



Determining the geographical origin of *Sechium edule* fruits by multielement analysis and advanced chemometric techniques



Melisa J. Hidalgo^a, Diana C. Fechner^a, Eduardo J. Marchevsky^{b,*}, Roberto G. Pellerano^{a,*}

^aInstituto de Química Básica y Aplicada del Nordeste Argentino (IQUIBA-NEA), UNNE-CONICET, Facultad de Ciencias Exactas y Naturales y Agrimensura, Av. Libertad 5400, (3400) Corrientes, Argentina

^bInstituto de Química San Luis (INQUISAL), UNSL-CONICET, Facultad de Química Bioquímica y Farmacia, Av. Ejército de los Andes 950, (5700) San Luis, Argentina

ARTICLE INFO

Article history:

Received 22 September 2015

Received in revised form 23 April 2016

Accepted 25 April 2016

Available online 26 April 2016

Keywords:

ICP OES

Mineral composition

Multivariate analysis

Classification

ABSTRACT

This paper describes the determination and evaluation of the major and trace element composition (Al, As, Ba, Ca, Cd, Co, Cu, Fe, K, Mg, Mn, Na, Pb, Sr and Zn) of *Sechium edule* (Jacq) Swartz fruits collected from four different places of production in Corrientes province, Argentina. Element concentrations were determined by using inductively coupled plasma optical emission spectrometry (ICP OES) after microwave digestion. The accuracy was confirmed with standard reference material of spinach leaves (NIST, 1570a) and spiking tests. Principal component analysis (PCA), linear discriminant analysis (LDA), k-nearest neighbors (kNN), partial least square-discriminant analysis (PLS-DA) and support vector machine (SVM) were applied to the results for discriminating the geographical origin of *S. edule* fruits. Finally, the LDA method was found to perform best with up to 90% accuracy rate based on the following elements: Ca, Ba, Cu, Mn, Na, Sr, and Zn.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Sechium edule (Jacq) Swartz fruits are appreciated for their nutritional quality. Those fruits contain multiple phytochemicals, as well as vitamins, minerals, and essential amino acids (Rao et al., 1990). These nutritional characteristics and the softness of the fruit flesh make it particularly suitable for its use as thickener in the manufacture of baby foods or hospital diets (Aung, Ball, & Kushad, 1990). In addition, the medicinal use of *S. edule* has also been documented throughout the literature. Several properties such as diuretic, antihypertensive, anti-inflammatory, and cardiovascular effects of this plant have been reported in the literature (Gordon, Guppy, & Nelson, 2000; Ordoñez, Gomez, Vattuone, & Isla, 2006).

The *S. edule* fruit is extensively produced and consumed in some Latin-American countries such as Brazil, Peru and Mexico. This crop can be grown with relative ease due to its adaptability to a wide range of climatic conditions (Saade, 1996). In other regions, including some U.S. states and Southern European countries, *S. edule* fruits are only produced on a small scale, as home-grown food crop. The characteristics of the fruit are highly depending on the geographical region and farming practices in which the fruit

garden has been located. Thus the origin of a fruit product is an important factor, affecting its quality.

Moreover, there are several important nutritional parameters affecting the quality of fruits. The mineral profile is an important quality parameter of horticultural products and depends of different factors, including soil type and environmental growing conditions. Thus, the mineral composition is a powerful tool for several types of food traceability. In this context, inductively coupled plasma optical emission spectrometry (ICP OES) has commonly been used to determine the geographic origin of crops (Mir-Marqués, Domingo, Cervera, & de la Guardia, 2015; Szymczycha-Madeja & Welna, 2013). ICP OES is a powerful tool for the simultaneous, rapid and accurate determination of metals and non-metals (inorganic elements) in a variety of samples at wide linear dynamic ranges (from trace to major elements). The application of chemometric methods combined with multielement analysis is an effective way to characterize and classify foods and vegetables from different geographical origin, allowing to detect the presence of fingerprint patterns in the samples (Barbosa et al., 2014; Drivelos, Higgins, Kalivas, Haroutounian, & Georgiou, 2014; Li et al., 2014).

The aim of this work was to evaluate the major and trace element data (Al, As, Ba, Ca, Cd, Co, Cu, Fe, K, Mg, Mn, Na, Pb, Sr and Zn) in combination with advanced chemometric techniques to investigate the potentiality of mineral composition as indicator of geographical origin of *S. edule* fruit samples, coming from four different Argentinean areas of production.

* Corresponding authors.

E-mail addresses: marchev@unsl.edu.ar (E.J. Marchevsky), roberto.pellerano@comunidad.unne.edu.ar (R.G. Pellerano).

2. Material and methods

2.1. Sample collection and preparation

In this paper, a total of 92 samples of *S. edule* fruits were analyzed for their major and trace element contents. The samples were provided from local growers of four cities from Corrientes province, namely: Bella Vista (Bv) (28°31'S, 59°02'W), Corrientes city (CT) (27°27'S, 58°46'W), Monte Caseros (MC) (30°07'S, 57°34'W) and Santo Tome (ST) (28°34'S, 56°06'W). A map of the studied sampling regions is provided in the [supplementary information \(SI-1\)](#). The province of Corrientes is located in the Northeast region of Argentina. It comes in form of a wedge between Uruguay and Brazil (to East), and Paraguay (to North). The Corrientes province covers a superficial area of 88.199 km². As part of the subtropical area of Argentine Mesopotamia, the province has humid climate, the annual average rainfalls is 1100 mm, with only two distinct seasons: summer from October to March and winter from May to August. April and September are transitional months in which temperatures are below the midsummer averages and minimums may dip below freezing.

The samples were obtained between September and October 2014. After collection the samples were immediately stored in the dark at 4 °C (refrigerator) until analysis. All fruits were first cleaned and washed with deionized water. The fresh fruits were processed in order to separate the seed and skin. The edible part of the fruits was weighed and mixed, and immediately after the samples were crushed and homogenized with a domestic mixer, and finally frozen at –20 °C in a freezer. Afterwards, they were freeze-dried for a minimum of 48 h at a chamber pressure of 0.05 mbar. The dried samples were pulverized with a domestic mixer, and the powdered samples were stored in polyethylene dry containers until analysis.

2.2. Reagents

Ultrapure deionized water (18.2 MΩ cm⁻¹) was obtained from a Milli-Q Pluswater purification system Millipore (Molsheim, France). Analytical reagent grade HNO₃ 65% and ultra-pure grade 30% (m/m) H₂O₂ were acquired from Sigma (St. Louis, MO, USA). Nitric acid was additionally cleaned by sub-boiling distillation. Prior to use, all plastic containers were soaked in 10% v/v sub-boiling HNO₃ for at least 24 h and then rinsed extensively with ultrapure deionized water. All types of glassware were avoided to prevent metal releases. All plastic containers, polyethylene flasks, pipette tips, PFA Teflon digestion vessels, and reagents exposed to samples, or standards, were tested for contamination.

2.3. Instrumentation

The samples were digested by microwave-assisted digestion using an Ethos One microwave system (Milestone, Chicago, USA), equipped with programmable power control (maximum power 1600 W) and segmented rotor HPR 1000/10 s (operating pressure up to 35 bar maximum; operating temperature 260 °C maximum) with 10 reaction vessels. Measurements were carried out with a Vista Pro (Varian, Australia) optical emission spectrometer with inductively coupled plasma and axial viewing configuration, equipped with a solid-state detector, Sturman-Masters mist chamber, and V-groove nebulizer. Selected emission wavelengths for the analysis were chosen according to previous interference studies (Larrea-Marín, Pomares-Alfonso, Gómez-Juaristi, Sánchez-Muniz, & de la Rocha, 2010). The lines that exhibited low interference, high analytical signal and background ratios were selected. These wavelengths were as follow: Al 309.271 nm, As 193.691 nm, Ba

455.403 nm, Ca 373.690 nm, Cd 226.499 nm, Co 228.615 nm, Cu 324.754 nm, Fe 238.204 nm, K 766.491 nm, Mg 285.213 nm, Mn 260.569 nm, Na 588.995 nm, Pb 220.350 nm, Sr 421.534 nm, and Zn 213.855 nm.

2.4. Sample digestion

Digestion of the samples and standard reference material (SRM) was carried out by weighing 500 mg of freeze-dried sample directly in the digestion vessels and adding 5 mL of sub-boiled HNO₃ (65%) and 2 mL of H₂O₂ (30%). Then, they were left open for 15 min for reacting in order to reduce the built up of gases which could increase the pressure inside the vessels. The power of the microwave for each digestion step was optimized, from 500 up to 1000 W. In all cases, the lowest power that provided the required temperature was selected. The digestion was complete in all cases, as checked by the analysis of SRM and recovery studies. After cooling at room temperature, all the digestion solutions were quantitatively transferred into plastic containers and diluted to 25 mL with ultrapure deionized water. A digestion blank was carried out in the same way. All samples (digest and blank solutions) were measured in triplicate.

2.5. Quality control

Reagent blanks were made regularly together with each batch of sample digestion. SRM was digested and measured together with the different sample batches in order to control the trueness of data found. Additionally, control standards were measured for every series of ten independent sample measurements.

The accuracy of the method was proved by analyzing a SRM: NIST 1570a (spinach leaves), obtained from the National Institute of Standards and Technology (Gaithersburg, MD, USA). This material was analyzed under the same conditions as the samples. Moreover, due to the Fe content was not certified in the NIST 1570a SRM, the accuracy of the determination of Fe was also evaluated by analysis of standard water solution for ICP, TraceCERT[®] obtained from Sigma (St. Louis, MO, USA). In addition, in order to further evaluate the proposed method, spike recovery studies were performed for all analytes in randomly selected samples.

Calibration curves were obtained at five different concentration levels in triplicate. Calibration ranges were modified according to the expected mineral concentration ranges. Limits of detection (LOD) and limit of quantification (LOQ) achieved with considered sample preparation procedure were calculated as three point three and ten times of the standard deviation of ten replicates of blank solutions, which were prepared in the same way as those of fruit samples.

The precision of the proposed procedure was also evaluated by measuring the repeatability and reproducibility. In the repeatability test (within-day precision), a digested sample was analyzed three times within one day; and in the reproducibility test (day-to-day precision), sample digestion and ICP OES analysis were studied by triplicate analyses of three samples on three days for a period of three weeks.

2.6. Multivariate data processing

Evaluation of the geographical origin of fruits can be facilitated using multivariate approach. Different chemometrics techniques were applied in order to visualize the data structure and to classify the samples according to their origin and to detect corresponding markers, namely principal component analysis (PCA), linear discriminant analysis (LDA), k-nearest neighbors (kNN), partial least square-discriminant analysis (PLS-DA) and support vector machine (SVM). All basic statistic and multivariate analysis were carried out

using R software version 3.2.0 (R Core Team, 2014) with caret package (Kuhn, 2008).

3. Results and discussion

3.1. Method validation

Although the analytical procedure we adopted is well established, it is important to validate the entire procedure. The accuracy of the procedure was made by analyzing a SRM (NIST 1570a) Spinach Leaves, for Al, As, Ca, Cd, Co, Cu, K, Mn, Na, Sr,

Table 1
Limit of detection (LOD) and quantification (LOQ) of elements achieved with ICP-OES, average precision for repeatability of solutions (RSD) and recoveries for spiked *S. edule* fruit samples ($n = 5$) digested by microwave-assisted wet digestion.

	LOD ^b (mg/kg)	LOQ ^b (mg/kg)	RSD (%)	Recovery (%)
Al	0.5	1.6	5.4	98
As	0.1	0.30	0.9	100
Ba	0.02	0.08	4.4	103
Ca ^a	1.4	4.6	1.5	96
Cd	0.03	0.10	2.7	95
Co	0.02	0.05	4.2	102
Cu	0.02	0.05	5.3	105
Fe	0.5	1.5	4.3	99
K ^a	2	7	6.2	98
Mg ^a	0.5	1.7	2.5	100
Mn	0.3	0.90	1.6	102
Na ^a	0.1	0.50	3.3	104
Pb	0.16	0.55	3.4	102
Sr	0.03	0.09	1.8	95
Zn	0.03	0.10	2.9	98

^a Major elements were spiked at 5.0 mg/kg, others were spiked at 0.5 mg/kg.
^b LOD and LOQ are referred to dry sample taking into account the sample mass 0.5 g and dilution carried out for measurement.

Table 2
Major and trace element concentrations of *S. edule* samples from Corrientes province (Argentina).

Element [mg/kg]		Corrientes $n = 19$	Bella Vista $n = 17$	Monte Caseros $n = 30$	Santo Tomé $n = 26$
Al	Mean	3.7	3.7	5.0	5.5
	Range	2.0–6.2	2.5–5.0	2.0–7.8	2.2–8.5
As	Mean	<LOQ	<LOQ	<LOQ	<LOQ
	Range	<LOQ	<LOQ	<LOQ	<LOQ
Ba	Mean	3.2	2.3	6.7	5.4
	Range	1.2–6.0	1.0–5.2	3.0–7.2	1.2–10.5
Ca	Mean	193	168	162	187
	Range	183–201	144–192	139–183	165–210
Cd	Mean	<LOQ	<LOQ	<LOQ	<LOQ
	Range	<LOQ	<LOQ	<LOQ	<LOQ
Co	Mean	0.25	0.22	0.25	0.25
	Range	0.22–0.50	0.20–0.55	0.20–0.58	0.20–0.50
Cu	Mean	0.37	0.40	0.35	0.44
	Range	0.20–0.66	0.32–0.60	0.25–0.68	0.20–0.72
Fe	Mean	2.9	3.0	4.7	4.9
	Range	2.2–4.2	2.3–4.1	2.7–6.9	2.8–6.6
K	Mean	1860	1850	1640	1690
	Range	1670–2000	1620–2120	1345–2075	1350–2190
Mg	Mean	202	175	165	198
	Range	156–236	125–208	121–225	112–223
Mn	Mean	0.7	0.6	0.4	0.4
	Range	0.4–1.0	0.2–1.2	0.2–0.5	0.2–0.8
Na	Mean	10.7	10.9	8.4	7.5
	Range	9.0–12.0	9.5–12.0	5.7–10.5	5.7–9.5
Pb	Mean	<LOQ	<LOQ	<LOQ	<LOQ
	Range	<LOQ	<LOQ	<LOQ	<LOQ
Sr	Mean	2.4	2.0	3.2	3.8
	Range	2.0–3.5	1.2–3.2	2.5–5.0	3.0–4.7
Zn	Mean	4.4	3.7	6.6	5.3
	Range	2.5–6.5	2.5–5.8	3.0–9.8	2.5–8.1

and Zn. Good agreement was achieved between the data obtained by the present method and certified values. The mean recoveries of elements ranged from 97% to 105% and standard deviations were less than 10%. In addition, the method was also validated for Fe determination with satisfactory results (97.5%).

On the other hand, spike tests were also executed to evaluate the accuracy of the proposed method. The recoveries, depicted in Table 1, were in the range of 95%–105% with relative standard deviation (RSDs) lower than 6.5% in all cases ($n = 5$). The obtained recoveries confirmed that no significant elemental losses occurred during the digestion process.

The capability of the method to detect the studied elements was estimated through the determination of the LODs and LOQs experimentally (Table 1). As can be seen, all LODs and LOQs were sufficiently below the typical levels of interest for the studied samples.

Finally, inter-day and intra-day repeatability data were obtained using the described procedure for all the analyzed elements. The inter-day data were generated on different days using new instrument settings and new calibrations curves. The relative standard deviations (RSD) were lower than 10.3 and 6.8% for all the elements for inter- and intra-day data, respectively.

3.2. *S. edule* fruits mineral contents

Table 2 summarizes the results obtained for mineral concentrations determined in *S. edule* fruit samples, according to their geographical origin. Each sample was independently microwave digested, and analyzed in triplicate. The levels of potentially toxic trace elements (Al, As, Cd, and Pb) were very low or even below the LOQ in all *S. edule* samples, except Al, which was measured from 2.0 mg/kg up to 8.5 mg/kg.

From the results, it is clear that K is the most abundant element in all *S. edule* fruit samples, followed by Mg, Ca, and Na as the

second, third and fourth major minerals, respectively. In general, fruits are considered as a good source of these major elements, especially K. In line with most of the previous studies, our results were at similar order than K, Ca, Mg, and Na levels determined in *S. edule* fruits collected from Colombian Andes (Leterme, Buldgen, Estrada, & Londoño, 2006) and Mexico (USDA (U. S. Department of Agriculture), 2014). On the other hand, in comparison with other commonly consumed fruits, the major element concentrations (K, Ca and Mg) in *S. edule* edible part (flesh) were lower than those determined in bananas from Tenerife Spain (3500–7500 mg/kg for K; 180–200 mg/kg for Ca; 320–680 mg/kg for Mg) (Hardisson et al., 2001), or apples from Croatia (3850–4590 mg/kg for K; 210–540 mg/kg for Ca, 1520–4190 mg/kg for Mg) (Juranić Cindrić et al., 2012). The Na concentrations in *S. edule* fruits were at levels similar to other four cucurbit fruits (the same plant family of *S. edule*), such as: caigua, pumpkin, cucumber and watermelon (Oliveira et al., 2014).

The concentrations of trace elements in all *S. edule* samples were in general at similar order of magnitude than the observed by other authors in these fruits (Leterme et al., 2006; Modgil, Modgil, & Kumar, 2004). Iron and Zn were the most abundant elements relatively to Mn and Cu contents. However, Fe and Zn contents were significantly lower than those reported in other fruits, such as banana, apples, citrus, quince or pear fruits (Hardisson et al., 2001; Juranić Cindrić et al., 2012; Özcan, Harmanakaya, & Gezgin, 2012).

3.3. Differentiation among *S. edule* fruits of different geographical origin

To verify the significance of the different geographical origin of samples, the results were evaluated by using multivariate tools. As preliminary stage, prior to classification modelling, PCA was carried out for exploratory analysis. This technique was used to visualize the distribution of samples in a reduced-dimensional space, in order to explore the presence of groups, outliers or trends in the data. PCA is a multivariate technique that allows an easy visualization of all the information contained in the data set, retaining most information of the original data set. Prior to the multivariate processing, the data matrix was standardized (autoscaled) because the element concentrations were at different orders of magnitude.

In addition, the data matrix was also tested by applying Grubb's test (Grubbs, 1969) detecting no outliers.

The results for elemental concentrations in 92 samples were arranged as a whole data matrix (92 × 12) using the element concentration as column and the samples as rows. The potentially toxic trace elements As, Cd and Pb, were not considered for multivariate analysis because of their non-detectable levels in all samples. At first, PCA was carried out on the whole data matrix, three principal components with eigenvalues exceeding one were extracted, which explains up to 64.3% of the total variance. The loadings of the original variables on the first two principal components (PCs) and the variance explained by each PC are shown in Fig. 1a.

As can be seen, the first principal component was strongly associated with the values of Ba, Fe, Mn, Na and Sr. On the other hand, Mg, Ca and Na were the dominant variables in the second principal component. The third principal component incorporates information from the remaining less abundant elements.

The score plot of two first PCs (Fig. 1b) retained 44.8% of the variability of the system. Despite the low explained variability retained in the two first PCs, this explorative analysis stage is crucial to check if the analytical technique is able to collect information on the geographical origin of samples (Li Vigni, Durante, & Cocchi, 2013). The results given in Fig. 1b showed a tendency of groupings between samples of the same origin. This figure clearly shows the systematic separation of samples in two principal groups according to PC1, the first group consisting of MC and ST samples, and the second group of Bv and CT samples. The MC and ST samples showed negative scores on PC1 indicating higher concentrations of Ba, Sr and Fe; furthermore, Bv and CT samples showed positive scores on PC1 indicating higher concentrations of K, Mn and Na. Then, the addition of the information resumed on PC2, allowed us to resolve the score projection in four groups corresponding to each selected site of origin with some overlapping among groups. Samples with negative scores on PC2 correspond to high concentrations of Ca and Mg; and positive scores to high concentrations of Na.

In summary, good separation among the samples of four regions was observed, inferring that they had inherent compositional differences and classifying samples from different geographical origins was feasible.

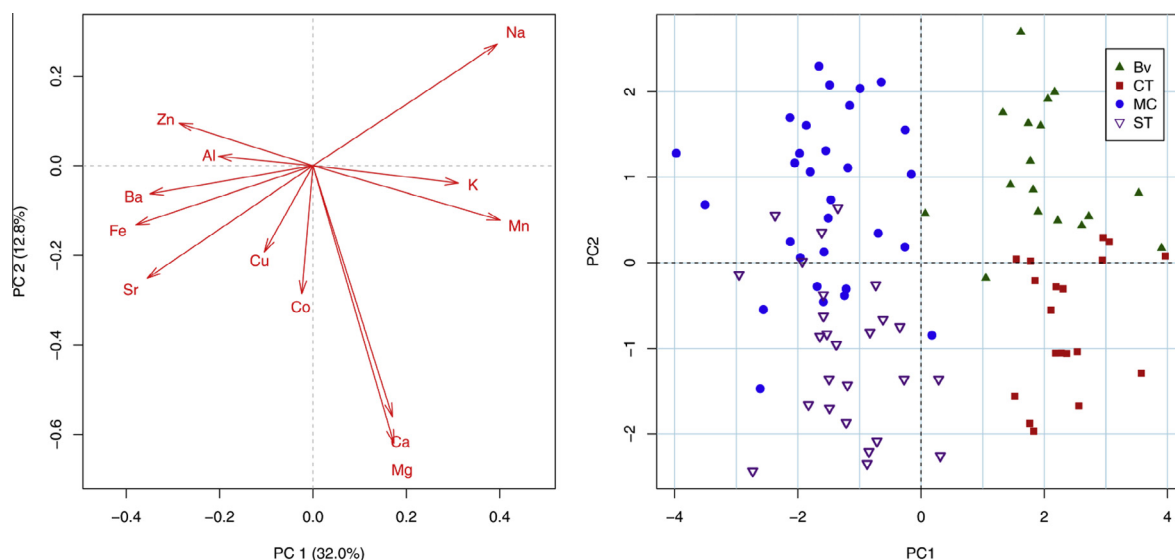


Fig. 1. Plot of the first principal component (PC1) versus the second principal component (PC2) results: (a) the 2D loading plot, and (b) corresponding score plot for the 92 *S. edule* fruit samples identified according to geographical origin.

PCA showed a natural grouping of the samples according to their geographical origin. As a result, we consider that the multivariate classification analysis can be performed by using as factor the geographical origin of samples. The goal of supervised pattern recognition techniques is to create classification rules. Training set with known class memberships is used to calculate a classifier. A test set, containing objects not included in the training and also with known class memberships, serves to validate the model built. In this work, we selected a training data set randomly (75% of the objects of the whole data matrix) and used the derived classification rule to predict the group membership of the remaining data. The random sampling was done according to the geographical origin of samples in order to balance the class distributions within the splits (stratified sampling). Since the results obtained in each split of whole matrix depend on the choice of the test data, the procedure was repeated 100 times. This procedure gave a distribution for the misclassification rate, allowing comparison between the different discriminatory methods tested. LDA, PLS-DA, kNN and SVM were used to construct models to classify *S. edule* fruits according to its geographical origin.

In this work, the models for discrimination of *S. edule* fruits according to their geographical origin were constructed by using different supervised pattern recognition techniques. The selected chemometric methods were as follow:

Linear discriminant analysis (LDA) is a linear classification tool used to differentiate groups of samples as a function of one or several combinations of experimental variables (Bevilacqua et al., 2013; Lavine & Rayens, 2009). Discriminant functions (DFs) are linear combinations of original variables. This method maximizes the variance between groups and minimizes the variance within categories.

Partial least square discriminant analysis (PLS-DA) is a linear classification method that combines the properties of partial least squares regression with the discrimination power of a classification technique (Ballabio & Consonni, 2013). The major advantage in PLS-DA comes from its flexibility, especially when coping with situations, unforeseen in traditional statistics, where the number of variables far outnumbers the samples (Breton & Lloyd, 2014). Within caret package, the number of significant PLS components (model dimensionality) is determined by ten-fold cross-validation.

K-nearest neighbor analysis (kNN) is a distance based non-parametric approach. This method assigns an object into the class most common among its k-nearest neighbors in terms of distance. This method is easy to implement and is computationally fast. The optimal size of neighbor k is estimated by a cross-validation procedure (Bevilacqua et al., 2013).

Support vector machines (SVM) classify the data by constructing a separate hyperplane in n-dimensional space, which maxi-

mizes the margin between classes. A detailed description of the theory of SVM can be referred in several excellent books and tutorials (Kivinen, Smola, & Williamson, 2004; Steinwart & Christmann, 2008). The SVM method has a number of interesting characteristics, including an effective prevention of overfitting, which improves its ability to build models using large numbers of variables with relatively few experimental samples in the training set.

PLS-DA, kNN and SVM need to optimize several parameters in such a way that a suitable number of parameters are selected to build the model. In this work, the choices of number of latent variables (LVs) for PLS-DA; number of neighbor k for kNN; penalty factor C, ϵ of the ϵ -insensitive loss function and kernel type for SVM, were calculated by using ten-fold cross-validation technique by which maximum accuracy is selected. In addition, we ran the SVM with four types of kernel function: linear, radial basis, polynomial and sigmoid. The best results were obtained when the radial kernel function was used for SVM.

Once selected the optimal values for each model, the sensitivity (samples belonging to the class and classified correctly in this class), specificity (samples not belonging to the modeled class and correctly classified as not belonging), and the mean accuracy rate were considered for evaluation of the classification achieved with the multivariate methods (Marcelo, Martins, Pozebon, Dressler, & Ferrão, 2014). Table 3 summarizes the results obtained after the application of the different classification models.

Through the above analysis, we can see that the four models displayed different degrees of success. Of the four techniques, the performance of the LDA was the best, with a success rate of 89.1% in the test set and better sensitivity for Bv, CT and MC groups. The order of successful identification rates was as follows: LDA > SVM > PLS-DA > kNN (Table 3). Moreover, the search time spending on LDA model was shorter. So it was an ideal model for discriminating different geographical origins of samples studied.

Then, in order to evaluate the performance of LDA technique, LDA algorithm was applied to the entirely data set, and their accuracy was analyzed. The objective of this process is not to lose information, although the system obtained is more complex.

Two DFs whose Eigenvalues were greater than 1 were obtained from the data. The first two canonical discriminant functions explain 98.7% of the variance. Fig. 2 shows the distribution patterns of all *Sechium* fruit samples according to their geographical origins in the plot defined by the first two DFs. This figure showed a good discrimination between the two principal groups formed by samples from CT + Bv (negative scores on DF1), and ST + MC (positive scores on DF1). This spatial distribution of samples in the score plot corresponding to the first two DFs, is compatible with the short distance (in terms of longitude) and similar geo-climatic characteristics among the studied areas.

Table 3
The discrimination results of different models for the test sets.

Groups	Number of samples		LDA		PLS-DA (VLs = 2) ^a		k-NN (k = 5) ^b		SVM (C = 1; ϵ = 0.053) ^c	
	Training set	Test set	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
Bv	13	4	100	100	75	100	50	100	75	100
CT	14	5	100	100	100	100	100	88	100	94
MC	22	8	100	93	86	86	71	100	100	100
ST	19	7	83	100	83	93	100	87	100	100
<i>Mean accuracy (%)</i>			89.1		84.7		82.0		87.0	

^a VLs: number of latent variables.

^b k: number of k neighbors.

^c C: penalty factor; ϵ : ϵ -insensitive loss function.

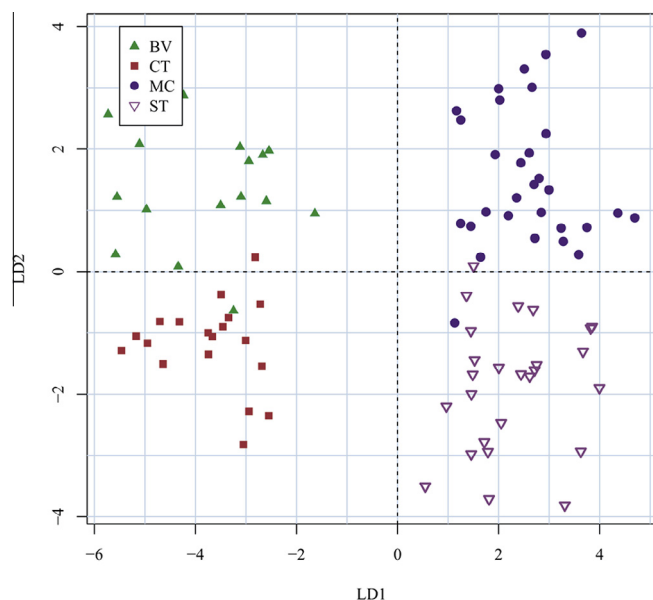


Fig. 2. Linear discriminant analysis of *S. edule* samples according to their geographical origin. Scatter plot representing the projection of the point of all *S. edule* samples and of each geographical origin separately on the plane formed by the first two discriminant functions.

Table 4

Mean daily intake and contribution to the daily requirements of essential elements for adult men and women from one serving of *S. edule* fruit (264 g/person/day).

Element		DRI (mg/day)	Intake (mg/day)	Contribution to the DRI (%)
<i>Major elements</i>				
Ca		1000 ^a	46.8	4.7
Mg		320–400 ^a	48.8	13.5–12.2
K		4700 ^b	465	9.8
Na		1500 ^b	2.4	0.16
<i>Trace elements</i>				
Cu	Male	0.9 ^a	0.10	11.4
	Female	0.9 ^a		
Fe	Male	8 ^a	1.02	12.8
	Female	18 ^a		
Mn	Male	2.3 ^b	0.14	6.1
	Female	1.8 ^b		
Zn	Male	11 ^a	1.32	12.0
	Female	8 ^a		

^a Recommended dietary allowances (RDAs).

^b Acceptable intakes (AIs).

It was found that the average prediction accuracies for all 92 samples were 88.2% for Bv samples, 100.0% for CT samples, 96.7% for MC samples and 96.2% for ST samples. As such, this method is sufficiently accurate to employ this technique as a screening method for discrimination of *S. edule* fruits according to their geographical origin.

In the final step, we looked to select a reduced number of chemical variables necessary to achieve good discrimination of fruit samples according to their geographical origin. Thus, we reduced or eliminated irrelevant information, selecting those parameters which are important for geographical classification by forward stepwise LDA (S-LDA). After application of the S-LDA to the data, most of the samples (93.5% and 90.7% after cross-validation) were correctly classified selecting the following variables: Ca, Ba, Cu, Mn, Na, Sr, and Zn. These results indicate that the major and trace element content in *S. edule* fruit samples make it possible to classify them according to the geographical origin reasonably well

(above 90%), by using only seven major and trace element concentrations.

3.4. Mineral contribution by *S. edule* fruits

Table 4 summarizes the average daily essential element intakes based on a 264 g/person/day *S. edule* consumption and their contribution to the dietary reference intakes (DRI) for adults (Institute of Medicine & Food and Nutrition Board, 2006). In general, the contribution to the essential mineral intakes for the consumption of one serving of two *S. edule* fruit a day may fulfill approximately 9–16% of the DRIs of several essential elements (Cu, Fe, K, Mg, and Zn). However, it is important to emphasize that the intake of Fe for adult women does not exceed 6.2% of DRI established.

In addition, the intake of one serving of *S. edule* fruits would not contribute significantly to dietary Na intake. Sodium is essential to humans but in excess dietary Na can lead to hypertension, the primary risk factor for cardiovascular disease. Even at the upper end of the range, the Na content of the samples studied was very low as expected for fruit contents.

4. Conclusions

In this paper, 15 major and trace elements were determined in samples of *S. edule* fruits by using ICP OES after microwave digestion. The multivariate analysis of the major and trace element data demonstrated that there is an impact of geographical origin on element concentrations in *S. edule* fruit. The variations in certain element concentrations (Ca, Ba, Cu, Mn, Na, Sr, and Zn) may serve as a classifier for determining geographical origin. A higher number of samples should be analyzed to have robust classification models, but multivariate analysis performed in the present work suggests a great potentiality of mineral composition for geographical traceability of *S. edule* fruits. Finally, the overall concentrations of minor and essential trace elements for *S. edule* fruits were found to have good nutritional contribution in accordance to RDA. The contribution of the toxic trace elements Al, As, Cd and Pb to the overall intake from the analyzed fruits was very low and could not pose any threat to the consuming population.

Acknowledgements

This research was supported by Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) and Secretaría General de Ciencia y Técnica of UNNE, Argentina.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.foodchem.2016.04.120>.

References

- Aung, L., Ball, A., & Kushad, M. (1990). Developmental and nutritional aspects of chayote (*Sechium edule*, Cucurbitaceae). *Economic Botany*, 44(2), 157–164.
- Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: Linear models. PLS-DA. *Analytical Methods*, 5(16), 3790.
- Barbosa, R. M., Batista, B. L., Varrigue, R. M., Coelho, V. A., Campiglia, A. D., & Barbosa, F. (2014). The use of advanced chemometric techniques and trace element levels for controlling the authenticity of organic coffee. *Food Research International*, 61, 246–251.
- Bevilacqua, M., Bucci, R., Magri, A. D., Magri, A. L., Nescatelli, R., & Marini, F. (2013). Chapter 5 – Classification and class-modelling. In F. Marini (Ed.), *Data handling in science and technology* (Vol. 28, pp. 171–233). Oxford: Elsevier.
- Brereton, R. G., & Lloyd, G. R. (2014). Partial least squares discriminant analysis: Taking the magic away. *Journal of Chemometrics*, 28(4), 213–225.
- Drivelos, S. A., Higgins, K., Kalivas, J. H., Haroutounian, S. A., & Georgiou, C. A. (2014). Data fusion for food authentication. Combining rare earth elements and trace

- metal to discriminate "Fava Santorinis" from other yellow split peas using chemometric tools. *Food Chemistry*.
- Gordon, E. A., Guppy, L. J., & Nelson, M. (2000). The antihypertensive effects of the Jamaican Cho-Cho (*Sechium edule*). *West Indian Medical Journal*, 49(1), 27–31.
- Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics*, 11(1), 1–21.
- Hardisson, A., Rubio, C., Baez, A., Martin, M., Alvarez, R., & Diaz, E. (2001). Mineral composition of the banana (*Musa acuminata*) from the island of Tenerife. *Food Chemistry*, 73(2), 153–161.
- Institute of Medicine & Food and Nutrition Board (2006). *Dietary reference intakes: The essential guide to nutrient requirements*. Washington, DC: The National Academies Press.
- Juranović Cindrić, I., Krizman, I., Zeiner, M., Kampić, Š., Medunić, G., & Stinger, G. (2012). ICP-AES determination of minor- and major elements in apples after microwave assisted digestion. *Food Chemistry*, 135(4), 2675–2680.
- Kivinen, J., Smola, A. J., & Williamson, R. C. (2004). Online learning with kernels. *Signal Processing, IEEE Transactions on*, 52(8), 2165–2176.
- Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, 28(5), 1–26.
- Larrea-Marín, M. T., Pomares-Alfonso, M. S., Gómez-Juaristi, M., Sánchez-Muniz, F. J., & de la Rocha, S. R. (2010). Validation of an ICP-OES method for macro and trace element determination in Laminaria and Porphyra seaweeds from four different countries. *Journal of Food Composition and Analysis*, 23(8), 814–820.
- Lavine, B. K., & Rayens, W. S. (2009). 3.16 – Statistical discriminant analysis. In S. D. B. T. Walczak (Ed.), *Comprehensive chemometrics* (pp. 517–540). Oxford: Elsevier.
- Leterme, P., Buldgen, A., Estrada, F., & Londoño, A. M. (2006). Mineral content of tropical fruits and unconventional foods of the Andes and the rain forest of Colombia. *Food Chemistry*, 95(4), 644–652.
- Li, C., Yang, S.-C., Guo, Q.-S., Zheng, K.-Y., Shi, Y.-F., Xiao, X.-F., & Long, G.-Q. (2014). Determining the geographical origin of the medicinal plant *Marsdenia tenacissima* with multi-element analysis and data mining techniques. *Chemometrics and Intelligent Laboratory Systems*, 136, 115–120.
- Li Vigni, M., Durante, C., & Cocchi, M. (2013). Chapter 3 – Exploratory data analysis. In F. Marini (Ed.), *Data handling in science and technology* (Vol. 28, pp. 55–126). Amsterdam: Elsevier.
- Marcelo, M. C. A., Martins, C. A., Pozebon, D., Dressler, V. L., & Ferrão, M. F. (2014). Classification of yerba mate (*Ilex paraguariensis*) according to the country of origin based on element concentrations. *Microchemical Journal*, 117, 164–171.
- Mir-Marqués, A., Domingo, A., Cervera, M. L., & de la Guardia, M. (2015). Mineral profile of kaki fruits (*Diospyros kaki* L.). *Food Chemistry*, 172, 291–297.
- Modgil, M., Modgil, R., & Kumar, R. (2004). Carbohydrate and mineral content of chayote (*Sechium edule*) and bottle gourd (*Lagenaria Siceraria*). *Journal of Human Ecology*, 15(2), 157–159.
- Oliveira, A. C., dos Santos, V. S., dos Santos, D. C., Carvalho, R. D. S., Souza, A. S., & Ferreira, S. L. C. (2014). Determination of the mineral composition of Caigua (*Cyclanthera pedata*) and evaluation using multivariate analysis. *Food Chemistry*, 152, 619–623.
- Ordoñez, A. A. L., Gomez, J. D., Vattuone, M. A., & Isla, M. I. (2006). Antioxidant activities of *Sechium edule* (Jacq.) Swartz extracts. *Food Chemistry*, 97(3), 452–458.
- Özcan, M., Harmankaya, M., & Gezgin, S. (2012). Mineral and heavy metal contents of the outer and inner tissues of commonly used fruits. *Environmental Monitoring and Assessment*, 184(1), 313–320.
- R Core Team (2014). *The R project for statistical computing. R: A language and environment for statistical computing* (vol. 2015).
- Rao, K. S., Dominic, R., Singh, K., Kaluwin, C., Rivett, D. E., & Jones, G. P. (1990). Lipid, fatty acid, amino acid, and mineral compositions of five edible plant leaves. *Journal of Agricultural and Food Chemistry*, 38(12), 2137–2139.
- Saade, R. L. (1996). *Chayote. Sechium edule* (Jacq.) Sw. Promoting the conservation and use of underutilized and neglected crops. 8. Rome, Italy/Gatersleben: Institute of Plant Genetics and Crop Plant Research/International Plant Genetic Resource Institute.
- Steinwart, I., & Christmann, A. (2008). *Support vector machines*. Berlin: Springer Science & Business Media.
- Szymczycha-Madeja, A., & Welna, M. (2013). Evaluation of a simple and fast method for the multi-elemental analysis in commercial fruit juice samples using atomic emission spectrometry. *Food Chemistry*, 141, 3466–3472.
- USDA (U. S. Department of Agriculture) (2014). Basic Report: 11149, Chayote, fruit, raw. *National Nutrient Database for Standard Reference. Release 27* (vol. 2015).