Guijarro-Díez, M., Nozal, L., Marina, M.L. & Crego, A.L. 2015, "Metabolomic fingerprinting of saffron by LC/MS: novel authenticity markers", *Analytical and Bioanalytical Chemistry,* vol. 407, no. 23, pp. 7197-7213.

The final publication is available at Springer via:

http://dx.doi.org/10.1007/s00216-015-8882-0

Metabolomic fingerprinting of saffron by LC/MS: Novel authenticity markers.

Miguel Guijarro-Díez, Leonor Nozal, María Luisa Marina, Antonio Luis Crego*.

Department of Analytical Chemistry, Physical Chemistry and Chemical Engineering, University of Alcalá, Ctra. Madrid-Barcelona Km. 33,600, 28871 Alcalá de Henares (Madrid), Spain.

*Corresponding author

E-mail: antonio.crego@uah.es

Fax: + 34-91-8854971

Keywords: Metabolomic fingerprinting, liquid-chromatography-time of flight-mass spectrometry, saffron, authenticity markers.

Abstract

An untargeted metabolomic approach using liquid chromatography coupled to electrospray ionization-time of flight-mass spectrometry was developed in this work to identify novel markers for saffron authenticity which is an important matter related to consumer protection, quality assurance, active properties, and also economical impact (saffron is the most expensive spice). Metabolic fingerprinting of authentic and suspicious saffron samples from different geographical origin was obtained and analysed. Different extracting protocols and chromatographic methodologies were evaluated to obtain the most adequate extracting and separation conditions. Using an ethanol/water mixture at pH 9.0 and an elution gradient with a fused core C18 column enabled obtaining the highest number of significant components between authentic and adulterated saffron. By using multivariate statistical analysis, predictive classification models for authenticity and geographical origin were obtained. Moreover, 84 and 29 significant metabolites were detected as candidates for markers of authenticity and geographical origin, respectively, from which only 34 metabolites were tentatively identified as authenticity markers of saffron, but none related to its geographical origin. Six characteristic compounds of saffron (*kaempferol 3-O-glucoside, kaempferol 3-O-sophoroside, kaempferol 3,7-O-diglucoside, kaempferol 3,7,4´-O-triglucoside, kaempferol 3-O-sophoroside-7-O-glucoside, and geranyl-O-glucoside)* were confirmed by comparing experimental MS/MS fragmentation patterns with those provided in scientific literature being proposed as novel markers of authenticity.

1. Introduction

Saffron is produced from dried stigmas of *Crocus sativus* L., and is the most expensive spice in the world due to the direct labour required for its cultivation, harvesting, and handling. It has long been used as flavouring and colouring in food preparation, and it is also known for a wide range of health promoting benefits in traditional and modern medicine [1]. The three main secondary metabolites in saffron are crocins (crocin and its derivatives as colour factors), a family of yellow pigments freely soluble in water, picrocrocin (taste factor), a colourless and bitter tasting glycoside and safranal (perfume factor), a volatile oil and the saffron's characteristics odour and aroma. In addition, saffron also contains flavonoids, proteins, sugars, vitamins, amino acids, mineral matter, gums, and other chemical compounds [2].

Due to its high price and limited production, saffron has been subjected to various types of adulteration over the centuries. Saffron adulterations are done in order to increase its weight with foreign matters, and/or to enhance its colour with natural or synthetic colorants to mask foreign matters addition or to improve their colouring properties. Common fraudulent practices include the addition of different plant stuffs with similar colour and morphology. Historically, the most frequently encountered materials have been *Crocus sativus* stamens (even styles or strips of the corolla), *Carthamus tinctorius* L. petals (safflower), *Curcuma longa* L. rhizomes (turmeric), *Calendula officinalis* L. and *Arnica montana* L. flowers, *Bixa orellana* L. seeds, *Hemerocallis lilioasphodelus* L. petals and *Crocus vernus* L. stigmas [3, 4]. More recently, a new adulteration by gardenia additions may have reached the European market. The use of gardenia (extract obtained from the fruits of *Gardenia jasminoides* J. Ellis) is a sophisticated method of adulteration, considering that gardenia and saffron differ merely in the pigments contained [5, 6].

Up to the moment, the quality of saffron is certified in the international trade market following the ISO 3632 normative [7, 8], which certifies by means of a combination of UV spectrophotometric measurements of picrocrocin at 250 nm and safranal at 310 nm, and chromatographic profiles of polar dyes and pigments (i.e. crocins) at 440 nm. The measured parameters allow controlling the quality of saffron through the contents of picrococin, safranal and crocins as well as the possible presence of some dyes that can be toxic. This ISO normative is currently under revision to incorporate non-polar dyes and pigments. However, this update will not solve the important drawbacks of the use of these standards, highlighted by recent literature, and their weak reliability in the detection of plant foreign matter. Unfortunately, the ISO 3632 standards are non-specific and unable to separate authentic and adulterated saffron adequately. In particular, it has been demonstrated that a contamination of ground saffron with amounts of up to 20% (w/w) of *Calendula* flowers, safflower or turmeric was not revealed by the ISO 3632 standards, as it was recently reported [9].

Various authors have proposed several analytical methods for the detection of plant adulterants in saffron, as UV-Vis spectrophotometric measurements [10-12], near infrared spectroscopy [13], raman and nuclear magnetic resonance spectroscopy (NMR) [14, 15], capillary electrophoresis [16], and high-performance liquid chromatography without and with Mass Spectrometry (MS) detection [17-20]. Most of these methods used to detect adulteration of saffron are based on targeted analysis of a number of compounds, and their main disadvantage is that usually only known (targeted) compounds can be detected as indicators of certain type of adulteration. Finally, the use of molecular methods to detect DNA markers has been employed so far with encouraging results [21-24]. Low amounts (up to 1%) of plant adulterants (safflower and turmeric) used as bulking agents were detected as a fraud in commercial saffron. Nevertheless, there is still an on going demand for the development of

faster, simple and robust screening methods suited for identifying saffron adulteration, especially at levels that make practical economic sense.

In conclusion, despite these attempts for quality control and standardization, there is a rich history of saffron adulteration constantly evolving. Therefore the search for markers of authenticity instead of markers of adulteration would be the most intelligent and definitive strategy to detect adulteration of saffron in which the implementation of metabolomics approaches provides the tools needed to face this challenge [25]. The general principle of metabolomics is to characterize biological samples by the production of a chemical signature or fingerprint. From an analytical point of view, the most widely used technique for this purpose has been NMR, but MS is becoming more widely used in this field [26, 27]. Indeed, MS offers higher performance in terms of sensitivity, which is extremely useful for measuring species with low abundance, as that provides valuable information. Moreover, the specificity of MS (through high-resolution and/or MS/MS techniques) can help and even facilitate elucidation of the chemical structures of potential metabolites of interest (i.e. identification of biomarkers).

These new strategies are also being studied for their application to solve current food fraud issues where classical methods fail to detect them. Thus, a methodology based on NMR metabolite fingerprinting has been published very recently proving to be efficient for determining and identifying fraudulent additions of bulking agents to saffron, considering the difficulties in detecting saffron fraud according to the ISO 3632 standard methods, especially when plant adulterants are involved and the spice is commercialized in powder form [28]. Taking into account the deficiency of established methodologies to detect saffron adulteration with plant adulterants, the method developed could be viable for dealing with saffron frauds at a minimum level of 20% (w/w) with four typical plant-derived materials employed as bulking agents in saffron (*Crocus sativus* stamens, safflower, turmeric and gardenia).

The main advantage of metabolomics in food authentication makes use of its untargeted nature, which can enable the detection of emerging frauds. Thus, the aim of this work was to explore, for the first time, the feasibility of applying liquid chromatography-quadrupole time of flight mass spectrometry (LC-(QTOF)MS) together with multivariate statistical analysis for the assessment of the authenticity of saffron using an untargeted metabolomics approach. Metabolomics combined with a sensitive analytical technique as MS and with multivariate analysis is a valid and powerful tool to investigate the quality and authenticity of saffron. The broad applicability of LC-(QTOF)MS to metabolites of all classes justified its choice for the problem under consideration. In the present study, authentic saffron samples from Spain and Iran, as high value samples and the most popular saffron consumed globally, frequently sold as 100% pure, were chosen as 'saffron model' for the study.

2. Materials and Methods

2.1 Chemicals and samples

Acetonitrile, ethanol, methanol, acetic and formic acid of HPLC grade were purchased from Scharlab (Barcelona, Spain). Sodium borate and phosphoric acid were obtained from Sigma (St. Louis, MO, USA). Ultrapure water for the chromatographic mobile phase and for preparing the saffron extracts was obtained from a Milli-Q system (Millipore, Madrid, Spain). Several standards, verapamil, niflumic acid, propranolol, terfenadine, geranic acid, genistein, baicalein, quercitrin, rutin, quercetin and kaempferol were purchased from Sigma (St. Louis, MO, USA).

Ten saffron samples (stigmas or powdered) from Spain and Iran were provided by "Carmencita" company (Alicante, Spain). Their quality and authenticity had been checked according to the ISO 3632 parameters and HPLC analysis of dyes. All authentic saffron

samples belonged to the commercial category I. In addition, a total of ten saffron samples suspected of adulteration (stigmas or powdered) purchased in Spain and Iran markets were also provided by "Carmencita" company. The suspicious samples were considered as such according to the criteria of the market based on their low cost and/or questionable origin.

2.2 Sample Preparation

Saffron stigmas were finely ground in a mortar with stainless steel balls Ultra Turrax (IKA, Staufen, Germany) for 2 min. Ground and powdered saffron samples were extracted by ultrasonic-assisted solid/liquid extraction (0.3 g in 3 ml) using ethanol:borate buffer at pH 9.0 (50:50 v/v) as extractant solvent for 15 min at room temperature. After extraction, samples were centrifuged for 15 min at 4000 *g*. 2 mL of supernatant fraction was collected, diluted 1:1 with the extractant solvent, and ultra-filtered through a 3 kDa cut-off filter (Amicon Ultra Filters, Merck, Darmstadt, Germany) to remove proteins and large molecules.

A quality control (QC) sample was prepared by combining equal aliquots from each saffron extract (both authentic and suspicious). In the case of untargeted metabolomic approaches, the performance of the chemometric models could be ensured if a "biological QC" sample (prepared by combining extracts of all samples employed in the metabolomic approach) analyzed repeatedly is predicted at the middle of the model [29]. Additionally, a "test sample" was prepared by adding four standards (verapamil, niflumic acid, propranolol, and terfenadine) to the QC sample at 0.1 µg/mL in order to characterize the LC-MS system.

2.3 Analytical setup

The analysis was completed using an HPLC system (1100 series, Agilent Technologies, Waldbronn, Germany) coupled to a quadrupole time-of-flight (QTOF) equipped with an orthogonal electrospray ionization source (ESI) with Jet Stream thermal

focusing technology (6530 series, Agilent) and operating in positive or in negative ion modes (i.e. polarity-switch was not used, and the samples were analyzed twice). The HPLC system consisted of a degasser, a quaternary pump, an automatic injector, and a thermostatic column compartment. Agilent Mass Hunter Qualitative Analysis software (B.04.00) was used for MS control, data acquisition, and data analysis. A solution of compounds whose masses are known with great accuracy was continually infused to the system to allow constant mass correction for accurate mass calibration. Thus, during all analysis, two reference masses were used, *m/z* 121.0509 ($C_5H_4N_4$) and *m/z* 922.0098 ($C_{18}H_{18}O_6N_3P_3F_{24}$) for positive ionization mode, and *m/z* 112.9856 ($C_2O_2F_3(NH_4)$) and m/z 1033.9881 ($C_{18}H_{18}O_6N_3P_3F_{24}$) for negative ionization mode.

2.4 LC-MS conditions

Analysis was completed at 40 ºC using different columns, all of them Ascentis Express (Sigma, St. Louis, USA), with the same dimension (100 x 2.1 mm i.d.) and type of packed bed (fused-core® particles with 0.5 µm thick porous shell and an overall particle size of 2.7 µm) but of different kinds of stationary phases (C18, Cyano or HILIC). In addition, Ascentis Express guard columns (5 x 2.1 mm i.d.) of the same material as the analytical column in each case were employed. The system was operated with an injected volume of 15 µL and a flow rate of 0.4 mL min^{-1}. The mobile phases consisted of 0.1% formic acid for ESI+ or 10 mM ammonium formate for ESI- in Milli-Q water (eluent A) and 0.1% formic acid for ESI+ or 10 mM ammonium formate for ESI- in acetonitrile (eluent B), using different elution gradients according to the selected stationary phase. For the C18 and Cyano columns, the linear gradient started from 5% B to 95% B in 33 min and returned to starting conditions in 1 min, keeping the re-equilibration at 5% B for 15 min. For the HILIC column the same linear gradient profile was used but eluent A was 0.1% formic acid for ESI+ or 10 mM

ammonium formate for ESI- in acetonitrile and B was 0.1% formic acid for ESI+ or 10 mM ammonium formate for ESI- in Milli-Q water.

The ionization source conditions were as follows: capillary voltage of 3 kV with a nozzle voltage of 0 V; nebulizer pressure at 2.7 bar; sheath gas of jet stream of 6.5 L/min at 300 ºC; and drying gas of 10 L/min at 300 ºC. The cone voltage (fragmentator) after sampling capillary was set at 175 V. The skimmer and octapole voltages were 60 V at 750 V, respectively.

MS analysis was performed in both positive and negative ESI modes with the mass range set at *m/z* 100-1700 (extended dynamic range) in full scan resolution mode at a scan rate of 2 scans per second. These conditions allowed to reach an average mass resolution of 0.0001 Da calculated by full width half maximum of standards in the “test sample”. Also, to characterize the LC-MS system in terms of mass accuracy and reproducibility, the "test sample" was injected at the beginning, middle, and end of the analysis sequence.

For sample analysis, replicate extractions (n = 3) of all samples were used in random sequence to ensure that any experimental trends observed were directly associated with the sample and not due to any change in the instruments performance or sample preparation over time. Also, QC sample was injected at the beginning of the run and after every 3 real samples analysis to ensure the stability and repeatability of the LC–MS system.

2.5 Data treatment

The use of dedicated software solutions for handling the large datasets typically produced in metabolomics is unavoidable. Hence a *data processing* to convert the initial three-dimensional raw data (*m/z,* retention time, and intensity of ion current) to a two-dimensional data table reporting time-aligned and mass-aligned abundances of chromatographic peaks was performed. First, the resulting data file was cleaned of

background noises and unrelated ions by the Molecular Feature Extraction (MFE) tool in Mass Hunter Qualitative Analysis software. MFE algorithm uses the accuracy of the mass measurements to group ions related by charge-state envelope, isotopic distribution, and/or the presence of adducts and dimmers. The MFE then creates a listing of all possible components (molecular features) as represented by the full TOF mass spectral data. For data extraction by MFE, the molecular features present in a sample were determined using the following parameters in the software: target data type of small molecules (chromatographic); peaks with height ≥ 500 counts; peak spacing tolerance = 0.0025 m/z, plus 7.0 ppm; isotope model = common organic molecular; limited assigned change = 3; and then, to find co-eluting adducts of the same feature the following adduct settings were applied: H^+, Na^+, K^+, and NH_4^+ in positive ionization, and $HCOO^-$ for negative ionization. Second, other *data processing* such as filtering and alignment were performed with Mass Profiler Professional software (B.02.00, Agilent). The retention time (RT) and *m/z* alignment of the respective peaks was carried out using the following parameters depending on the mass accuracy and reproducibility obtained in the analysis of the "test sample": initial retention time 3 min, final retention time 33 min, mass tolerance 0.02 Da, mass window 0.02 Da, and retention time window 0.1 min. To clean the data matrix from random signals and to select only features with biological meaning filter by frequency was applied. Choosing the data present in all quality control samples performed primary filtering. Also, features were filtered by choosing masses that were present in all samples at least in one of two groups for comparison authentic vs. suspicious and Spanish vs. Iranian. A secondary filtering was performed by choosing the data with a coefficient of variation below 35%.

Further *data pretreatment* was performed to the data sets obtained from the previous data-processing stage before application of statistics. First, the data sets were transformed by applying common logarithm to intensities in order to reduce the influence of a few

particularly intense signals that can strongly influence statistical analysis and subsequent interpretation. This aspect is of particular concern in metabolomics, due to the huge dynamic range in terms of metabolite-concentration levels. Second, the data sets were pre-treated using the Pareto scaling (the square root of the standard deviation is used as the scaling factor).

The data tables generated in the previous stage of data processing was analyzed comprehensively with appropriate statistical tools. Thus, *data analysis* was performed with *univariate methods* using Microsoft Excel software (2010, Redmon, WA, USA) to reveal potential candidate compounds with significant differences in terms of abundance between two groups of samples (authentic and suspicious). The normality of distribution for each data set was assessed using the "Wilk-Shapiro's test", the homogeneity of variances was studied using the "Levene´s test", and "t-test" (assumed equal variance) or "Welch's test" (assumed unequal variance) could reveal potential candidate compounds with significant differences in terms of abundance between two groups of samples (authentic and suspicious). Likewise, *multivariate statistical analysis* was performed on mass spectral data sets using SIMCA-P+12.0 software (Umetrics, Umeå, Sweden). Unsupervised principal component analysis (PCA) was applied to represent the sample distribution in the multivariate space. Supervised partial least squares-discriminant analysis (PLS-DA) and orthogonal partial least squares-discriminant analysis (OPLS-DA) were used in order to reduce the model complexity by removing the systematic variations in the X matrix that were not related to Y response (structured noise) maximizing the separation among samples.

The quality of the models was described by the goodness of fit (R^2 value), and the predictive ability (Q^2 value). Recognition ability (R^2) represents a percentage of successfully classified samples in the training set. Prediction ability (Q^2) is a percentage of correctly classified samples in the test set by using the model developed during the training step. In addition, to ensure the performance of the models, QC analysed at the beginning of the run

and after every 3 real samples analysis should be predicted at the middle of the model since it was prepared by combining equal aliquots from each saffron extract (both authentic and suspicious). Also, to avoid the risk of over fitting for a discriminant analysis models used for the selection of statistically significant metabolites according to jack-knifed confidence intervals, models were validated by the use of a cross-validation tool using a 1/3 out approach [29, 30]. The data set was divided into three parts and 1/3rd of samples were excluded to build a model with the remaining 2/3rd of samples. This new model then predicted excluded samples and the procedure was repeated until all samples had been predicted at least once. The percentage of correctly classified samples was calculated each time.

Finally, selection of potential biomarkers was maintained for each comparison based on different tests: (i) "t-test" or "Welch's test" depending on the homogeneity of variances, calculated by using Microsoft Excel; and (ii) S-plot and jack-knife confidence intervals obtained for OPLS-DA models in SIMCA-P+12.0.

2.6 Databases for Identification

Metabolites were tentatively identified by searching by mass accuracy against the online available databases such as the METLIN (http://metlin.scripps.edu), HMDB (http://hmdb.ca), KEGG (http://genome.jp/kegg), FooDB (http://foodb.ca/), and lipidMAPS (http://www.lipidmaps.org/data/databases.html). For features, individual searching in databases was performed employing an error of 10 ppm.

3. Results and discussion

3.1 LC-MS metabolomic analysis

The extremely wide diversity of potential metabolites present in a sample in terms of chemical structures and concentrations means that it is unrealistic to have the goal of

measuring all of them in metabolomics. Thus, minimum sample preparation is preferred, especially for untargeted applications without any presupposed hypothesis. For solid matrices such saffron samples an extraction step is required for transferring the metabolome compounds into a liquid phase [31], and a subsequent protein-elimination step to limit ion suppression when using ESI for liquid chromatography with MS is employed [32]. Considering the fact that any sample treatment step can potentially result in its alteration and consecutive losses of some metabolites, only three sample preparation steps after extraction, centrifugation, dilution and ultrafiltration, essential to avoid instrument problems like column clogging or MS system contamination, were carried out.

In order to obtain the greatest number of compounds from saffron samples, the nature of extraction solvent was considered during the extraction procedure. As the saffron compounds can have varied polarities, different extraction solvents were investigated using different mixtures between ethanol (0, 50 and 100 %) and aqueous buffer at two different pHs (low pH at 2.5 and high pH at 9.0). Also, chromatographic separation was investigated using different stationary phases, two reversed phases (C18 and CN) and one polar phase (HILIC), to allow separation of a range of compounds of different polarities, from low to high polarity. In addition, generic HPLC methods (see section 2.4) were applied to cover a wide range of metabolites with diverse chemical and physical properties. Finally, MS analysis was done in both positive and negative ionization modes to ensure that metabolites extracted from saffron samples amenable to either positive or negative ionization were covered, in order to monitor as many ions as possible.

To evaluate the different analytical approaches, the total number of possible components (molecular features) present in an authentic and other suspicious saffron sample was compared when the different conditions described before were used in the MS analysis for both positive and negative ionization modes (see **Table 1**). Initially, ethanol and aqueous

buffer at different pH (2.5 and 9.0) were investigated as extraction solvents to increase the efficiency of saffron components extraction. As shown in **Table 1**, ethanol and borate buffer at pH 9.0 worked better, so the complementarity of both was tested. A substantial increase in the number of features was obtained with the mixture ethanol and borate buffer at pH 9.0 (50:50 v/v). On the other hand, the information obtained using the reversed phase columns (chromatographic separation and features) was more useful than data obtained using the HILIC column. Between the reversed phase columns, the C18 was the best option due mainly to better resolution between features, since the number of molecular features extracted was only increased by 110%, being this number between 30-60% less in ESI- than ESI+ (see **Table 1**). In conclusion, the mixture ethanol and borate buffer at pH 9.0 (50:50 v/v) as extraction solvent and a C18 column were chosen for further studies, using MS analysis in both positive and negative ion modes to expand the possibilities of discovering new markers.

On the other hand, the influence of extraction time in sonic bath was also investigated. Samples were extracted for 5, 10, 15 and 20 min. Similar results were obtained with the extraction time of 15 and 20 min, therefore 15 min was chosen as the optimum extraction time (results not shown).

Finally, other ESI parameters were also studied: i) only depending on analytes as nozzle voltage (0-1000 V) and fragmentator voltage (100-200 V), and ii) depending on mobile phase flow-rate and composition but also limited by analytes thermal stability: drying gas temperature (200-350 ºC) and sheath gas temperature (250-400 ºC). The optimized ESI parameters (results not shown) obtained with the previously selected mobile-phase composition and a flow-rate of 0.4 mL/min were: nozzle voltage, 0 V; fragmentator voltage 175 V; drying gas flow-rate, 10 L/min; drying gas temperature, 350 ºC; sheath gas flow-rate, 7.5 L/min, and sheath gas temperature, 350 ºC.

Table 1. Total number of possible components (molecular features) present in an authentic and other suspicious saffron samples obtained by MFE software when different extraction solvents for sample preparation and different chromatographic columns were used in the MS analysis for both positive and negative ionization modes. Experimental conditions as in section 2.4.

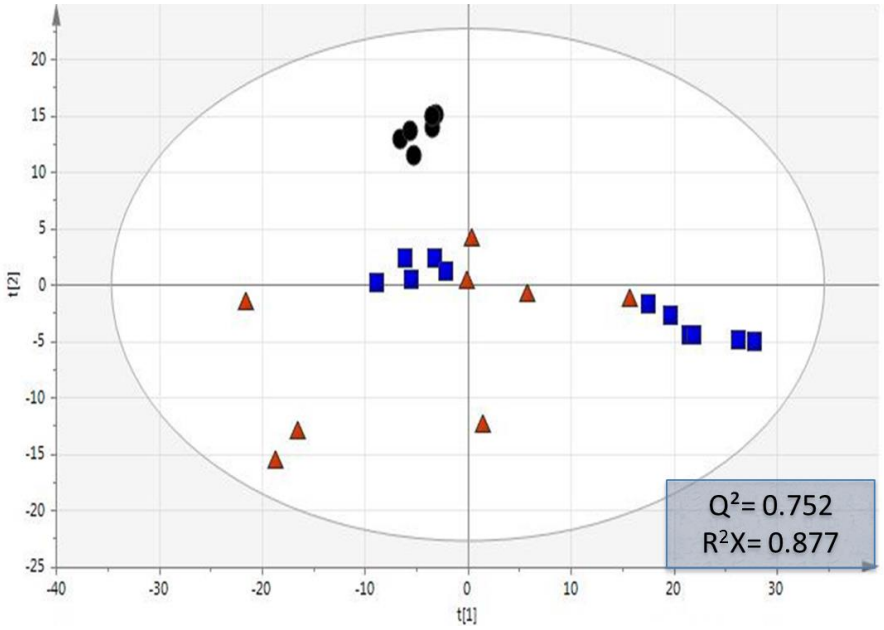
C18								
	Phosphoric buffer pH2.5		Borate buffer pH 9.0		Ethanol		Ethanol/borate buffer pH 9.0 (50:50 v/v)	
SAMPLES	ESI+	ESI-	ESI+	ESI-	ESI+	ESI-	ESI+	ESI-
Authentic	9663	6764	9975	6987	10056	6164	14673	7588
Suspicious	8438	4586	9380	3996	1223	648	10304	4578
CN								
	Phosphoric buffer pH2.5		Borate buffer pH 9.0		Ethanol		Ethanol/borate buffer pH 9.0 (50:50 v/v)	
SAMPLES	ESI+	ESI-	ESI+	ESI-	ESI+	ESI-	ESI+	ESI-
Authentic	8857	5677	9344	6254	9754	5632	12456	7040
Suspicious	8336	4245	9055	3327	1275	543	9687	4176
HILIC								
	Phosphoric buffer pH2.5		Borate buffer pH 9.0		Ethanol		Ethanol:borate buffer pH 9.0 (50:50 v/v)	
SAMPLES	ESI+	ESI-	ESI+	ESI-	ESI+	ESI-	ESI+	ESI-
Authentic	3452	1085	3654	1132	3987	1129	4967	1235
Suspicious	2968	354	3088	376	1234	166	4674	476

3.2 Chemometric analysis

Chromatograms from all saffron samples and QCs using positive and negative ionization modes were aligned, revealing a number of features in total of 25645 and 17153, respectively. The PCA represents a highly useful tool when interpreting complex multivariate data sets, as it allows dimensionality reduction and visualization of the information present in the original data in the form of a few principal components while retaining the maximum possible variability [33]. As shown in **Figure 1**, score plots from unsupervised PCA of the data set representing all saffron samples and QC samples showed more pronounced clustering and significantly better differentiation among sample clusters for positive ionization data (see **Figure 1A**) compared to those acquired in the negative mode (see **Figure 1B**). Therefore, only data acquired in positive ionization mode were further used in this statistical evaluation with univariate and multivariate methods, as it was found fit for purpose, in addition as

discussed above, the ESI+ gave almost a 200% increase in molecular features compared to the ESI-.

(A)



(B)

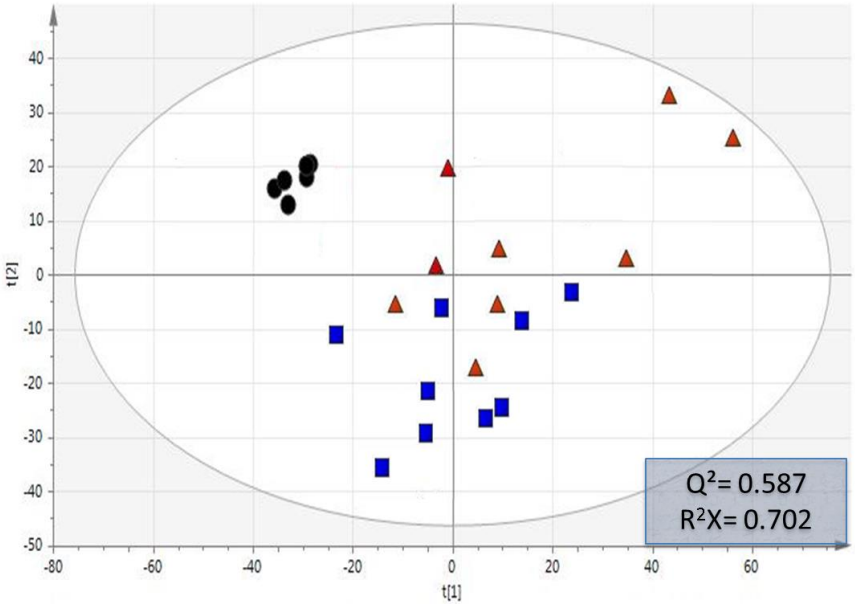


Figure 1. PCA score plot obtained in positive (A) and negative (B) ionization mode data from saffron metabolic profiles obtained for authentic samples, samples suspected of adulteration, and QC samples**.** ▲ Suspicious samples. ■ Authentic samples. ● QC samples.

For quality checking, a PLS-DA model of positive ionization data was built for the two groups of authentic and suspicious samples, taking into account all variables generated

from the mass spectra. As it can be observed in **Figure 2**, even without any filtering or scaling, the samples were clustered clearly, and the quality of the model built for two components was very good with excellent values for variance explained (R^2_X = 0.990, and R^2_Y = 0.986) and variance predicted (Q^2 = 0.975). The robustness of the analytical procedure was tested by prediction of the QC samples in the model, and proved by the tight clustering of the QC samples in the middle of the plot (see **Figure 2**). Since QC samples were obtained by mixing equal volumes of all the samples, the model proved that separation between groups was not random but due to real variability. In addition, the fact that QCs clustered together proved the stability and repeatability of the methodology, confirming that clustering (or separation) was due to the sample content and not to the analytical conditions and therefore, data are suitable for further statistical analysis.

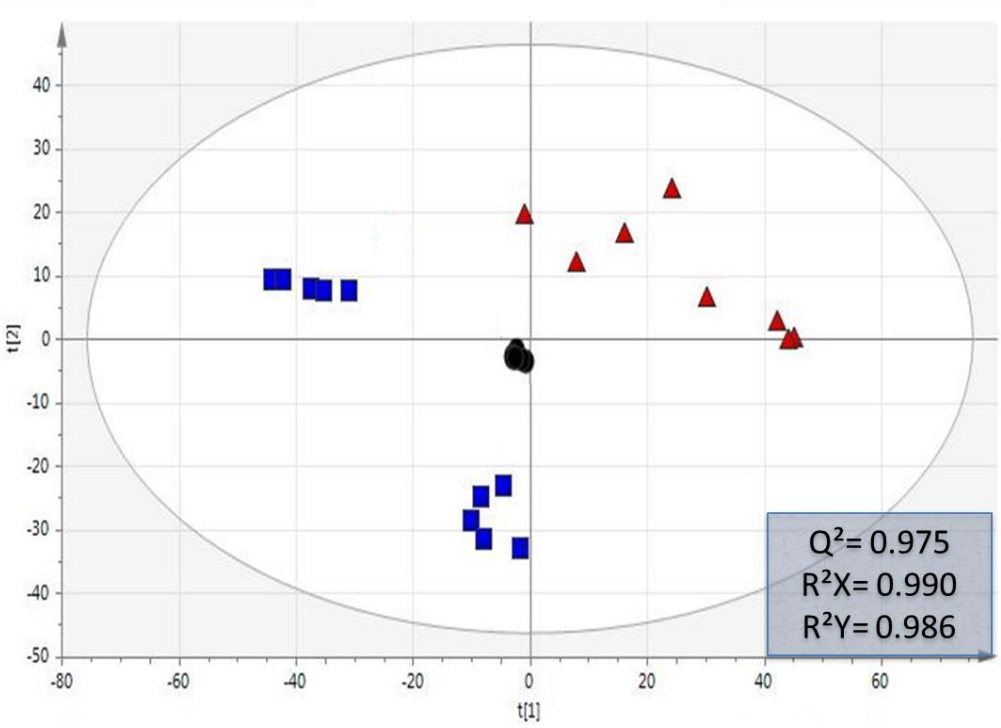


Figure 2. Score plot for a PLS-DA model built with the whole data of saffron metabolic profiles obtained for authentic samples and samples suspected of adulteration with prediction for QC samples. ▲ Suspicious samples. ■ Authentic samples. ● QC samples.

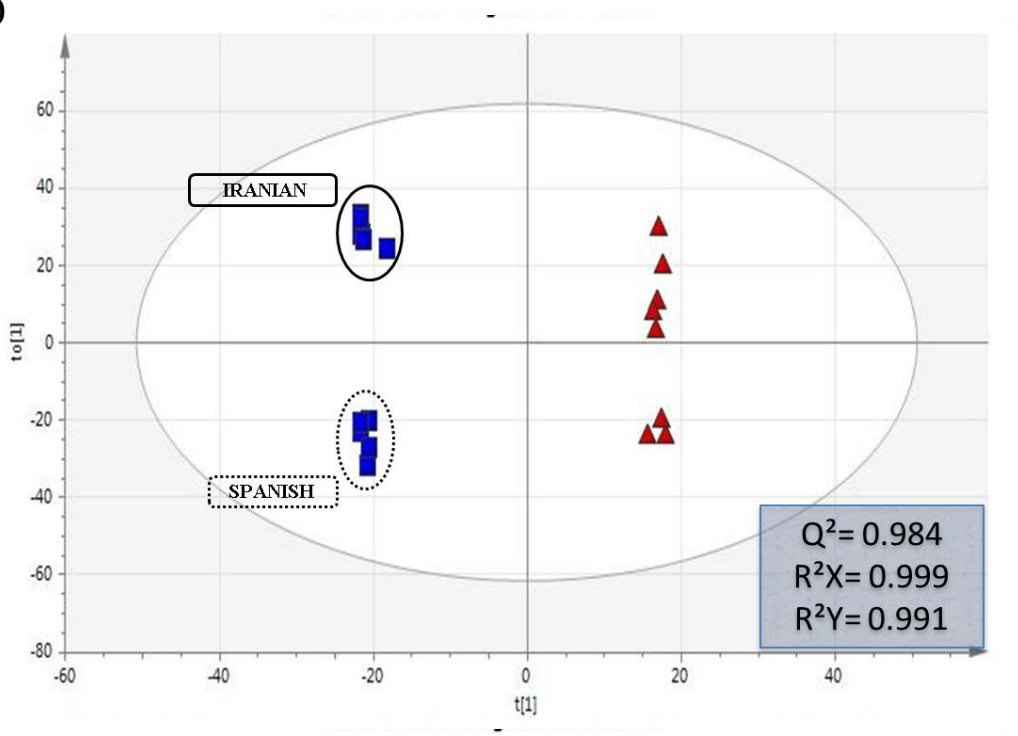
OPLS-DA model, which comes from PLS-DA model, was employed to discriminate between groups of samples and to obtain the significance of the discriminatory compounds by

using jack-knife test. In OPLS-DA a regression model is calculated between the multivariate data and a response variable that only contains class information. The advantage of OPLS-DA compared to PLS-DA is that a single component is used as a predictor for the class, while the other components describe the variation orthogonal to the first predictive component [34-36]. To perform the OPLS-DA model for saffron samples classification, chromatograms from every saffron sample (authentic samples and samples suspected of adulteration) were aligned with the same parameters described in the *Material and Methods* section. First, features were filtered by choosing the data present in 100 % of QC samples and present in at least one of two groups for comparing authentic vs. suspicious. 1126 features (out of 24768) were selected after primary filtering and chosen for further filtering. In total, 490 different features fulfilled the requirements of a secondary filtering performed by choosing the data with a coefficient of variation below 35%.

The OPLS-DA model with two groups (authentic and suspicious) was built with those 490 features that fulfilled the filtering requirements and using pareto scaling [37]. As shown in **Figure 3A**, the quality of the model built was excellent regarding variance explained (R^2_X = 0.999, and R^2_Y = 0.991) and variance predicted (Q^2 = 0.984). In addition, the large orthogonal variation divided the authentic samples into two groups, both clearly differentiated from suspicious samples. Reviewing the origin of authentic samples, it was observed that each group corresponded to Spanish and Iranian saffron. Therefore, new OPLS-DA model was performed according to the geographical origin of the authentic samples (see **Figure 3B**). In that case, suspicious samples were excluded from the model. The same parameters for alignment and filtering were employed to perform the OPLS-DA model for origin saffron classification. In this case, 1127 features (out of 19351) were selected after primary filtering and chosen for further filtering. In total, 489 different features fulfilled the requirements of a

secondary filtering performed. Again, the quality of the model built was excellent regarding variance explained (R^2_X = 0.998, and R^2_Y = 0.989) and variance predicted (Q^2 = 0.982).

(A)



(B)

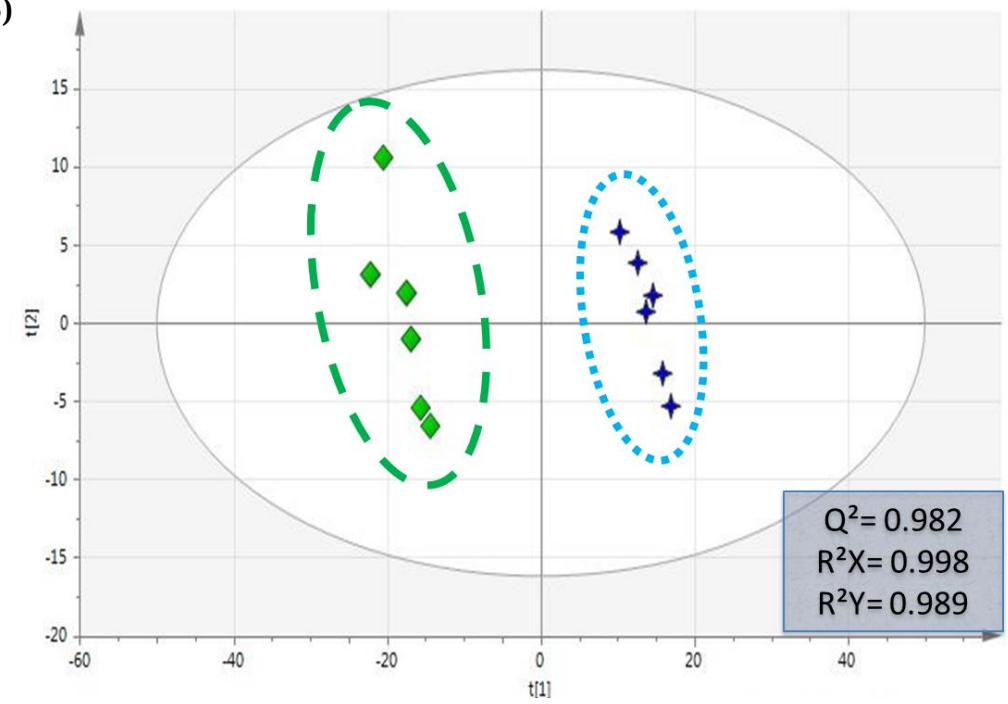


Figure 3. OPLS-DA score plot of saffron metabolic profiles obtained for: (A) authentic samples and samples suspected of adulteration, and (B) authentic saffron samples from Spain and Iran.

▲ Suspicious samples. ■ Authentic samples. ♦ Spanish authentic samples. + Iranian authentic samples.

In conclusion, separations by authenticity (authentic and suspicious) and by origin (Spain and Iran) of saffron samples were possible. In addition, to avoid the risk of over fitting, both models (authenticity and origin) were validated by cross-validation tool as described in the *Materials and Methods* section, and the percentages of samples correctly classified for models (authenticity and origin) were tested [38]. In the validation of authenticity model, all the authentic samples were 100 % predicted with appropriate classification, and from suspicious samples only one out of 10 samples was misclassified. In the validation of origin model, all the Iranian samples were 100 % predicted with appropriate classification, and from Spanish samples only one out of 5 samples was misclassified enabling these results to be considered satisfactory.

Finally, all potential unique and high intensity markers giving a significant contribution were selected with unpaired "Welch's test" performed for log-transformed intensities and statistical significance by s-plot and jack-knifed confidence intervals of both models. When the interval includes the zero value, the covariance is not significant and the compound should not be considered as a potential marker (see **Figure 4**). For *authentic vs. suspicious* (see **Figure 4A**) and *spanish vs. iranian* (see **Figure 4B**) comparisons, a total of 84 and 29 metabolites (with p-value < 0.05) were obtained as candidates for markers of authenticity and origin, respectively. It is noteworthy for *authentic vs. suspicious* comparison that all significant markers obtained by OPLS-DA model were up regulated in authentic group according to the statistical significance of these metabolites using the jack-knife interval (see **Figure 4A)**. This fact implies that the suspicious samples were not adulterated with common compounds that may serve for discriminant analysis by OPLD-DA and therefore the significant markers obtained were compounds with higher content in authentic saffron, i.e, they are markers of authenticity. However, for the *spanish vs. iranian*

comparison, the significant markers obtained by OPLS-DA model were regulated in both groups according to the statistical significance of these metabolites using the jack-knife interval (see **Figure 4B)**, that is, there are both significant markers for Spanish samples and for Iranian samples.

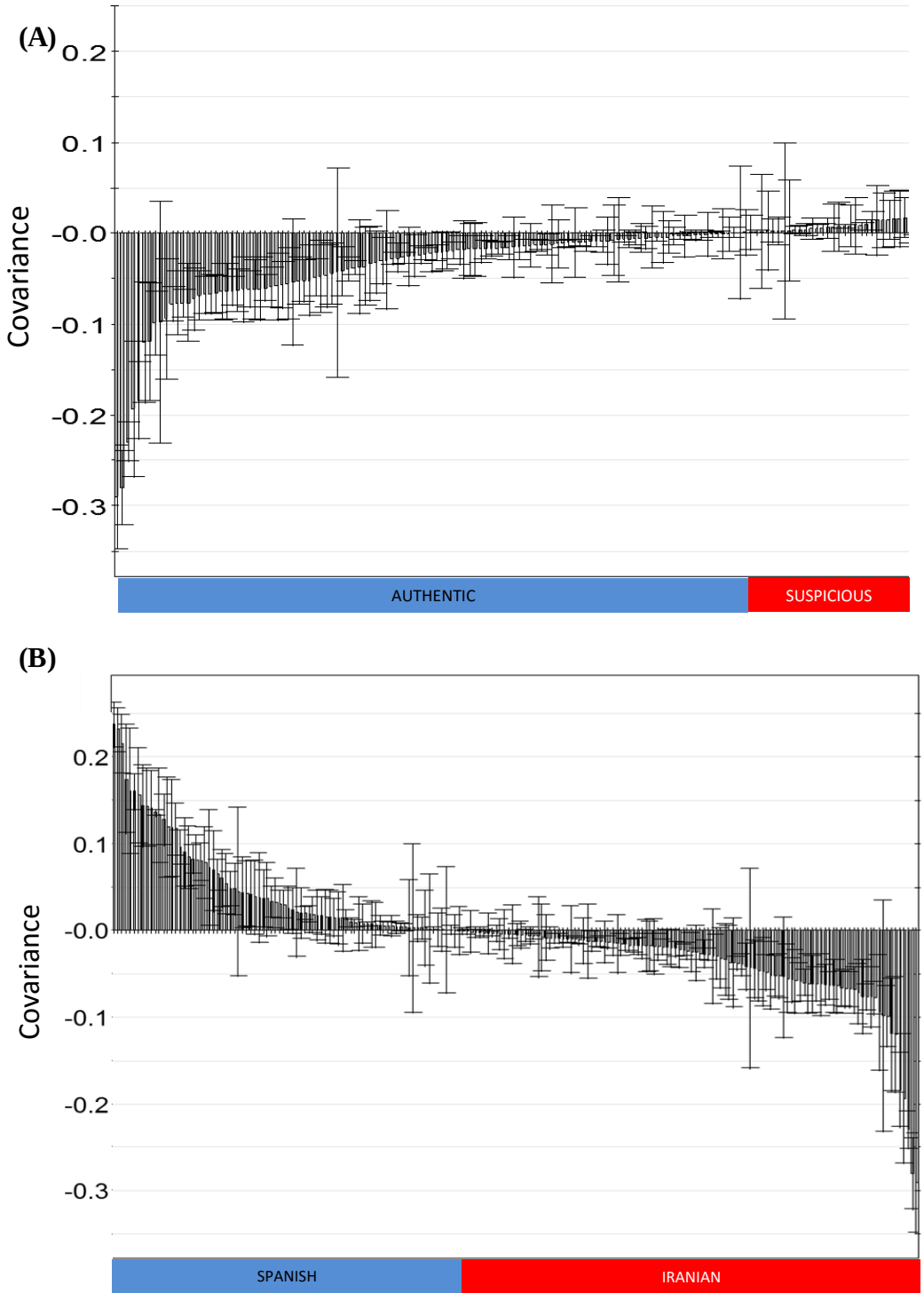


Figure 4. Covariance for discriminant variables including jack-knife interval (p > 95%) in (A) authentic samples and samples suspected of adulteration, and (B) authentic saffron samples from Spain and Iran.

3.3 Identification of marker compounds

The identification of discriminating marker compounds represents probably the most laborious and time-consuming step of the metabolomic workflow. The accurate mass information was used to propose the elemental formula of each marker using a database search (see 2.6 section). The search for the elemental formula can be complicated because multiple hits can be matched with the same exact mass. For reliable elemental formula estimation, the suggested formulas are filtered based on matching of experimental and theoretical isotopic profile in terms of relative intensities. Finally, out of the numerous compound hits obtained, only those whose presence was probable in plants were taken into account (e.g., hits of drugs and synthetic compounds obtained in databases were cleared). Using this procedure, 34 of 113 marker compounds (from both models, authenticity and origin) were tentatively identified. It should be highlighted that no marker related to origin model was tentatively identified. Therefore, the 34 possible metabolites tentatively identified were all related to the authenticity model, i.e, they are authenticity markers of saffron.

Table 2 reports all significant metabolites tentatively identified detected as candidates for authenticity markers of saffron and includes their retention time, the *m/z* value obtained in the LC-TOF system and the type of ion responsible for it, the calculated mass error when comparing with the database, the probable ion elemental formula, the percentage change among the groups and its statistical significance (p-value), and their coefficient of variance according to QC. All suggested formulas have very good mass accuracy, usually less than 2 mDa, thus increasing the confidence in predicted formulas.

It should be noted that a mass match with a metabolite in the database does not provide a conclusive identification. The confirmatory analysis of a predicted compound standard is generally required for full and unequivocal identity confirmation by comparison of

Table 2. Overview of significant metabolites (tentatively identified) detected as candidates to be authenticity markers in saffron.

Tentative identification	RT (min)	Experimental m/z	Ion	Mass error (ppm)	Mass error (mDa)	Probable ion elemental formula	Fold change (p-value)	CV for QCs [%]
Kaempferol 3,7,4'-triglucoside	4.8	773.2134	$[M+H]^+$	-0.1	-0.1	$C_{33}H_{41}O_{21}$	37 (0.007)	9
Kaempferol 3-sophorotrioside			$[M+H]^+$			$C_{33}H_{41}O_{21}$		
Kaempferol 3-sophoroside-7-glucoside			$[M+H]^+$			$C_{33}H_{41}O_{21}$		
Kaempferol 3-glucoside-7-sophoroside			$[M+H]^+$			$C_{33}H_{41}O_{21}$		
Cyanidin 3,5,3'-triglucoside			M^+			$C_{33}H_{41}O_{21}$		
Delphinidin 3-rutinoside-5-glucoside			M^+			$C_{33}H_{41}O_{21}$		
Nepetalic acid	4.9	185.1176	$[M+H]^+$	5.9	0.4	$C_{10}H_{17}O_3$	6 (0.005)	7
Kaempferol 3,7,4'-triglucoside	5.9	773.2142	$[M+H]^+$	1.0	0.8	$C_{33}H_{41}O_{21}$	24 (0.01)	12
Kaempferol 3-sophorotrioside			$[M+H]^+$			$C_{33}H_{41}O_{21}$		
Kaempferol 3-sophoroside-7-glucoside			$[M+H]^+$			$C_{33}H_{41}O_{21}$		
Kaempferol 3-glucoside-7-sophoroside			$[M+H]^+$			$C_{33}H_{41}O_{21}$		
Cyanidin 3,5,3'-triglucoside			M^+			$C_{33}H_{41}O_{21}$		
Delphinidin 3-rutinoside-5-glucoside			M^+			$C_{33}H_{41}O_{21}$		
Kaempferol 3,7-diglucoside	6.3	611.1601	$[M+H]^+$	-0.9	-0.6	$C_{27}H_{31}O_{16}$	27 (0.001)	6
Kaempferol 3-sophoroside			$[M+H]^+$			$C_{27}H_{31}O_{16}$		
Quercetin 3-rutinoside (Rutin)			$[M+H]^+$			$C_{27}H_{31}O_{16}$		
Cyanidin 3,5-diglucoside (Cyanin)			M^+			$C_{27}H_{31}O_{16}$		
Cyanidin 3-sophoroside			M^+			$C_{27}H_{31}O_{16}$		
Geranic acid	6.4	169.1448	$[M+H]^+$	5.3	0.9	$C_{10}H_{17}O_2$	21 (0.002)	5
Dihydrojasmonic acid, Methyl ester	7.5	227.1636	$[M+H]^+$	-1.3	-0.3	$C_{13}H_{23}O_3$	13 (0.009)	10

Table 2.Continued.

Tentative identification	RT (min)	Experimental m/z	Ion	Mass error (ppm)	Mass error (mDa)	Probable ion elemental formula	Fold change (p-value)	CV for QCs [%]
Kaempferol 3,7-diglucoside	7.7	611.1604	$[M+H]^+$	-0.4	-0.3	$C_{27}H_{31}O_{16}$	16 (0.003)	8
Kaempferol 3-sophoroside			$[M+H]^+$			$C_{27}H_{31}O_{16}$		
Quercetin 3-rutinoside (Rutin)			$[M+H]^+$			$C_{27}H_{31}O_{16}$		
Cyanidin 3,5-diglucoside (Cyanin)			M^+			$C_{27}H_{31}O_{16}$		
Cyanidin 3-sophoroside			M^+			$C_{27}H_{31}O_{16}$		
Kaempferol 3-glucoside	8.6	449.1050	$[M+H]^+$	-2.0	-0.9	$C_{21}H_{21}O_{11}$	29 (0.004)	4
Quercetin 3-rhamnoside (Quercitrin)			$[M+H]^+$			$C_{21}H_{21}O_{11}$		
Angoluvarin	9.8	485.1991	$[M+H]^+$	2.6	1.3	$C_{30}H_{29}O_6$	4 (0.009)	15
Isococculidine	9.9	286.1792	$[M+H]^+$	8.1	2.3	$C_{18}H_{24}NO_2$	11 (0.005)	13
Isobrucein A	9.9	523.2172	$[M+H]^+$	-0.4	-0.2	$C_{26}H_{35}O_{11}$	4 (0.007)	9
Apimaysin	9.9	561.1618	$[M+H]^+$	2.1	1.2	$C_{27}H_{29}O_{13}$	9 (0.008)	11
4,2-Dihydroxy-4,6-dimethoxychalcone 4-apiosyl-glucoside	10.1	595.2038	$[M+H]^+$	2.5	1.5	$C_{28}H_{35}O_{14}$	9 (0.03)	20
Eriojaposide B	10.4	517.2631	$[M+H]^+$	-3.5	-1.8	$C_{25}H_{41}O_{11}$	9 (0.04)	18
Cinncassiol-glucoside	10.4	531.2810	$[M+H]^+$	1.9	1.0	$C_{26}H_{43}O_{11}$	9 (0.006)	16
Kaempferide 3,7-dirhamnoside	10.6	593.1859	$[M+H]^+$	-1.0	-0.6	$C_{28}H_{33}O_{14}$	12 (0.009)	9
5-Hydroxypseudobaptigenin 7-O-Glucoside	11.3	461.1085	$[M+H]^+$	1.3	0.6	$C_{22}H_{21}O_{11}$	7 (0.007)	17
Karatavigenin B	11.4	569.3478	$[M+H]^+$	0.9	0.5	$C_{34}H_{49}O_7$	15 (0.003)	18
Anhalonidine	11.6	224.1285	$[M+H]^+$	1.8	0.4	$C_{12}H_{18}NO_3$	11 (0.007)	12
Nupharamine	11.6	252.1231	$[M+H]^+$	0.0	0.0	$C_{13}H_{18}NO_4$	7 (0.04)	15

Tentative identification	RT (min)	Experimental m/z	Ion	Mass error (ppm)	Mass error (mDa)	Probable ion elemental formula	Fold change (p-value)	CV for QCs [%]
Resokaempferol	11.8	271.0605	$[M+H]^+$	1.5	0.4	$C_{15}H_{11}O_5$	21 (0.007)	9
Baicalein			$[M+H]^+$			$C_{15}H_{11}O_5$		
Genistein			$[M+H]^+$			$C_{15}H_{11}O_5$		
Dihydrojasmonic acid	12.4	213.1491	$[M+H]^+$	2.3	0.5	$C_{12}H_{21}O_3$	8 (0.004)	10
Aconine	12.8	500.2834	$[M+H]^+$	4.2	2.1	$C_{25}H_{42}NO_9$	10 (0.009)	8
3-Hydroxyethylbacteriochlorophyllide A	12.9	635.2700	$[M+H]^+$	1.6	1.0	$C_{35}H_{39}MgN_4O_6$	6 (0.04)	13
4,6,8-Megastigmatriene	13.7	177.1645	$[M+H]^+$	3.9	0.7	$C_{13}H_{21}$	3 (0.0007)	6
1-O-beta-D-Glucopyranose	13.9	361.1984	$[M+H]^+$	-7.5	-2.7	$C_{17}H_{29}O_5$	26 (0.005)	10
28-Hydroxyglycyrrhetic acid	14.4	487.3394	$[M+H]^+$	-4.9	-2.4	$C_{30}H_{47}O_5$	9 (0.007)	13
Cinnamylisovalerate	14.5	219.1377	$[M+H]^+$	-1.4	-0.3	$C_{14}H_{19}O_2$	7 (0.004)	9
Fissinolide	14.5	513.2438	$[M+H]^+$	8.0	4.1	$C_{29}H_{37}O_8$	14 (0.03)	12
15,16-Dihydrobiliverdin	14.5	585.2666	$[M+H]^+$	-4.7	-2.8	$C_{32}H_{41}O_{10}$	6 (0.007)	10
Octadecanedioic acid	18.4	315.2504	$[M+H]^+$	7.3	2.3	$C_{18}H_{35}O_4$	6 (0.007)	11
Hexafluoro-25-hydroxycholecalciferol	26.5	509.2824	$[M+H]^+$	-4.9	-2.5	$C_{27}H_{39}F_6O_2$	7 (0.004)	9
1-octadecatrienoyl-2-octadecatetraenosyl-glycero-3-phosphate	29.5	691.4350	$[M+H]^+$	2.5	1.7	$C_{39}H_{63}O_8P$	11 (0.01)	9
4-dimethylaminophenyl-25 dihydroxycholecalciferol	30.9	536.4104	$[M+H]^+$	0.9	0.5	$C_{35}H_{53}NO_3$	3 (0.04)	15

retention time, accurate mass data and isotopic profile of commercially available reagents with those obtained in real samples. However, metabolites frequently have very complex chemical structures and may be difficult to synthetize, so commercial standards are not available or have huge prices.

Many of the compounds grouped in **Table 2** are flavonols and anthocyanins (molecules responsible of color) and substances responsible of flavor. This fact is in concordance with the main expected differences in a saffron sample adulterated, since the objective of adulteration is to mask saffron properties such as color and flavor. However, in several cases different marker compounds were matched with the same elemental formula. Note especially those having a retention time of 4.8, 5.9, 6.3, 7.7, and 8.6 min, corresponding to compounds naturally conjugated with sugars (glycosides) with three, two and one hexoses, respectively. Due to the lack of supplier or huge prices of standards of the majority of these glycosides, they were confirmed by studies focused on MS and MS/MS analysis in the positive and negative modes (see **Table 3**), and comparing experimental MS/MS fragmentation patterns with those provided in scientific literature.

According to **Table 3**, the tentatively identified compound as delphinidin 3-rutinoside-5-glucoside that could elute at retention time 4.8 or 5.9 min was discarded due to the absence of an ion in positive mode with *m/z* 303.0499 which should appear as M^+ for its aglycone delphinidin. Likewise, the quercetin glycosides (rutin at 6.3 or 7.7 min, and quercitrin at 8.6 min) were discarded because in all cases signals with *m/z* about 287 (287.0532, 287.0537 or 287.0527) were obtained in positive mode instead of an ion with *m/z* 303.0499 corresponding to its protonated aglycone quercetin. In addition, the standards of the rutin and quercitrin were analyzed for further confirmation of their absence by their retention times and MS/MS fragmentation data.

Table 3. MS and MS/MS data (in positive and negative ESI modes) for several compounds tentatively identified in Table 2 as glycosides.

RT (min)	ESI +		ESI -	
	MS ions *m/z* (%)[a]	**MS/MS[b] ions *m/z* (%)**	**MS ions *m/z* (%)**	**MS/MS ions *m/z* (%)**
4.8	773.2076 (...) $[M+H]^+$ 611.1546 (...) $[M-162+H]^+$	611.1543 (...) $[M-162+H]^+$ 449.1049 (...) $[M-324+H]^+$ 287.0531 (...) $[M-486+H]^+$	771.1811 (...) $[M-H]^-$ 609.1387 (...) $[M-162-H]^-$	609.1386 (...) $[M-162-H]^-$ 447.0854 (...) $[M-324-H]^-$ 284.0263 (...) $[M-487-H]^-$
5.9	773.2081 (...) $[M+H]^+$ 611.1541 (...) $[M-162+H]^+$	611.1539 (...) $[M-162+H]^+$ 449.1051 (...) $[M-324+H]^+$ 287.0529 (...) $[M-486+H]^+$	771.1880 (...) $[M-H]^-$ 609.1374 (...) $[M-162-H]^-$	609.1375 (...) $[M-162-H]^-$ 429.0736 (...) $[M-342-H]^-$ 284.0258 (...) $[M-487-H]^-$
6.3	611.1543 (...) $[M+H]^+$	449.1056 (...) $[M-162+H]^+$ 287.0532 (...) $[M-324+H]^+$	609.1371 (...) $[M-H]^-$	447.0824 (...) $[M-162-H]^-$ 285.0331 (...) $[M-324-H]^-$ 284.0253 (...) $[M-325-H]^-$
7.7	611.1538 (...) $[M+H]^+$	449.1054 (...) $[M-162+H]^+$ 287.0537 (...) $[M-324+H]^+$	609.1345 (...) $[M-H]^-$	429.0729 (...) $[M-180-H]^-$ 284.0259 (...) $[M-325-H]^-$
8.6	449.1057 (...) $[M+H]^+$	287.0527 (...) $[M-162+H]^+$	447.0752 (...) $[M-H]^-$	284.0250 (...) $[M-163-H]^-$

[a] Percentage of signal intensity (in parentheses)

[b] MS/MS experiments using as precursor ion the quasimolecular ion of MS experiments ($[M+H]^+$ or $[M-H]^-$)

The other anthocyanins tentatively identified that eluted at retention times 4.8, 5.9, 6.3 and 7.7 min were also discarded because although signals with *m/z* about 287 (287.0531, 287.0529, 287.0532 or 287.0537) appeared in the positive mode as M^+ for their aglycone cyanidin, the ion at *m/z* 286.0483 as $[M-H]^-$ was absent in the negative mode. Therefore, the ions at *m/z* about 287 in positive mode corresponded to a protonated aglycone of kaempferol. These results confirmed the well-known fact that kaempferol glycosides represent 84.0% of total flavonol content in flowers of *Crocus sativus* L. whereas quercetin and isorhamnetin glycosides represent only 9.3 and 8.7%, respectively [39, 40].

The tentative identification of the different glycosides of kaempferol listed as significant authenticity markers of saffron in **Table 2** was confirmed on the basis of MS/MS fragmentation patterns by comparison with previously published data. These studies were focused on MS analysis in the negative mode (see **Table 3**) because this mode is more sensitive than positive mode to establish the differences between interglucosidic linkages (see MS/MS fragments in Table 3 for compounds that eluted at 4.8 and 5.9 min, and for compounds that eluted at 6.3 and 7.7 min), and positional isomers of O-glycosylated kaempferols [41]. Thus, the identification of these markers was achieved in a similar manner, and therefore, the discussion was focused on two characteristic examples of the five possible markers, the kaempferols glycosylated with two hexose residues detected at 6.3 min and 7.7 min.

To differentiate kaempferols with the same degree of glycosylation, the characterization of the (glucosyl(1 → 2)glucosides) interglycosidic linkage was defined by the presence at 7.7 min of the fragment ion $[M-180-H]^-$ characteristic of sophoroside flavonoids (at *m/z* 429.0729 formed from the loss of the terminal sugar) although at low abundance (see **Figure 5A**), which is absent in a gentiobioside flavonoid (glucosyl (1 → 6) glucosides) and in a kaempferol di-O-glucoside with two sugar moieties linked to different

phenolic hydroxyl positions of the kaempferol nucleus. In addition, the MS fragmentation pattern of kaempferol 3-O-diglycosides was detected by the presence of the ion [M-162-163-H]$^-$ (*m/z* 284.0259) as base peak, ion obtained as a result of the loss of two sugar moieties. The presence of a fragment ion at *m/z* 284.0259 corresponding to a kaempferol moiety, instead of the expected ion at *m/*z 285 (loss of two glucosyl units, [M-162-162-H]$^-$), was recently described as a characteristic fragment ion for the kaempferol 3-O-glucoside ([M-163-H]$^-$), but not in the case of kaempferol 7-O-glucoside [~~43~~ 42]. All these data (see **Table 3** and **Figure 5A**), according to previously published data [41], confirmed the structure of *kaempferol 3-O-sophoroside* as the compound eluting at 7.7 min, the main saffron flavonoid (about 55% of total flavonols content), whose presence was reported previously in *Crocus sativus* in several works [40, 42-46]. In addition, Carmona et al [47] stated that saffron samples from different geographical origin were clearly separated by their kaempferol 3-sophoroside contents. However, according to the results obtained in this work, kaempferol 3-O-sophoroside was not detected as a significant marker of the origin of saffron because its interval of covariance includes zero value according to the jack-knife interval.

On the other hand, the MS/MS spectrum of the [M-H]$^-$ ion (*m/z* 609.1317) provided the ion [M-162-H]$^-$ (*m/z* 447.0824) as base peak (see **Figure 5B**), which was formed by a loss of one glucose, and the ion [M-162-162-H]$^-$ (*m/z* 285.0331) formed by the loss of two glucoses. However, the characteristic ion [M-180-H]$^-$ of the sophoroside of kaempferol was not detected. These results confirmed that it is a kaempferol di-O-glucoside with sugar moieties linked to different phenolic hydroxyl positions of the kaempferol nucleus. This sugar substitution takes place at the hydroxyls in the following order of preference (from more to less) at the 3-, 7- and 4'- positions of the flavonoid nucleus [48]. Therefore, these data, according to previously published data [41], confirmed the structure of *kaempferol 3,7-O-*

diglucoside as the compound eluting at 6.3 min. The presence of this compound was also described in *Crocus sativus* in two recent papers [40, 42].

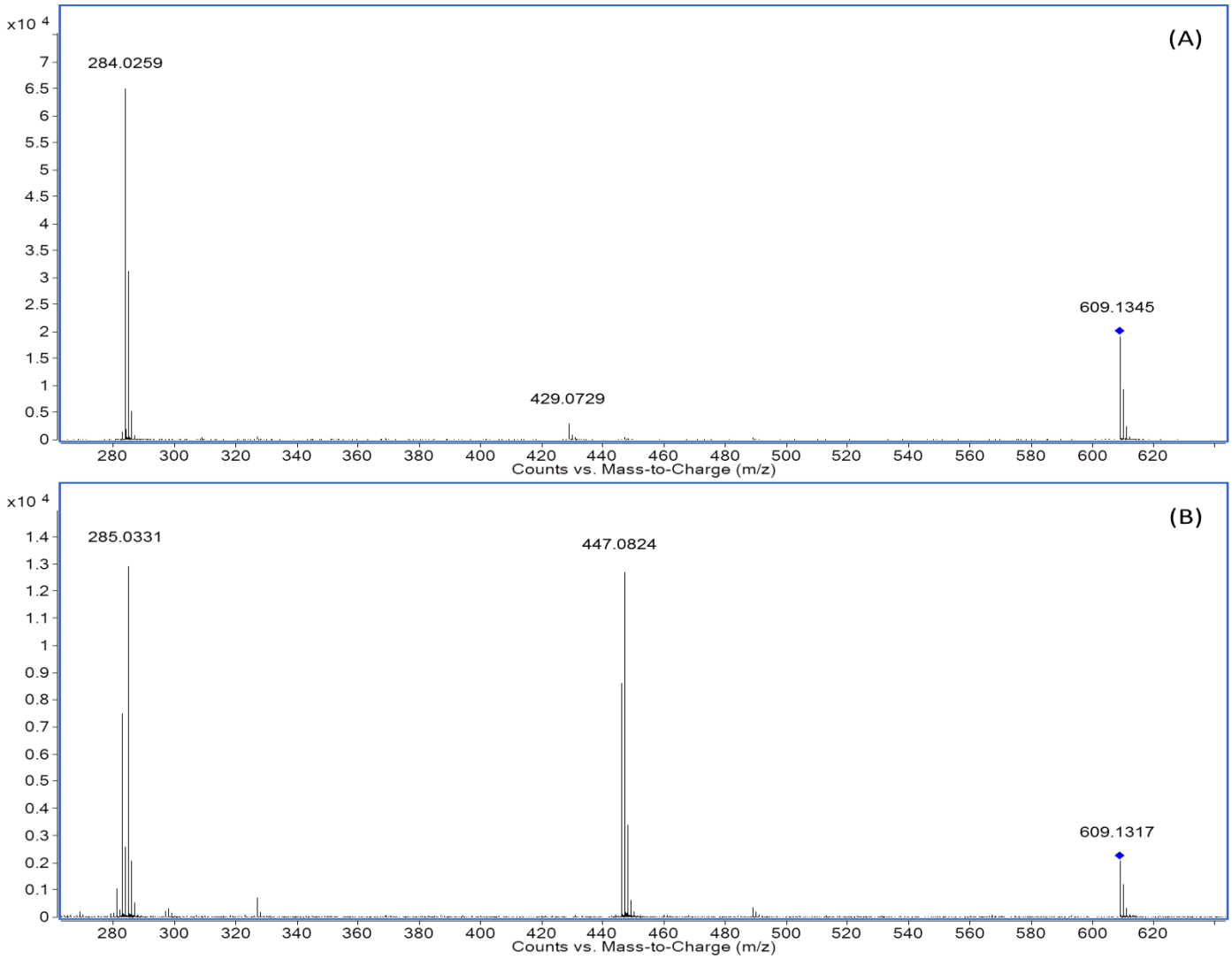


Figure 5. MS/MS spectra in ESI- for compounds eluted at 7.7 min (A) and at 6.3 min (B) using as precursor ion the deprotonated molecular ion of the kaempferols glycosylated with two hexoses, at *m/z* 609.1345 in (A) and at *m/z* 609.1317 in (B).

With respect to other glycosylated kaempferols tentatively identified, compounds eluted at retention times of 4.8, 5.9 and 8.6 min, their MS and MS/MS data are also listed in **Table 3**.

For the last of them (at 8.6 min), its MS spectrum showed the deprotonated molecular ion at *m/z* 447.0752, characteristic of a kaempferol glycosylated with only one hexose. This fact together with the presence of a MS/MS fragment ion at *m/z* 284.0250 confirmed that it was *kaempferol 3-O-glucoside*, according to that described above. This kaempferol was also identified in *Crocus sativus* in a recent work of Goupy et al [40]. In the case of the other two compounds, their MS spectra showed the deprotonated molecular ions at *m/z* 771.1811 (at 4.8 min) and 771.1880 (at 5.9 min), characteristic of kaempferols glycosylated with three hexoses. As described above, in the first case, the presence of the ion with *m/z* 447.0854 and the absence of an ion at *m/z* 429 confirmed that it was a tri-O-glucoside (i.e., *kaempferol 3,7,4´-O-triglucoside*), while for the second compound the presence of the ion with *m/z* 429.0736 and the absence of an ion at *m/z* 447 confirmed that it was the *kaempferol 3-O-sophoroside-7-O-glucoside*.

Finally, three standards were analyzed in order to confirm them as markers of authenticity, geranic acid at 6.4 min, and genistein and baicalein at 11.8 min. The latter two could not be confirmed for not matching their retention times and MS/MS spectra with the results obtained in samples of saffron, but the *Geranic acid* tentatively identified in **Table 2** as marker from $[C_{10}H_{16}O_2+H]^+$ ion (*m/z* 169.1196), was confirmed as *Geranyl-O-glucoside* after analysis of commercially available geranic acid and studies focused on MS/MS analysis. When geranic acid was analyzed its retention time was significantly higher (15.8 min) than that obtained from the sample (6.4 min), but the same MS/MS mass spectra (product ions spectra) of the precursor ion at *m/z* 169 were obtained from standard and sample (see **Figure 6A, 6B**). A detailed study of MS spectra obtained with the saffron sample at 6.4 min (see **Figure 6C**) allowed to observe an ion of small intensity at *m/z* 331.1717corresponding to glycoside of geranic acid with very good mass accuracy (less 4 mDa), along with the protonated ion of the geranic acid at *m/z* 169.1196 and two of its characteristic fragments at

m/z 151.1090 and *m/z* 123.1151. This result justifies the differences between the retention times, because as it was expected, the more polar glycoside of geranic was eluted earlier than geranic acid. Therefore, using the available experimental MS and MS/MS data it is possible to accomplish structural confirmation of *Geranyl-O-glucoside* as authenticity marker of saffron according to the fragmentation pattern shown in **Figure 7** based on the MS/MS spectra in **Figure 6**.

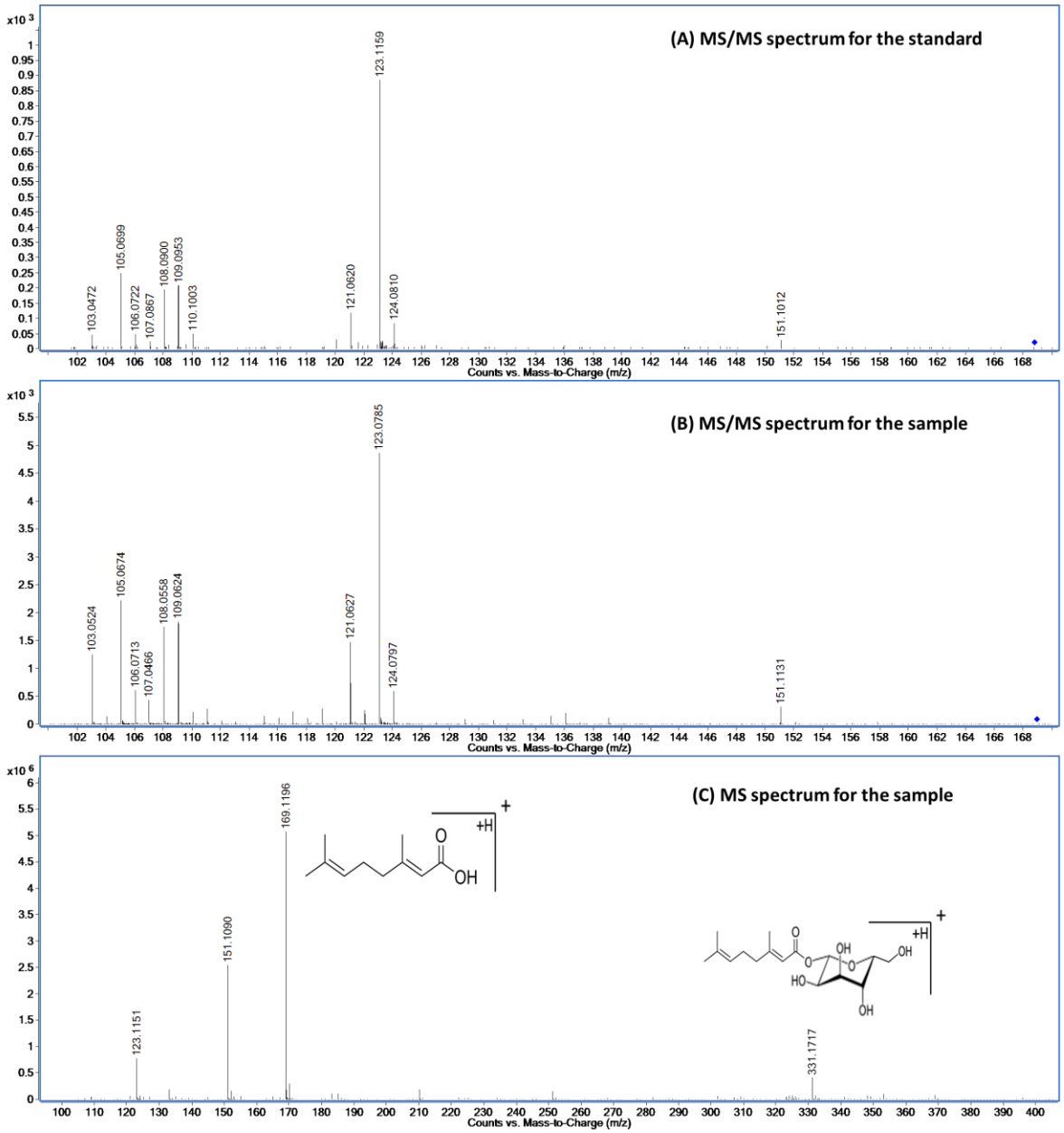


Figure 6. MS/MS spectra in ESI+ of the precursor ion with *m/z* 169.1 for geranic acid standard eluted at 15.8 min (A) and for a compound of a saffron sample eluted at 6.4 min (B). MS spectra in ESI+ for the same compound of a saffron sample eluted at 6.4 min tentatively identified as geranic acid in a saffron sample (C).

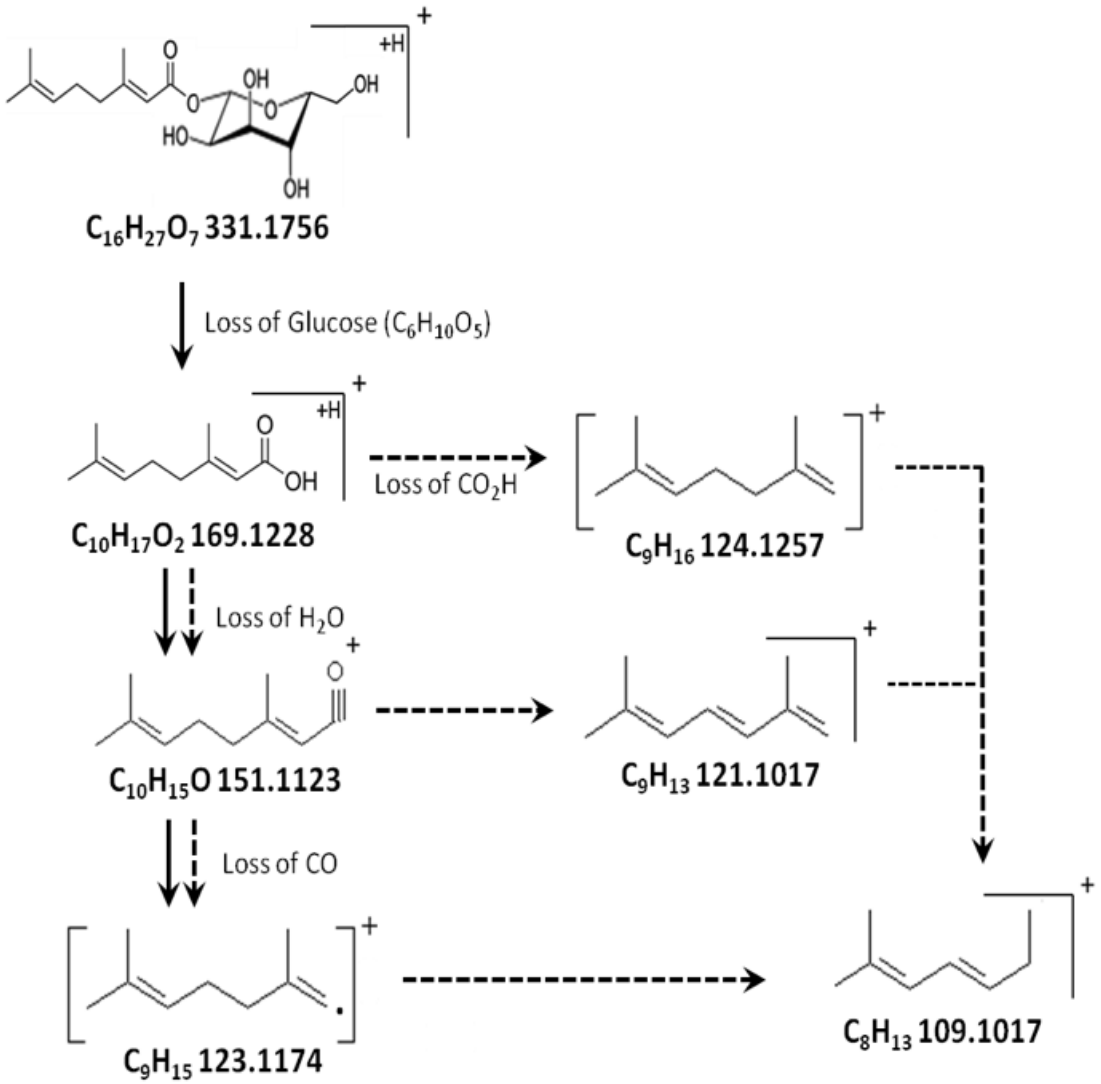


Figure 7. Fragmentation pattern for *geranyl-O-glucoside* with data related with the molecular formulas and their theoretical accurate masses.

4. Conclusions

The comprehensive and non-targeted LC–MS metabolomic fingerprinting coupled to chemometric methods was demonstrated for the first time to be a powerful tool for saffron authenticity testing. An attempt to identify novel marker compounds was carried out based on MS/MS mass spectra obtained within the LC-(QTOF)MS analysis. Using a combination of

experimental data and information available in scientific literature, and mass spectral databases, six novel metabolites related to metabolism of kaempferol and geranic acid as glycosides (*kaempferol 3-O-glucoside, kaempferol 3-O-sophoroside, kaempferol 3,7-O-diglucoside, kaempferol 3,7,4´-O-triglucoside, kaempferol 3-O-sophoroside-7-O-glucoside, and geranyl-O-glucoside*) were identified.

The fact that all possible markers obtained from the comparison between authentic *vs*. suspicious samples were regulated in authentic group implies that the suspicious samples were not adulterated with common elements which may serve as markers of adulteration. In addition, the large heterogeneity in adulterated samples increased the value of authenticity markers in saffron. These authenticity markers of saffron have the potential to be a useful tool for detecting novel adulteration practices, as more advanced and sophisticated adulteration methods are continuously developing.

However, no marker related to geographical origin was tentatively identified, so a larger set of samples with differences in origin (environmental conditions such as diverse soil types, cultivation environments, altitude, etc.) would be required to validate this model and assign an appropriate confidence level before it could be concluded the provenience of a saffron sample.

In conclusion, the results obtained in this work demonstrate that metabolomics, in conjunction with a comprehensive database, has a great potential as a screening tool for the detection of food fraud, and may be used in the future to enable a rapid reaction in the global saffron market and to help regulators to stay one step ahead of fraudsters.

Acknowledgements

Authors thank financial support from the Spanish Ministry of Science and Innovation (project CTQ2009-09022), the Comunidad of Madrid (Spain) and European funding from

FEDER program (project S2013/ABI-3028, AVANSECAL-CM), and the University of Alcalá (project UAH2011/EXP-020). M. Guijarro-Díez thanks the Ministry of Science and Innovation for his predoctoral contract (BES-2010-033465). Authors also thank the kind gift of saffron samples by Carmencita Company and the group of Prof. Coral Barbas for the one month stay of M.G.D. in her laboratory to work with the different chemometric techniques used in this work.

References

1 Melnyk JP, Wang S, Marcone MF (2010) Chemical and biological properties of the world's most expensive spice: Saffron. Food Res Int 43:1981-1989.

2 Winterhalter P, Straubinger MS (2000) Saffron-renewed interest in an ancient spice. Food Rev Int 16:39-59.

3 Hagh-Nazari S, Keifi N (2007) Saffron and various fraud manners in its production and trades. Acta Hortic 739:411-416.

4 Kanti R, Sushma K, Ghandi R (2011) Adulteration and marketing of saffron (Crocus sativus L.). J Trop Med Plants 12:135-140.

5 Carmona M, Zalacain A, Sanchez AM, Novella JL, Alonso GL (2006) Crocetin ester, picrocrocin and its related compounds present in Crocus sativus stigmas and Gardenia Jasminoides fruits. Tentative identification of seven new compounds by LC-ESI-MS. J Agr Food Chem 54:973-979.

6 Sanchez AM, Maggi L, Carmona M, Alonso GL (2011) In: Ebeler S (ed) Progress in authentication of food and wine. American Chemical Society, Washington DC.

7 International standard ISO 3632-2: Saffron (Crocus sativus L.) test methods.

8 International standard ISO 3632-1: Saffron (Crocus sativus L.) specification.

9 Sabatino L, Scordino M, Gargano M, Belligno A, Traulo P, Gagliano G (2011) HPLC/PDA/ESI-MS evaluation of saffron (Crocus sativus L.) adulteration. Nat Prod Commun 6:1873-1876.

10 Zalacain A, Ordoudi AS, Blazquez I, Diaz-olaza EM, Carmona M, Tsimldou MZ (2005) Screening method for the detection of artificial colours in saffron using derivative UV-Vis spectrometry after precipitation of crocetin. Food Addit Contam 22:607-615.

11 Sanchez AM, Carmona M, Zalacain A, Carot TM, Jabaloyes JM, Alonso GL (2008) Rapid determination of crocetin esters and picrocrocin from saffron spice (Crocus sativus L.) using UV visible spectrophotometry for quality control. J Agr Food Chem 56:3167-3175.

12 Maggi L, Sanchez MA, Carmona M, Kanakis DC, Anastasaki E, Tarantilis AP (2011) Rapid determination of safranal in the quality control of saffron spice (Crocus Sativus L). Food Chem 127:369-373.

13 Zalacain A, Ordoudi SA, Blazquez I, Diaz-Plaza EM, Carmona M, Blazquez MZI (2005) Near-infrared spectroscopy in saffron quality control: determination of chemical composition and geographical origin. J Agr Food Chem 53:9337-9341.

14 Assimiadis MK, Tarantilis PA, Polissiou MG (1998) UV-vis FT-Raman and 1H NMR spectroscopies of cisetrans carotenoids from saffron (Crocus sativus L). J Appl Spectroscopy 52:519-522.

15 Tarantilis PA, Polissiou MG (2004) Chemical analysis and antitumor activity of natural and semi-natural carotenoids of saffron. Acta Hortic 650:447-461.

16 Zougagh M, Simonet BM, Rios A, Valcarcel M (2005) Use of nonaqueous capillary electrophoresis for the quality control of commercial saffron samples. J Chrom A 1085:293-298.

17 Caballero-Ortega H, Pereda-Miranda R, Abdullaex FI (2007) HPLC quantification of mayor active components from 11 different saffron (Crocus sativus L.) sources. Food Chem 100:1126-1131.

18 Lage M, Cantrell CL (2009) Quantification of saffron (Crocus sativus L.) metabolites crocins, picrocrocin and safranal for quality determination of the spice grown under different environmental Moroccan conditions. Sci Hortic 121:366-373.

19 Loskutov AV, Beninger CW, Hoseld GL, Sink KC (2000) Development of an improved procedure for extraction and quantitation of safranal in stigmas of Crocus sativus L. using high performance liquid chromatography. Food Chem 69:87-95.

20 Sabatino L, Scordino M, Gargano M, Belligno A, Traulo P, Gagliano G (2011) HPLC/PDA/ESI-MS evaluation of saffron (Crocus sativus L.) adulteration. Nat Prod Commun 6:1873-1876.

21 Javanmardi N, Bagheri A, Moshtaghi N, Sharifi A, Hemati-Kakhki A (2011) Identification of safflower as a fraud in commercial saffron using RAPD/SCAR marker. J Cell Mol Res 3:31-37.

22 Marieschi M, Torelli A, Bruni R (2012) Quality control of saffron (Crocus sativus L). Development of SCAR markers for the detection of plant adulterants used as bulking agents. J Agr Food Chem 60:10998-11004.

23 Babaei S, Talebi M, Bahar M (2014) Developing an SCAR and ITS reliable multiplex PCR-based assay for safflower adulterant detection in saffron samples. Food Control 35:323-328.

24 Torelli A, Marieschi M, Bruni R (2014) Authentication of saffron (Crocus sativus L.) in different processed, retail products by means of SCAR markers. Food Control 36:126-131.

25 Cubero-Leon E, Peñalver R, Maquet A (2014) Review on metabolomics for food authentication. Food Res Int 60:95-104.

26 Antignac JP, Courant F, Pinel G, Bichon E, Monteau F, Elliott C, Le Bizec B (2011) Mass spectrometry-based metabolomics applied to the chemical safety of food. Trends Anal Chem 30:292-301.

27 Hu C, Xu G (2013) Mass-spectrometry-based metabolomics analysis for foodomics. Trends Anal Chem 52:36-46.

28 Petrakis EA, Cagliani LR, Polissiou MG, Consonni R (2015) Evaluation of saffron (Crocus sativus L.) adulteration with plant adulterants by ^{1}H NMR metabolite fingerprinting. Food Chem 173:890-896.

29 Ciborowski M, Teul J, Martin-Ventura JL, Egido J, Barbas C (2012) Metabolomics with LC–QTOF-MS permits the prediction of disease stage in aortic abdominal aneurysm based on plasma metabolic fingerprint. PLoS One. Doi: 10.1371/journal.pone.0031982.

30 Westerhuis J, Hoefsloot H, Smit S, Vis D, Smilde A, van Velzen E, van Duijnhoven J, van Dorsten F (2004) Assessment of PLSDA cross validation. Metabolomics 4:81-89.

31 El Rammouz R, Letisse F, Durand S, Portais JC, Moussa ZW, Fernandez X (2010) Analysis of skeletal muscle metabolome: evaluation of extraction methods for targeted metabolite quantification using liquid chromatography tandem mass spectrometry. Anal Biochem 398:169-177.

32 Bruce S, Tavazzi I, Parisod V, Rezzi S, Kochhar S, Guy P (2009) Investigation of human blood plasma sample preparation for performing metabolomics using ultrahigh performance liquid chromatography/mass spectrometry. Anal Chem 81:3285-3296.

33 Berrueta LA, Alfonso-Salces RM, Hérberger K (2007) Supervised pattern recognition in food analysis. J Chrom A 1158:196-214.

34 Trygg J (2002) Orthogonal projections to latent structures (O-PLS). J Chemometr 16:119-128.

35 Bylesjo M, Rantalainen M, Cloarec O, Nicholson JK, Holmes E, Trygg J (2006) OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. J Chemometr 20:341-351.

36 Westerhuis JA, Velzen EJJ, Hoefsloot HCJ, Smilde AK (2010) Multivariate paired data analysis: multilevel PLSDA versus OPLDA. Metabolomics 6:119-128.

37 van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, van der Werf MJ (2006) Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics 7:142-147.

38 Rubingh CM, Bijlsma S, Derks E, Bobeldijk I, Verheij ER (2006) Assessing the performance of statistical validation tools for megavariate metabolomics data. Metabolomics 2:53-61.

39 Nørbæk R, Brandt K, Nielsen JK, Ørgaard M, Jacobsen N (2002) Flower pigment composition of Crocus species and cultivars used for a chemotaxonomic investigation. Biochem Syst Ecol 30:763-791.

40 Goupy P, Vian MA, Chemat F, Caris-Veyrat C (2013) Identification and quantification of flavonols, anthocyanins and lutein diesters in tepals of Crocus sativus by ultra performance liquid chromatography coupled to diode array and ion trap mass spectrometry detections. Ind Crop Prod 44:496-510.

41 Ferreres F, Llorach R, Gil-Izquierdo A (2004) Characterization of the interglycosidic linkage in di-, tri-, tetra- and pentaglycosylated flavonoids and differentiation of positional isomers by liquid chromatography/electrospray ionization tandem mass spectrometry. J Mass Spectrom 39:312-321.

42 Serrano-Díaz J, Sańchez AM, Martínez-Tomé M, Winterhalter P, Alonso GL (2014) Flavonoid Determination in the Quality Control of Floral Bioresidues from Crocus sativus L. J Agric Food Chem 62:3125-3133.

43 Straubinger M, Jezussek M, Waibel R, Winterhalter P (1997) Novel glycosidic constituents from saffron. J Agr Food Chem 45:1678-1681.

44 Montoro P, Tuberoso CIG, Maldini M, Cabras P, Pizza C (2008) Qualitative profile and quantitative determination of flavonoids from Crocus sativus L. Petals by LC–MS/MS. Nat Prod Commun 3:2013-2016.

45 Termentzi A, Kokkalou E (2008) LC–DAD–MS (ESI+) analysis and antioxidant capacity of Crocus sativus petal extracts. Planta Med 74:573-581.

46 Vignolini P, Heimler D, Pinelli P, Ieri F, Sciullo A, Romani A (2008) Characterization of by-products of saffron (Crocus sativus L.) production. Nat Prod Commun 3:1959-1962.

47 Carmona M, Sánchez AM, Ferreres F, Zalacain A, Tomás-Barberán F, Alonso GL (2007) Identification of the flavonoid fraction in saffron spice by LC/DAD/MS/MS: Comparative study of samples from different geographical origins. Food Chem 100:445-450.

48 Harborne JB (1999) In: Baxter H (ed) The Handbook of Natural Flavonoids. Wiley, New York.