: 1-13.

Vidoudez, C. and G. Pohnert. 2012. Comparative metabolomics of the diatom *Skeletonema marinoi* in different growth phases. Metabolomics. 8: 654-669.

Villarreal, D., A. Laffargue, H. Posada, B. Bertrand, P. Lashermes and S. Dussert. 2009. Genotypic and environmental effects on coffee (*Coffea arabica* L.) bean fatty acid profile: impact on variety and origin chemometric determination. J. Agric. Food Chem. 57: 11321-11327.

Wassenaar, L.I. and K.A. Hobson. 2000. Improved method for determining the stable-hydrogen isotopic composition (δD) of complex organic materials of environmental interest. Environ. Sci. Technol. 34: 2354-2360.

Watling, R.J., G.S. Lee, C.J. Scadding, T.S. Pilgrim, R.L. Green, A.E. Martin, C.D. May, and J.L. Valentin. 2010. The application of solution and laser ablation based ICP-MS and solution based AES for the provenance determination of selected food and drink produce. The Open Chem. Biomed. Methods J. 3: 179-196.

Weckerle, B., E. Richling, S. Heinrich and P. Schreier. 2002. Origin assessment of green coffee (*Coffea arabica*) by multi-element stable isotope analysis of caffeine. Anal. Bioanal. Chem. 374: 886-890.

Ye, N.S. 2012. A minireview of analytical methods for the geographical origin analysis of teas (*Camellia sinensis*). Crit. Rev. Food Sci. Nutr. 52: 775-780.

Food and Agriculture Organization of the United Nations (2011) FAOSTAT.
URL http://faostat.fao.org/site/609, retrieved 15 September 2011.