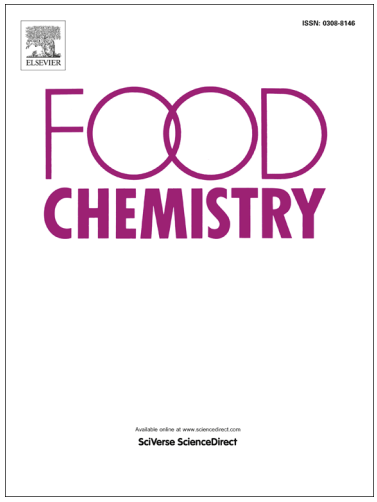


Intra-regional classification of grape seeds produced in Mendoza province (Argentina) by multi-elemental analysis and chemometrics tools

Brenda V. Canizo, Leticia B. Escudero, María B. Pérez, Roberto G. Pellerano, Rodolfo G. Wuilloud

PII: S0308-8146(17)31517-0
DOI: http://dx.doi.org/10.1016/j.foodchem.2017.09.062
Reference: FOCH 21731

To appear in: *Food Chemistry*

Received Date: 24 February 2017
Revised Date: 8 September 2017
Accepted Date: 12 September 2017

Please cite this article as: Canizo, B.V., Escudero, L.B., Pérez, M.B., Pellerano, R.G., Wuilloud, R.G., Intra-regional classification of grape seeds produced in Mendoza province (Argentina) by multi-elemental analysis and chemometrics tools, *Food Chemistry* (2017), doi: http://dx.doi.org/10.1016/j.foodchem.2017.09.062

Intra-regional classification of grape seeds produced in Mendoza province (Argentina) by multi-elemental analysis and chemometrics tools

Brenda V. Canizo[a], Leticia B. Escudero[a], María B. Pérez[a], Roberto G. Pellerano[b] and Rodolfo G. Wuilloud[a,*]

[a]*Laboratorio de Química Analítica para Investigación y Desarrollo (QUIANID), Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Cuyo, Instituto Interdisciplinario de Ciencias Básicas (ICB), UNCUYO-CONICET, Padre J. Contreras 1300, (5500) Mendoza, Argentina.*

[b]*Instituto de Química Básica y Aplicada del Nordeste Argentino (IQUIBA-NEA), CONICET, Facultad de Ciencias Exactas y Naturales y Agrimensura, Universidad Nacional del Nordeste (UNNE), Av. Libertad 5470, (3400) Corrientes, Argentina.*

*Corresponding author. Tel: +54-261-4259738

E-mail address: rwuilloud@mendoza-conicet.gob.ar; rodolfowuilloud@gmail.com (R.G. Wuilloud)

Abstract

The feasibility of the application of chemometric techniques associated with multi-element analysis for the classification of grape seeds according to their provenance vineyard soil was investigated. Grape seed samples from different localities of Mendoza province (Argentina) were evaluated. Inductively coupled plasma mass spectrometry (ICP-MS) was used for the determination of twenty-nine elements (Ag, As, Ce, Co, Cs, Cu, Eu, Fe, Ga, Gd, La, Lu, Mn, Mo, Nb, Nd, Ni, Pr, Rb, Sm, Te, Ti, Tl, Tm, U, V, Y, Zn and Zr). Once the analytical data were collected, supervised pattern recognition techniques such as linear discriminant analysis (LDA), partial least square discriminant analysis (PLS-DA), k-nearest neighbors (k-NN), support vector machine (SVM) and random forest (RF) were applied to construct classification/discrimination rules. The results indicated that nonlinear methods, RF and SVM, perform best with up to 98% and 93% accuracy rate, respectively and therefore are excellent tools for classification of grapes.

Keywords: Grape seeds; Multi-elemental analysis; Multivariate classification; Chemometrics; ICP-MS.

1. Introduction

The identification of the geographical origin of wine grapes (*Vitis vinifera L.*) is of remarkably important in today's globalized trade, because it is directly related with wine provenance and wine-making practices belonging to a specific region. Likewise, the knowledge of grape origin is highly valuable to winemakers in order to assure the quality of the wines being produced. In South America, Argentina is one of the main producers and exports high quality wines to different countries around the world. Among several regions of Argentina, Mendoza province is the most important grape and wine producing region, representing more than 75% of national production (Castex, Tejeda, & Beniston, 2015; International Organization of Vine and Wine (OIV), 2016). The natural conditions of Mendoza province, such as altitude, arid and rocky soils, and differentiated thermal amplitude between day and night, are ideal for the successful growing of several varieties of high quality grapes (white, red, and rose cultivars) (Fundación ProMendoza, 2016). However, besides high quality, ensuring the geographical origin of grapes is considered an important aspect by consumers, and hence, of great strategic value to the economies of the regions producing grapes and wines. Within this framework, the traceability of grapes and wines produced in Argentina is being intensively controlled in recent years to certify their origin.

Recently, regional classification using chemometrics has been widely explored for wines from different parts of the world. Among the several factors that are strongly correlated and affect wine composition, mention can be made of the variety or varieties of grapes, the location where grapes are grown, the seasonal weather conditions, agricultural practices in vineyards, and the techniques used by winemakers, among

others. However, the most influential variable on wine flavor is the grape variety due to differential chemical composition (Reynolds, 2010; van Leeuwen, 2010). Grapes show a high complex matrix composed of water, sugars, organic acids, phenolic compounds, vitamins, and minerals (Jackson, 2014). Moreover, it is well known that grapes are used for wine production and play an important role in the quality of this alcoholic beverage. In general, the elemental composition of wines reflects the geochemistry of the producing area, chemical composition of raw materials, as well as manufacture factors such as agricultural practices, or winemaking processes (Versari, Laurie, Ricci, Laghi, & Parpinello, 2014). Inorganic elements have several advantages as chemical markers because they are not metabolized and remain unchanged or transformed during enological processes (Pohl, 2007; Saurina, 2010). Considering these aspects, the study of mineral profiles along with chemometrics tools could be used for classification of grapes. In fact, it could be a powerful strategy to discriminate wines and other grape-products coming from different geographical origins. Furthermore, trace elements in grapes are mainly located in their seeds (Fabani, Toro, Vázquez, Díaz, & Wunderlin, 2009; Rogiers, Greer, Hatfield, Orchard, & Keller, 2006), which is also very important because they are major end-products resulting from wine industry, accounting for 15% of the solid waste (Lachman et al., 2013; Mironeasa, Leahu, & Codin, 2010).

Fingerprinting techniques combine chemical analysis with multivariate statistical analysis or advanced data mining methods to solve complex problems that require multidisciplinary approaches (Galgano, Favati, Caruso, Scarpa, & Palma, 2008; González-Centeno et al., 2010; Mutihac & Mutihac, 2008). In general, to generate reliable data on mineral contents of different kind of samples, elemental mass spectrometry is used. Currently, ICP-MS is the most used technique due to its high sensitivity and robustness for reliable multi-elemental determinations (Tanner

& Günther, 2009). However, although several works have reported the application of ICP-MS along with chemometrics to discriminate wines, the classification of grapes according to their geographical origin has not been extensively studied (Cugnetto et al., 2014; Ferrer-Gallego, Hernández-Hierro, Rivas-Gonzalo, & Escribano-Bailón, 2013; Versari et al., 2014). Moreover, to the knowledge of the authors, there are no studies where grape seeds have been used for fingerprinting grapes. Most probably, due to the fact that normally the entire grape berry is analyzed, without establishing clearly which part is more suitable to analyze in order to develop the chemometrics classification.

Accordingly, the purpose of the current work was to assess the potential use of several multivariate statistics tools combined with multi-element data obtained by ICP-MS analysis of grape seeds as an original source for the differentiation of grapes according to their geographical origin. The potential of this strategy was tested to discriminate grapes cultivated in nearby locations within Mendoza province. Thus, this work is different from others reporting the classification of grapes sourced from regions separated by large distances, where significant differences in soil composition might be expected and had a remarkable impact on chemical composition of grapes and their easy classification. Finally, this is the first work reporting the classification of grapes grown in Mendoza province according to their geographical origin.

2. Materials and methods

2.1. Instrumentation

Analytical measurements were performed using an inductively coupled plasma mass spectrometer, PerkinElmer-SCIEX, ELAN DRC-e model (Thornhill, Canada) fitted with an HF-resistant and high performance perfluoracetate (PFA) nebulizer, coupled to a baffled quartz-made cyclonic spray chamber, cooled with the PC^3 system from ESI (USA). Nickel sampler and skimmer cones were used.

The instrumental conditions were as follows: auto lens mode on, peak hopping measure mode, dwell time of 50 ms, 10 sweeps per readings, 1 reading per replicate and each analysis consisted of 3 replicates. After optimization of the instrument, the following conditions were chosen for all determinations: 1000 W RF power, sample introduction flow of 0.8 mL min^{-1}, nebulizer gas flow rate of 0.85 L min^{-1}. The instrument was calibrated against external certified standard solutions. To correct non-spectral interferences, Rh was used as internal standard for all determinations. The isotopes detected, in order of mass number, were as follows: ^{47}Ti, ^{51}V, ^{55}Mn, ^{57}Fe, ^{59}Co, ^{60}Ni, ^{63}Cu, ^{66}Zn, ^{69}Ga, ^{75}As, ^{85}Rb, ^{89}Y, ^{90}Zr, ^{93}Nb, ^{98}Mo, ^{107}Ag, ^{130}Te, ^{133}Cs, ^{139}La, ^{140}Ce, ^{141}Pr, ^{142}Nd, ^{152}Sm, ^{153}Eu, ^{158}Gd, ^{169}Tm, ^{175}Lu, ^{205}Tl, and ^{238}U.

2.2. Reagents and standards

All the reagents used were of analytical grade. Ultrapure water (18 MΩ·cm) obtained from a Milli-Q water purification system (Millipore, Paris, France) was used in the preparation of all solutions. Ultrapure concentrated nitric acid (65% v/v) purchased from Merck (Darmstadt, Germany) and hydrogen peroxide (30% v/v) from Biopack (Argentina) were used throughout. All the glassware was washed in 0.5 mol L^{-1} HNO_3 solution for 24 h and later rinsed with ultrapure water before use. Argon (99.996%) from Air Liquide (Córdoba, Argentina) was used for ICP-MS

determinations. Certified multi-element standard solutions 2, 3 and 5, and rhodium (^{103}Rh) mono-elemental standard solution from Perkin Elmer Pure Plus Atomic Spectroscopy Standards, (Norwalk, USA) were used.

2.3. Sample collection and analytical procedure

Vitis vinifera L. grapes were collected from vineyards located in five winemaking regions of Mendoza (Argentina): Rivadavia (33º11´S, 68º28´W), San Martín (33º04´S, 68º19´W), Guaymallén (32º54´S, 68º47´W), Junín (33º15´S, 68º43´W) and Maipú (32º58´S, 68º46´W), during the 2011 growing season. Sampling locations are shown in Fig. 1. The following grapevine varieties were collected: Cabernet Sauvignon, Malbec, Bonarda, Aspirant Bouchet (as red cultivars), Chardonnay, Sauvignon Blanc, and Pedro Jiménez (as white cultivars).

In order to obtain representative samples, individual bunches were randomly collected from each specific vineyards plot, obtaining a total of 408 samples. The grape bunches were manually harvested at their optimum maturity stage with an appropriate pair of scissors, cutting them carefully at the point of insertion of peduncle with vine shoot. Subsequently, the samples were refrigerated and carried immediately to the laboratory for analysis.

The grapes collected for multi-elemental determination were first washed with tap water and then rinsed with Milli-Q water. Grape seeds were separated from the rest of the whole bunches using plastic tweezers, and then washed with Milli-Q water to remove pulp residues. Then, seeds were lyophilized and finally pulverized with a mill.

Lyophilized seed samples were mineralized following the next procedure: 0.5 g were weighted and added with 6 mL of concentrated HNO_3 and 2 mL of concentrated H_2O_2. This mixture was left at room temperature for 12 hours. Afterwards the following heat treatment was applied using a heating plate: 60 min at 50°C, 90 min at 100°C, and finally 90 min at 150°C. The digested samples were left to cool until room temperature was reached and then transferred into a volumetric flasks and taken to a final volume of 25 ml with Milli-Q water. Subsequently, aliquots of 2 mL of the previous solution were taken and placed into a 50 mL flask and the volume was completed with Milli-Q water. These solutions were finally analyzed by ICP-MS.

2.4. Chemometrics techniques

All seeds samples were quantitatively analyzed by ICP-MS and characterized by twenty-nine descriptors (element concentrations). The data matrix for the chemometric treatment contained 408 rows and 29 columns. The matrix rows represented the number of samples, and the columns corresponded to the trace element concentrations. The preprocessing of the dataset in the matrix was autoscaled due to wide differences in data dimensionality. For further chemometrical processing, grape seed samples of the different producing regions were identified as follows: SM for San Martín, RV for Rivadavia, JN for Junín, GY for Guaymallén, and MP for Maipú.

Multivariate statistical methods allowed verification of the contribution of each variable to the model and their ability to discriminate the different categories. Basic chemometrical characterization of the investigated grape seed samples was made by principal component analysis (PCA), which is an unsupervised technique. It was used as an exploratory analysis to split the dataset into a smaller number of variables to

provide a simpler representation of all the information. This method is able to detect groups of observations, trends, outliers, and uncover the relationships between observations and variables (Beebe, K. R., Pell, R. J., & Seasholtz, 1998; Serafim, Pereira-Filho, & Franco, 2016).

In addition, the following chemometric strategies were used to assess different models for classification of samples in accordance with their geographical origin: linear discriminant analysis (LDA), partial least square discriminant analysis (PLS-DA), k-nearest neighbor (K-NN), support vector machine (SVM), and random forest (RF).

LDA is a supervised learning method in which the model is constructed using the data of the pre-categorized objects into known categories (training data set), for finding linear combination to discriminated classes of objects (Massart et al., 1997). PLS-DA is a linear classification tool that calculates predictive models based on partial least squares regression algorithm that searches for latent variables with maximum covariance and represent the relevant sources of data variability with linear combinations of the original variables (Ballabio & Consonni, 2013; Barker & Rayens, 2003). K-NN is a nonparametric method of classification, used to improve discriminant analysis when reliable parametric estimates of probability densities are unknown or difficult to determine and, to categorize an unknown sample based on its distance to samples already classified (Fabani et al., 2009; Serafim et al., 2016). SVM represent a nonlinear classification technique, which works with supervised learning models associated to learning algorithms that analyze data used for classification and regression analysis, producing linear boundaries between objects groups in a transformed space of the x-variables. (Spring, 2008). Finally, RF consists of a set of tree decision

predictors to classify, where a random vector is generated independently from the input vector, and each tree casts a vote to choose the class to classify an input vector (Liaw & Wiener, 2002).

All statistical procedures were carried out using the statistical R software version 3.3.0 with caret package for pattern recognition classification (R Core Team, 2016).

3. Results and discussion

3.1. Multielemental determination in grape seeds

In this work, multielemental determination was performed in grape seeds by ICP-MS technique. Table 1 shows the concentrations of the different elements in the grape seed samples under study, expressed as mean values with their corresponding standard deviation, as a function of their geographical origin.

From the analysis of the results shown in Table 1, it is clear that Fe was the most abundant element found in the seeds, followed by Mn, Zn, and Cu. The lowest Fe mean content was obtained from San Martín (37.5 µg/g) whereas the highest level of this element was found in Rivadavia (44.2 µg/g). Also, Mn content was in the range of 15.7 µg/g and 20.3 µg/g, whereas Zn and Cu contents in grape seeds were in the range of 16.5 - 18.5 µg/g and 9.0 - 13.0 µg/g, respectively. These concentrations were in good agreement with those reported by other authors in seed samples, generally at a similar order of magnitude (Lachman et al., 2013; Tangolar, Ozoğul, Tangolar, & Torun, 2009; Yang, Duan, Du, Tian, & Pan, 2010). With respect to other works on grape seeds, differences were found in the levels of these elements (Ozcan, 2010; Rogiers et

al., 2006; Spanghero, Salem, & Robinson, 2009). These minerals are considered essential and determine that grape seeds can be considered as an excellent source of bio elements, which is important because they play a key role for the normal functioning of the human body. For instance, Fe, among other functions, is linked to the production of blood cells, and Zn is essential for the immune system. Both nutrients are also considered potent antioxidants. Meanwhile Mn is an essential element for both animals and plants, and deficiencies result in severe skeletal and reproductive abnormalities in mammals (Mironeasa et al., 2010; Mitić et al., 2011).

The next elements found in this study at micrograms levels were Mo, Ni, Rb, Te and Ti, while Ag, As, Co, Cs Ga, Nb, Tl, U, V and Zr were at concentrations below the microgram range. For these elements, it has to be mentioned that there is not enough data available in the literature about their content in grape seeds to be reasonably compared with the results obtained in the present work.

On the other hand, rare earth elements (REE) determined in this work, such as Y, La, Ce, Pr, Nd, Sm, Eu, Gd, Tm and Lu were at ultra-trace levels or were not detectable in some grape seed samples, this is in accordance with the fact that seeds are not good reservoirs of this kind of elements. In fact, other authors have reported only 3% of REEs is accumulated in grape seeds (Bertoldi, Larcher, Nicolini, Bertamini, & Concheri, 2009; Yang et al., 2010).

The comparison of elemental concentrations in grape seeds coming from different geographical locations (Table 1) showed similar concentration profiles of the majority elements found in the samples. Furthermore, a slight tendency to higher concentrations of REEs, as Eu, Gd, La, Lu, Nd, Pr, Sm, Tm and Y and of the minor elements such as Ag, As, Co, Cs, Ga, Tl, V and Zr can be noted for samples of GY region.

3.2. Geographical classification of grape seeds from Mendoza province

3.2.1. Statistical exploratory analysis

Basic chemometric characterization of the investigated grape seed samples was performed by PCA. As mentioned previously, this method allows the determination of the major sources of variability in the data set, and it can display a natural grouping of the studied objects (grape seed samples) in the plane or 3D space of the most important principal components, which are created by an uncorrelated linear combination of all original variables. Therefore, it distinguishes between studied objects, according to the region, and defines the causes of variability (Kruzlicova, Fiket, & Kniewald, 2013).

Concentrations data of all 29 elements was used to perform the PCA analysis. The total information content of the given number of principal components was expressed by cumulative percent (cum. %) value of the total variance. For instance, when all 29 variables were used, the first two principal components (PCs) represented 64.2%. The first principal component (PC1) represented 53.6% of the total variance and the next principal components, 10.6% (PC2).

In Fig. 2 are shown the most important PCA graphs, PC2 *vs.* PC1 score and loading plots, where the natural grouping of the studied 408 grape seed samples and the orientation of the variables are demonstrated. Figure 2a shows the score-plot for the first two PCs. It was observed a great overlap among the scores corresponding to different samples identified according to their origin. However, samples from the GY and JN departments showed positive scores on the considered PC1 and were differentiated from the rest of the groups having negative scores on PC1.

Figure 2b shows the loading plot in the PC2–PC1 plane, illustrating the orientation of the variables (concentrations of elements) with respect to the most informative principal components PC1 and PC2. The first principal component was strongly associated with the values of Eu, Gd, Pr, Tm, Lu, Nd, Sm, Y and Co, indicating higher concentrations in samples that showed positive scores (GY samples). On the other hand, Fe, Cu, Zn, Ti and Mn were the dominant variables in the second principal component, while samples with negative scores on PC2 correspond to high concentrations of these elements and positive scores to high concentrations of Nb and Tl.

In summary, the results obtained by PCA showed that only the samples from GY and JN regions could be differentiated considering their trace element profiles, but other seed samples with different geographical origin could not be solved by this unsupervised chemometric method. These findings suggested the need of using further complex chemometric algorithms to achieve the geographical classification of the samples.

3.2.2. Statistical classification analysis

Supervised pattern recognition techniques involve primarily classification to assign objects into groups or categories by creating classification rules (Brereton, 2015). Therefore, a classical method such as LDA was first applied considering all grape seed samples and their accuracy was analyzed. The first two canonical discriminant functions (DFs) explained 78.2% of the variance. Figure 3 shows the distribution patterns of the grape seed samples according to their geographical origins in the plot defined by the first two discriminant functions. This figure showed a notably overlap between groups. Three main groups were distinguished: GY (negative scores on DF1 and DF2), JN (positive scores on DF1 and negative on DF2) and the third group MP + RV + SM in the center of the graph.

The average prediction accuracies for all 408 samples were 91.6% for GY samples, 95% for JN samples, 78.3% for MP samples, 60.5% for RV samples and 55.1% for SM samples. It has to be mentioned that the samples from GY and JN regions were classified correctly as predicted from the graph. The main difficulties came up with the samples from the remaining localities. Nonlinear methods were clearly needed to solve the prediction of geographical origin of samples.

In order to perform a predictive classification analysis, data matrix was randomly divided into a training set (70% of the objects of the whole data matrix), with known class memberships to calculate a classifier and the test set (30%) containing the remaining objects not included in the training and with class memberships that serves to validate the models performance.

In the random sampling the division into subgroups was performed in a stratified manner, i.e. the proportion of each class in the original matrix was kept in the new subgroups. Since the results obtained were conditioned by the point where the division of the data matrix was made, this operation was repeated *n* times to obtain average values for success rate, allowing comparison between the different discriminatory methods assayed. The cases included in each set were randomly changed for each reproduced model.

In this work, five chemometric models were selected and tested, i.e. LDA, PLS-DA, k-NN, SVM and RF, to classify grape seeds samples in accordance with their geographical origins. The methods PLS-DA, kNN, SVM and RF, were required to optimize some parameters and build the model prior to their evaluation as a prediction tool. Each of the models was trained by using a *k*-fold cross-validation (repeated *n* times) on

training set to build the different classifiers. This procedure was repeated *n* times, so each subset was used for testing at least one time. In this work, $k = 10$ and $n = 5$ were applied.

When optimizing the parameters, the choices of the right number of significant components (*ncomp*) for PLS-DA; number of neighbor *k* for kNN; penalty factor C, ε of the ε-insensitive loss function and kernel type for SVM, and number of variables evaluated at each split (stratified sampling) (*mtry*) and number of trees (*nt*) for RF, were calculated by using the cross-validation technique described above. Thus, the maximum accuracy rate was selected based on this criterion. Once the optimal values for each model were selected, 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 overall accuracy rate of the proposed model was considered for evaluation of the classification performance with the supervised methods on the testing set. The results indicating the performance of the different classification methods are shown in Table 2.

Based on the analysis of Table 2, it was observed that the five chemometric methods displayed different degrees of success in the prediction of test samples. The order of successful identification rates was as follows: RF>SVM >k-NN> LDA>PLS-DA. Linear models such as LDA and PLS-DA, presented the worst performance from the overall accuracy point of view, therefore the use of nonlinear methods resolved the classification problem, and RF was the ideal model for discriminating the grape seed samples according to their geographical origins, with an overall classification accuracy of 98.3%. Therefore, good discrimination of the samples under study was obtained with nonlinear models, which

can be explained by the flexibility and ability of the algorithm for creating a generalized model, even for small training groups (Gaiad, Hidalgo, Villafañe, Marchevsky, & Pellerano, 2016).

Finally, an additional comparison of the performance of the five classification methods applied in this work was performed by using the repeated one-third holdout estimator. Since the result depend on the right choice of the test data, data matrix was 30-folds divided and the results were compared each other. This procedure provided a distribution for the overall accuracy to compare the performance of each method at its optimal parameter configuration. The distributions represented by Box and Whisker plots can be observed in Fig. 4. Thus, the limits around the mean represent the amplitude of the dispersion of the results obtained for each resample cycle. The results obtained by this approach confirmed that the best method for this data set was RF, with a median accuracy of almost 97% and kappa value of 11.

4. Conclusions

The first intra-regional classification of grapes produced in different vineyards of Mendoza province (Argentina) was achieved in this work based on multielemental analysis of grape seeds. The application of modern and multielemental techniques such as ICP-MS, combined with modern chemometrics techniques, such as RF, provided a robust tool for geographical classification of grapes produced in the different locations of this region. The prediction results of both training and tested samples demonstrated that the proposed RF model is an effective and efficient approach for the classification of grape seed samples. Moreover, 500 trees and 29 variables were optimized and selected as the best parameters for the classification of the grape seed samples. The predicted models for the grape seed sample contain information about their

chemical composition that contributes to classification of different geographical origin. The models established using the LDA, PLS-DA, k-NN, SVM, and RF methods were applied to the predicted geographical origin, showing that SVM and RF methods have a high prediction accuracy rate. In fact, the average prediction accuracy rate of the RF method was generally higher than other methods. Therefore, the RF method was proved to be a promising tool for classification and quality control of raw materials used in wine-making industry.

Acknowledgments

This work was supported by Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Agencia Nacional de Promoción Científica y Tecnológica (FONCYT) (PICT-BID) and Universidad Nacional de Cuyo (Argentina). The authors would like to thank Andrés Lo Vecchio for drawing the map containing geographical locations of sampling vineyards.

References

Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: linear models. PLS-DA. *Analytical Methods*, *5*(16), 3790–3798.

Barker, M., & Rayens, W. (2003). Partial least squares for discrimination. *Journal of Chemometrics*, *17*(3), 166–173.

Beebe, K. R., Pell, R. J., & Seasholtz, M. B. (1998). Chemometrics: A practical guide. In John Wiley Sons (Ed.). New York.

Bertoldi, D., Larcher, R., Nicolini, G., Bertamini, M., & Concheri, G. (2009). Distribution of rare earth elements in Vitis vinifera L.

"Chardonnay" berries. *Vitis*, *48*(1), 49–51.

Brereton, R. G. (2015). Pattern recognition in chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *149*, 90–96.

Castex, V., Tejeda, E. M., & Beniston, M. (2015). Water availability, use and governance in the wine producing region of Mendoza, Argentina. *Environmental Science and Policy*, *48*, 1–8.

Cugnetto, A., Santagostini, L., Rolle, L., Guidoni, S., Gerbi, V., & Novello, V. (2014). Tracing the "terroirs" via the elemental composition of leaves, grapes and derived wines in cv Nebbiolo (Vitis vinifera L.). *Scientia Horticulturae*, *172*, 101–108.

Fabani, M. P., Toro, M. E., Vázquez, F., Díaz, M. P., & Wunderlin, D. A. (2009). Differential absorption of metals from soil to diverse vine varieties from the valley of tulum (Argentina): Consequences to evaluate wine provenance. *Journal of Agricultural and Food Chemistry*, *57*(16), 7409–7416.

Ferrer-Gallego, R., Hernández-Hierro, J. M., Rivas-Gonzalo, J. C., & Escribano-Bailón, M. T. (2013). A comparative study to distinguish the vineyard of origin by NIRS using entire grapes, skins and seeds. *Journal of the Science of Food and Agriculture*, *93*(4), 967–972.

Fundación ProMendoza. (2016). Accessed June 15, 2016, from http://www.promendoza.com/

Gaiad, J. E., Hidalgo, M. J., Villafañe, R. N., Marchevsky, E. J., & Pellerano, R. G. (2016). Tracing the geographical origin of Argentinean lemon juices based on trace element profiles using advanced chemometric techniques. *Microchemical Journal*, *129*, 243–248.

Galgano, F., Favati, F., Caruso, M., Scarpa, T., & Palma, A. (2008). Analysis of trace elements in southern Italian wines and their classification

according to provenance. *LWT - Food Science and Technology*, *41*(10), 1808–1815.

González-Centeno, M. R., Rosselló, C., Simal, S., Garau, M. C., López, F., & Femenia, A. (2010). Physico-chemical properties of cell wall materials obtained from ten grape varieties and their byproducts: Grape pomaces and stems. *LWT - Food Science and Technology*, *43*(10), 1580–1586.

International Organization of Vine and Wine (OIV). (2016). Aspectos de la Coyuntura Mundial. Accessed June 24, 2016, from http://www.oiv.int/es/normas-y-documentos-tecnicos/analisis-estadisticos/analisis-de-la-coyuntura

Jackson, R. S. (2014). 6 - Chemical Constituents of Grapes and Wine BT - Wine Science (Fourth Edition). In *Food Science and Technology* (pp. 347–426). San Diego: Academic Press.

Kruzlicova, D., Fiket, Ž., & Kniewald, G. (2013). Classification of Croatian wine varieties using multivariate analysis of data obtained by high resolution ICP-MS analysis. *Food Research International*, *54*(1), 621–626.

Lachman, J., Hejtmánková, A., Hejtmánková, K., Horníčková, Š., Pivec, V., Skala, O., Dědina M., & Přibyl, J. (2013). Towards complex utilisation of winemaking residues: Characterisation of grape seeds by total phenols, tocols and essential elements content as a by-product of winemaking. *Industrial Crops and Products*, *49*, 445–453.

Liaw, a, & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2*, 18–22.

Massart, D. L., Vandeginste, B. G. M., Buydens, L. M. C., De Jong, S., Lewi, P. J., & Smeyers-Verbeke, J. (1997). *Handbook of chemometrics*

and qualimetrics (Vol.B). Amsterdam: Elsevier.

Mironeasa, S., Leahu, A., & Codin, G. (2010). Grape Seed : physico-chemical , structural characteristics and oil content, *16*(1), 1–6.

Mitić, S. S., Obradović, M. V., Mitić, M. N., Kostić, D. a., Pavlović, A. N., Tošić, S. B., & Stojković, M. D. (2011). Elemental Composition of Various Sour Cherry and Table Grape Cultivars Using Inductively Coupled Plasma Atomic Emission Spectrometry Method (ICP-OES). *Food Analytical Methods*, *5*(2), 279–286.

Mutihac, L., & Mutihac, R. (2008). Mining in chemometrics. *Analytica Chimica Acta*, *612*(1), 1–18.

Ozcan, M. M. (2010). Mineral Contents of Several Grape Seeds. *Asian Journal of Chemistry*, *22*(8), 6480–6488.

Pohl, P. (2007). What do metals tell us about wine? *TrAC - Trends in Analytical Chemistry*, *26*(9), 941–949.

R Core Team. (2016). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Accessed from https://www.r-project.org/

Reynolds, A. G. (2010). 11 – Viticultural and vineyard management practices and their effects on grape and wine quality. In *Managing Wine Quality* (pp. 365–444).

Rogiers, S. Y., Greer, D. H., Hatfield, J. M., Orchard, B. A., & Keller, M. (2006). Mineral sinks within ripening grape berries (Vitis vinifera L.). *Vitis - Journal of Grapevine Research*, *45*(3), 115–123.

Saurina, J. (2010). Characterization of wines using compositional profiles and chemometrics. *TrAC - Trends in Analytical Chemistry*, *29*(3), 234–

245.

Serafim, F. A. T., Pereira-Filho, E. R., & Franco, D. W. (2016). Chemical data as markers of the geographical origins of sugarcane spirits. *Food Chemistry*, *196*, 196–203.

Spanghero, M., Salem, A. Z. M., & Robinson, P. H. (2009). Chemical composition, including secondary metabolites, and rumen fermentability of seeds and pulp of Californian (USA) and Italian grape pomaces. *Animal Feed Science and Technology*, *152*(3–4), 243–255.

Spring, M. L. (2008). *Support Vector Machines* (Vol. 1). New York, NY: Springer New York.

Tangolar, S. G., Ozoğul, Y., Tangolar, S., & Torun, A. (2009). Evaluation of fatty acid profiles and mineral content of grape seed oil of some grape genotypes. *International Journal of Food Sciences and Nutrition*, *60*(1), 32–39.

Tanner, M., & Günther, D. (2009). Short transient signals, a challenge for inductively coupled plasma mass spectrometry, a review. *Analytica Chimica Acta*, *633*(1), 19–28.

van Leeuwen, C. (2010). 9 – Terroir: the effect of the physical environment on vine growth, grape ripening and wine sensory attributes. In *Managing Wine Quality* (pp. 273–315).

Versari, A., Laurie, V. F., Ricci, A., Laghi, L., & Parpinello, G. P. (2014). Progress in authentication, typification and traceability of grapes and wines by chemometric approaches. *Food Research International*, *60*, 2–18.

Yang, Y., Duan, C., Du, H., Tian, J., & Pan, Q. (2010). Trace element and rare earth element profiles in berry tissues of three grape cultivars.

American Journal of Enology and Viticulture, *61*(3), 401–407.

Figure captions

Fig.1. Geographical locations of sampling vineyards selected for the collection of grape seeds studied in this work.

Fig.2. Score (a) and loading (b) plots of the first principal component (PC1) versus the second principal component (PC2).

Fig.3. Scatter plot of the first two discriminant functions of linear discriminant analysis of grape seed samples according to their geographical origin.

Fig.4. Box and whisker plot comparing the chemometrics models applied for grape seeds classification.

Fig. 1

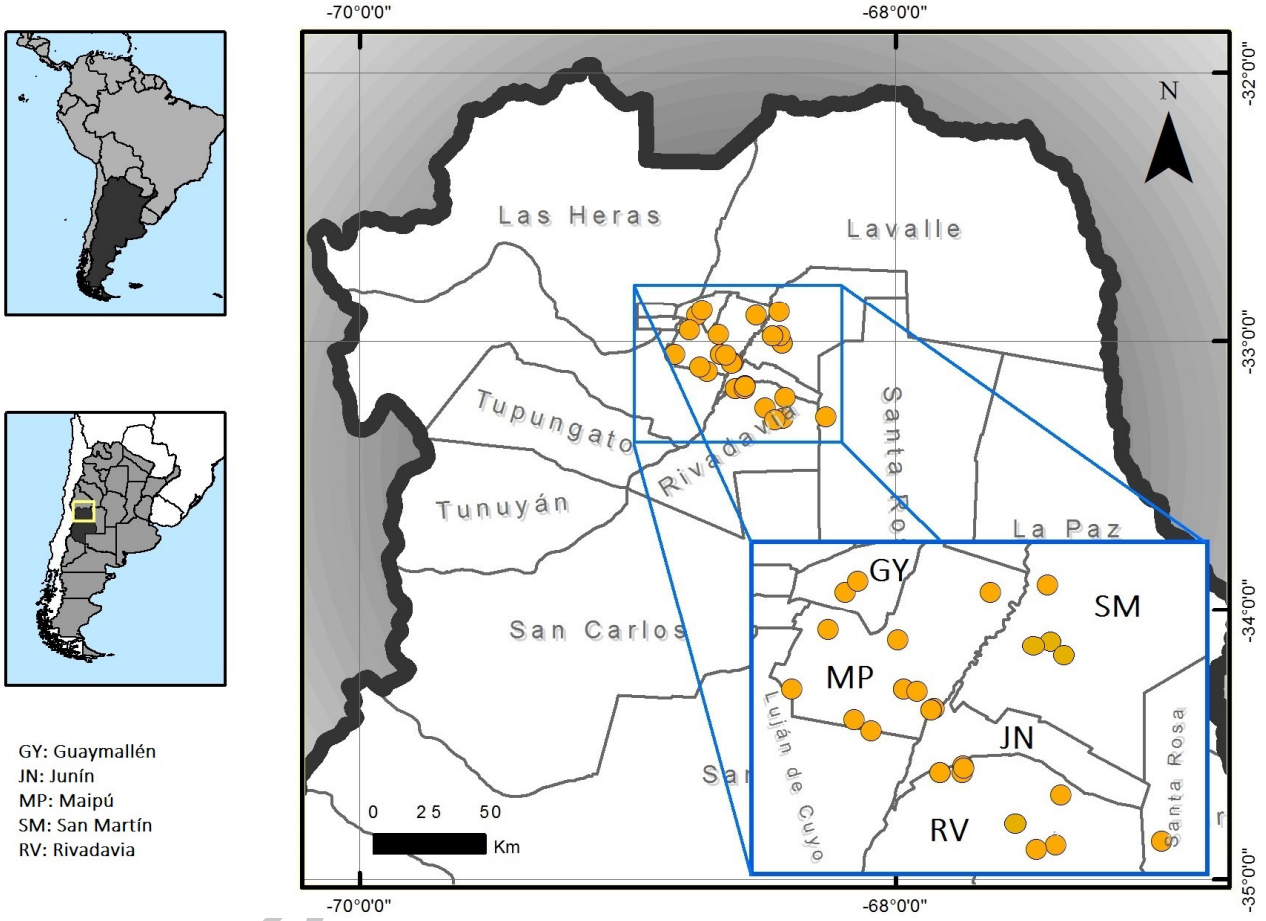
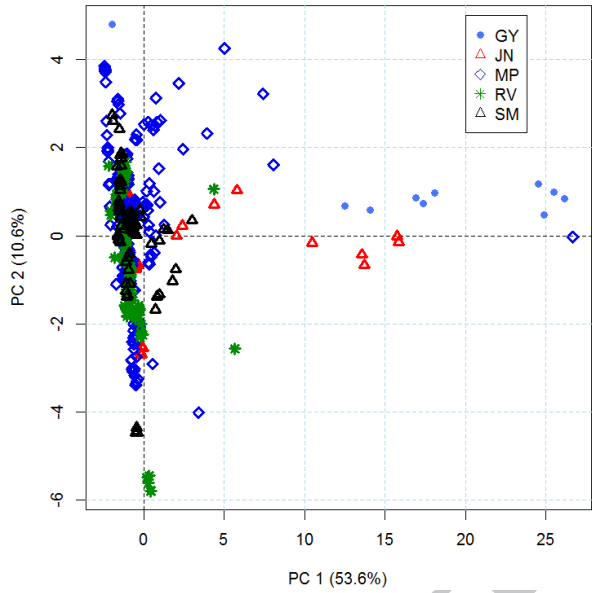


Fig. 2

a)



b)

Fig. 3

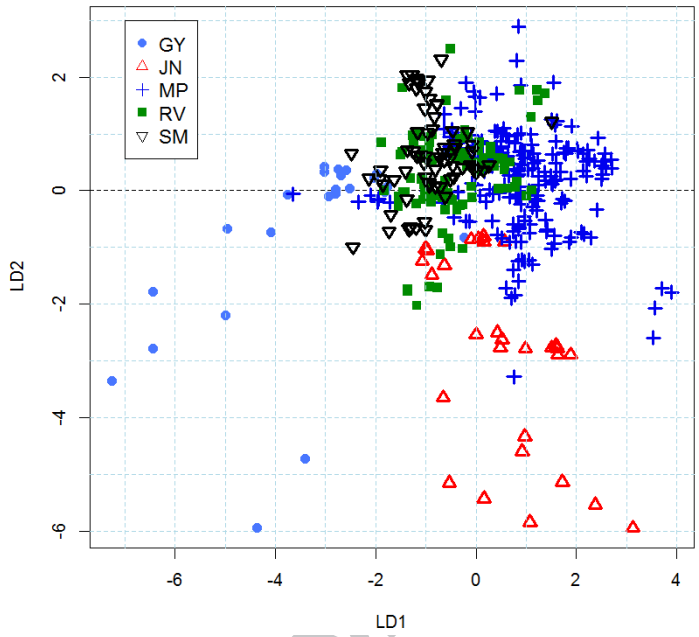


Fig. 4

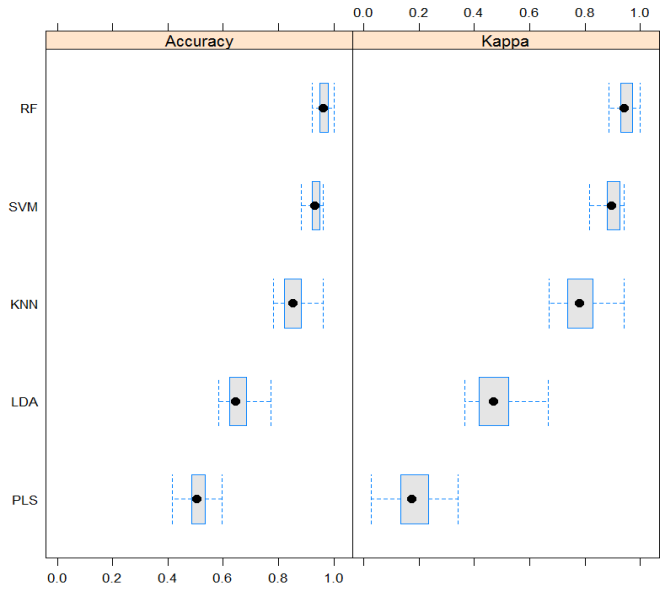


Table 1. Elemental composition of grape seed samples according to their geographical origin (Concentrations are expressed as µg/g of dry matter basis).

Element	LOD[a] (µg/L)	Rivadavia (n = 92)		San Martín (n = 69)		Guaymallén (n = 25)		Junín (n = 29)		Maipú (n = 193)	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Ag	1.11	0.074	0.042	0.04	0.018	0.255	0.307	0.212	0.193	0.067	0.087
As	0.48	0.076	0.091	0.103	0.055	0.696	0.961	0.251	0.283	0.186	0.371
Ce	0.45	0.148	0	<LOD	<LOD	0.016	0.011	0.041	0.013	0.029	0.037
Co	0.48	0.013	0.008	0.016	0.017	0.079	0.1	0.034	0.036	0.013	0.02
Cs	0.12	0.009	0.015	0.003	0.002	0.038	0.023	0.017	0.015	0.018	0.0018
Cu	0.9	13.0	3.3	12.0	2.732	9.002	2.345	12.9	2.129	12.4	3.225
Eu	0.18	0.003	0.006	0.007	0.008	0.143	0.028	0.052	0.056	0.01	0.024
Fe	2.73	44.2	11.6	37.5	10.7	39.3	7.053	38.7	11.7	39.5	12.2
Ga	0.33	0.024	0.017	0.016	0.014	0.134	0.133	0.053	0.045	0.036	0.057
Gd	0.57	0.004	0.007	0.009	0.007	0.238	0.069	0.058	0.064	0.015	0.043
La	0.24	0.037	0.03	0.024	0.017	0.102	0.061	0.057	0.038	0.025	0.021
Lu	0.03	0.003	0.005	0.003	0.003	0.042	0.007	0.016	0.015	0.005	0.01
Mn	0.27	18.2	4.36	19.6	4.752	18.41	4.117	20.3	4.247	15.7	6.683
Mo	0.36	2.391	0.756	2.36	0.625	3.844	2.469	6.914	4.026	4.34	3.116
Nb	0.15	0.094	0.142	0.156	0.162	0.122	0.1	0.303	0.381	0.21	0.331
Nd	1.38	0.03	0.055	0.011	0.011	0.247	0.055	0.09	0.078	0.034	0.078
Ni	0.27	3.379	2.808	3.424	3.874	2.898	1.443	4.971	5.074	3.21	2.235
Pr	0.18	0.006	0.011	0.012	0.008	0.093	0.026	0.042	0.04	0.014	0.025
Rb	0.42	1.898	1.027	2.027	1.219	6.247	6.776	5.217	8.872	3.364	2.519
Sm	0.78	0.004	0.007	0.009	0.007	0.191	0.056	0.034	0.043	0.012	0.036
Te	2.85	1.168	1.315	2.001	1.868	22.5	13.91	3.56	5.48	2.029	3.991
Ti	3.45	2.972	1.268	2.649	1.231	3.256	1.874	3.099	0.987	2.649	1.119

Tl	0.03	0.093	0.047	0.117	0.079	0.182	0.045	0.145	0.033	0.111	0.071
Tm	0.06	0.007	0.009	0.004	0.003	0.04	0.012	0.015	0.012	0.007	0.011
U	0.06	0.047	0	0.01	0.006	0.021	0.007	0.036	0.012	0.023	0.023
V	0.42	0.017	0.01	0.008	0.004	0.152	0.064	0.032	0.028	0.033	0.046
Y	0.18	0.009	0.02	0.014	0.012	0.111	0.022	0.04	0.046	0.009	0.019
Zn	0.18	18.5	6.34	17.6	7.163	16.5	3.403	18.0	10.2	16.8	5.834
Zr	0.51	0.136	0.047	0.157	0.072	0.211	0.144	0.197	0.12	0.168	0.091

[a]LOD: instrumental limit of detection.

Table 2. Discrimination results obtained with the different chemometrics models

Groups	Number of samples		LDA		PLS-DA (*ncomp* = 3)[a]		k-NN (k = 5)[b]		SVM (C = 32; ε = 0.09)[c]		RF (nt = 500; mtry = 11)[d]	
	Training set	Test set	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
GY	18	7	71	100	57	100	71	100	43	100	100	100
JN	20	9	62	98	0	100	87	95	100	100	100	100
MP	135	58	80	77	89	30	91	92	98	89	98	98
RV	64	28	40	77	26	90	85	95	89	100	96	100
SM	48	21	40	92	0	95	75	97	100	99	100	99
Mean accuracy (%)			**63.0**		**52.1**		**85.7**		**93.3**		**98.3**	

[a] ncomp: number of significant components.

[b] k: number of k neighbors.

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

[d] nt: number of trees; mtry: number of variables tried at each split.

Highlights

- Multi-element contents were determined in grape seeds by ICP-MS.
- LDA, PLS-DA, k-NN, SVM, and RF chemometrics methods were applied for classification.
- Random Forest (RF) method was useful for geographical classification.
- Intra-regional classification of grape seeds from Mendoza province was achieved.