Determination of Phenolic Compounds in Paprika by Ultra-High Performance Liquid Chromatography-Tandem Mass Spectrometry. Application to Product Designation of Origin Authentication by Chemometrics.

Sergio Barbosa[a,#], Guillem Campmajó[a,b,#], Javier Saurina[a,b], Lluis Puignou[a,b] and Oscar Núñez[a,b,c,*]

[a] Department of Chemical Engineering and Analytical Chemistry, University of Barcelona. Martí i Franquès 1-11, E08028 Barcelona, Spain.

[b] Research Institute in Food Nutrition and Food Safety, University of Barcelona. Av. Prat de la Riba 171, Edifici Recerca (Gaudí), E-08901 Santa Coloma de Gramanet, Barcelona, Spain.

[c] Serra Hunter Fellow. Generalitat de Catalunya (Spain).

* Corresponding author: Oscar Núñez

(Phone: 34-93-403-3706, Fax: 34-93-402-1233, e-mail: oscar.nunez@ub.edu)

[#] These authors contributed equally in the research of this work.

Abstract

A UHPLC-ESI-MS/MS method was developed for the determination of 36 phenolic compounds in paprika. The proposed method showed good method performance with limits of quantitation between 0.03 – 50 µg/L for 16 compounds, and between 50 µg/L and 1 mg/L for 12 compounds. Good linearity ($r^2 > 0.995$), run-to-run and day-to-day precisions (%RSD values <12.3% and <19.2%, respectively), and trueness (relative errors <15.0%) were obtained. The proposed method was applied to the analysis of 111 paprika samples from different production regions: Spain (La Vera PDO and Murcia PDO) and Czech Republic, each one including different flavor varieties (sweet, bittersweet, spicy). Phenolics profiles and concentration levels showed to be good chemical descriptors to achieve paprika classification and authentication according to the production region by principal component analysis (PCA) and partial least squares regression – discriminant analysis (PLS-DA). In addition, perfect classification among flavor varieties for Murcia PDO and Czech Republic samples was also obtained.

Keywords: Paprika; UHPLC-ESI-MS/MS; Phenolic compounds; Food Classification; Principal component analysis (PCA); Partial least squares regression-discriminant analysis (PLS-DA); Food Authentication

INTRODUCTION

Paprika is a spice obtained after drying and grinding fruits of the genus *Capsicum* that belongs to the *Solanaceae* family.[1] Within this genus there are approximately 39 species, including wild, semi-domestic and domestic ones, such as *C. annuum*, *C. chinense*, *C. baccatum*, *C. frutescens*, and *C. pubescens*, growing in different parts of the world, being *C. annuum* the most usual.[2,3] Paprika is commonly used to add flavor and color to many foods such as baked goods, beverages, meat, soup, ice cream, candy and seasoning mixes,[4] but is also used in medicine, cosmetics, protective sprays or even as adsorbents to remove contaminants.[5–9] Paprika contains a large number of bioactive compounds with great health-promoting properties such as carotenoids (provitamin A), ascorbic acid (vitamin C), tocopherols (vitamin E), capsaicinoids and phenolic compounds.[10] Among them, it is worth noting the importance of phenolic compounds that are widely distributed in plants, many of which are essential secondary metabolites that contribute to the sensory properties of foods such as color and aroma.[11] These phenolic and polyphenolic compounds have a high antioxidant activity and show potential health benefits such as vascular protection, antihepatotoxic, antiallergic, antiproliferative, antiosteoporotic, anti-inflammatory, antitumor, anti-diabetic, anti-obesity, etc.[12–14]

Current methods for the determination of polyphenols include global tests for the total polyphenolic content based on colorimetric or fluorimetric methodologies, or more specific ones employing capillary electrophoresis, liquid chromatography or gas chromatography techniques.[1,14–16] By far, liquid chromatography with either UV-detection or coupled to mass spectrometry is the most widely used technique for the determination of polyphenols.[13,17] Nevertheless, the great chemical diversity of these compounds and the low concentration levels in which they are found make liquid chromatography coupled to mass spectrometry or tandem mass spectrometry (LC-MS(/MS)) the most effective method for the characterization, identification and determination of polyphenols in paprika samples.[11,18,19] Previous studies have reported that the main phenolic compounds found in paprika are vanillic, caffeic, ferulic, *p*-coumaric and *p*-hydroxybenzoic acids.[20]

Food manufacturers, as well as the public in general, are increasingly concerned about food quality attributes and, therefore, the demand for food products of a specific geographical origin grows. Within this context, and with the aim of preserving the

reputation of the products and supporting good practices in rural and agricultural activities, the European legislation has established several quality parameters related to the protection of geographical indications and appellations of origin of agricultural and food products (Council Regulation, EEC No. 510/2006[21]): Protected Designation of Origin (PDO) that links the products with the defined geographical area where they are produced; Protected Geographical Indication (PGI) that links products to a geographical area where at least one step of production occurred; and Traditional Specialties Guaranteed (TSG) that protects traditional production methods.[22]

In Spain, there are two production areas of paprika with PDO recognized by the European Union: La Vera, from the north of the province of Cáceres (Extremadura), and the province of Murcia. Despite having a common origin and practically parallel development, the production process is different in each of these areas.[23] In both cases, the product is the result of drying and grinding the fruits of *Capsicum* species, but differences in fruit varieties and drying processes provide different organoleptic characteristics. The red fruits used for the production of La Vera paprika are dried with oak or holm oak firewood, by the traditional Vera system, and belong to the *Capsicum* annuum varieties of the Ocales group (Jaranda, Jariza and Jeromín) and Bola.. In contrast, red fruits of *Capsicum annuum* from the Bola variety are used for the production of Murcia Paprika PDO, of sweet flavor and with little weight, dried under sun conditions.[1,21,24–26]

Paprika is a worldwide consumed species susceptive of adulteration practices to attain economic benefits. The substitution of ingredients, the addition of (illegal) substances and false declarations of origin are important and challenging issues facing the authorities of the food industry.[27] Moreover, the characteristics of paprika, as well as the content of phenolic compounds, may differ due to multiple factors such as the varieties, climatic conditions, growing areas, water resources, ripening stage, agronomy conditions, pre- and post-harvest treatment, etc.[11] As a result, the content of phenolic and polyphenolic compounds in paprika products can be exploited as a source of analytical data to establish the product classification and authentication, both in the prevention of fraudulent adulterations and in the correct assignment of the PDO declarations.

In this work, a ultra-high performance liquid chromatography-electrospray-tandem mass spectrometry (UHPLC-ESI-MS/MS) method using a triple quadrupole (QqQ) analyzer has been developed for the determination and quantification of 36

phenolic and polyphenolic compounds in paprika, and subsequent characterization, classification and authentication of paprika samples by multivariate chemometric methodologies. Chromatographic and electrospray ion source conditions were optimized, and the method performance was established by determining quality parameters such as linearity, limits of detection, limits of quantitation, run-to-run and day-to-day precision, and trueness. A total of 111 paprika samples belonging to La Vera PDO and Murcia PDO (Spain) and to Czech Republic were analyzed with the proposed methodology after applying a simple extraction method using acetonitrile/water (80:20 v/v) solution as extracting agent. Then, contents of the 36 phenolic and polyphenolic compounds were employed as chemical descriptors of the analyzed paprika samples to their classification and authentication by principal component analysis (PCA) and partial least squares-discriminant analysis (PLS-DA).

MATERIALS AND METHODS

Reagents and solutions

All standards and chemicals used in this work were of analytical grade, unless otherwise indicated. Structures, family group, CAS number and supplier of the 36 phenolic and polyphenolic compounds under study are indicated in Table S1 (Supporting information). Individual stock standard solutions (ca. 1000 mg/L) were prepared in methanol in amber glass vials. Intermediate standard working solutions were prepared weekly from these individual stock standard solutions by appropriate dilution with water. All stock and intermediate working solutions were stored at 4 ^{o}C for no more than 1 month. LC-MS quality water, methanol and acetonitrile (Chromasolv^{TM} quality) were purchased from Honeywell (Riedel-de-Haën, Seelze, Germany). Formic acid (≥98%) was obtained from Sigma-Aldrich (St. Louis, MO, USA).

Instrumentation

The determination of polyphenols and phenolic acids was carried out on an Open Accela UHPLC instrument (Thermo Fisher Scientific, San José, CA, USA), equipped with a quaternary pump and a CTC autosampler. The separation was performed by reversed-phase chromatography using an Ascentis Express C18 fused-core (100 x 2.1 mm, 2.7 µm partially porous particle size) column from Supelco

(Bellefonte, PA, USA), and gradient elution using 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B) as mobile phase components, with a mobile phase flow-rate of 300 µL/min. The elution gradient program was as follows: 0–5.5 min, isocratic elution at 5% solvent B; 5.5–6.5 min, linear gradient up to 10% solvent B; 6.5–12 min, isocratic elution at 10% followed by a 1 min-increase to 20% solvent B; 13–18 min, isocratic elution at 20% solvent B; 18–19 min, linear gradient raising up to 50% solvent B and then 2 min elution at this percentage; 21-22 min, linear gradient to 95% solvent B and 3 min keeping these composition of the mobile phase. Afterwards, return to initial conditions for a 5 min-column re-equilibration, and completing a total elution program time of 30 min. The chromatographic column was kept at room temperature, and an injection volume of 10 µL, full loop mode, was employed.

The UHPLC instrument was coupled to a TSQ Quantum ultra AM triple quadrupole (QqQ) mass analyzer (Thermo Fisher Scientific), equipped with hyperbolic quadrupoles and an heated-electrospray ionization (H-ESI) source. Nitrogen with a purity of 99.98% was employed for the ESI sheath gas, ion sweep gas, and auxiliary gas at flow rates of 60, 20, 0 a.u. (arbitrary units), respectively. Other H-ESI parameters were as follow: capillary voltage in negative ion mode, -2.5 kV; H-ESI vaporizer temperature, 350 ^{o}C; ion transfer tube temperature, 350 ^{o}C. For compound quantitation and confirmation, multiple reaction monitoring (MRM) acquisition mode by recording two selected reaction monitoring (SRM) transitions (quantifier and qualifier transitions) was employed for all studied compounds except betulinic acid that showed no fragmentation under working conditions. A mass resolution of 0.7 *m/z* full width at half maximum (FWHM) on both quadrupoles (Q1 and Q3), and a scan width of 0.5 *m/z* were used. Fragmentation was carried out by using argon as collision gas at a pressure of 1.5 mtorr, and the optimal normalized collision energies (NCE) for each SRM transition monitored (quantifier and qualifier) are shown in Table 1. The precursor ion selected, precursor and product ion assignments, quantifier/qualifier ion ratios, and the tube lens offset voltage for each compound under study are also summarized in Table 1. To improve sensitivity, the acquired chromatogram was segmented into 4 windows (Table 1), and a dwell time of 50 ms, and 1 microscans were employed. The control of the UHPLC-ESI-MS/MS system and the data processing was performed by using the Xcalibur software version 2.1 (Thermo Fisher Scientific).

Samples and sample treatment

A total of 111 paprika samples, purchased from local markets in Spain and Czech Republic, were analyzed. The set included 72 La Vera PDO paprika samples (26 sweet, 23 bittersweet, and 23 spicy flavor), 24 Murcia PDO paprika samples (12 sweet and 12 spicy flavor), and 15 Czech Republic paprika samples (5 sweet, 5 smoked-sweet, and 5 spicy flavor).

Sample treatment was performed following a previously described method.[1,28] Briefly, 0.3 g of paprika were extracted with 3 mL of a water:acetonitrile (20:80 *v/v*) solution in a 15 mL PTFE tube. Extraction was performed by stirring in a Vortex for 1 min (Stuart, Stone, United Kingdom) followed by sonication for 15 min (2510 Branson ultrasonic bath, Hampton, NH, USA). Then, sample extracts were centrifuged for 30 min at 4500 rpm (Rotana 460 HR centrifuge, Hettich, Germany), and the supernatant extract filtered through 0.45 µm nylon filters (Whatman, Clifton, NJ, USA) and stored at –18 ^{o}C in 2 mL glass injection vials until analysis.

A quality control (QC) solution was prepared by mixing 50 µL of each sample extract. This QC was employed to evaluate the repeatability of the method and the robustness of the chemometric results.

Samples were randomly analyzed with the proposed UHPLC-ESI-MS/MS method. Moreover, a QC and an instrumental chromatographic blank of acetonitrile were also injected every ten analyzed samples.

Data analysis

Principal component analysis (PCA) and partial least squares regression-discriminant analysis (PLS-DA) calculations were performed using Stand Alone Chemometrics Software (SOLO) from Eigenvector Research.[29] Detailed description about the theoretical background of these methods can be found elsewhere.[30]

X-data matrices in both PCA and PLS-DA consisted of the concentration levels of the 36 phenolic and polyphenolic compounds quantified in the set of paprika samples and QCs, whereas the Y-data matrix in PLS-DA defined the membership of each sample in the corresponding class. Data was autoscaled to equalize the influence of

major and minor compounds on the descriptive models. Scatter plots of scores and loadings from principal components (PCs), in PCA, and from latent variables (LVs), in PLS-DA, were employed to study the distribution of samples and variables (quantified compounds), revealing patterns that could be correlated to their characteristics.

RESULTS AND DISCUSSION

UHPLC chromatographic separation

As commented in the introduction section, one of the objectives of the present work is the development of a LC-MS/MS method for the determination of a total of 36 phenolic and polyphenolic compounds, which belong to different phenolic classes, in paprika samples. The separation of polyphenols and phenolic acids in food products by LC-MS techniques is normally addressed by reversed-phase chromatography under gradient elution conditions using acidified water and methanol or acetonitrile as mobile phase components.[18] For that purpose, as a first attempt in this work, the separation was carried out with an Ascentis Express C18 fused-core (100 x 2.1 mm, 2.7 µm partially porous particle size) column, using water and acetonitrile, both acidified with 0.1% formic acid), as mobile phase components, and applying a universal gradient elution profile from 0 to 90% acetonitrile in 25 min. Under these conditions, multiple co-elutions were observed, and almost all the analyzed compounds eluted within the first 5 min, showing that when acetonitrile was used as organic mobile phase modifier low elutropic strength was needed for the elution of this family of compounds by reversed-phase chromatography. Therefore, the separation of the studied compounds was optimized by combining isocratic and linear gradient elution steps at low acetonitrile content (between 5 to 50%) to improve separation among the more polar phenolic acids, increasing then the acetonitrile content to elute all the compounds. It should be noted that due to the high number of compounds under study, a compromise between chromatographic resolution and analysis time was considered. Figure 1 shows the proposed UHPLC chromatographic separation for the 36 studied phenolic and polyphenolic compounds (see elution program in the instrumentation section). As can be seen, an acceptable chromatographic separation was obtained in less than 26 min, although still some partial and total co-elutions were found for some compounds, such as for homovanillic and syringic acids (peaks 11 and 12), *p*-coumaric acid, (-)-epigallocatechin gallate and syringaldehyde (peaks 16, 17 and 18), and veratric and

ferulic acids (peas 21 and 22). However, the use of MS detection under MRM acquisition mode allowed to overcome problems dealing with partial and total co-elutions for the correct determination of the studied compounds. In this regard, different SRM transitions were monitored for the co-eluting compounds and no ion-suppression effects within ESI were present (that will be addressed in the next section).

UHPLC-ESI-MS/MS acquisition conditions

The ionization of the studied compounds under H-ESI conditions was thoroughly investigated. First, ion source parameters were tuned to generate the highest number of ions and to improve the obtained signal. For that purpose, these parameters were optimized by infusion of 100 mg/L standard solutions of each one of the studied compounds at a flow rate of 15 µL/min and using the syringe pump integrated in the TSQ QqQ instrument, mixed with 200 µL/min of a 0.1% formic acid acidified water/acetonitrile (1:1 *v/v*) solution by means of a Valco zero dead volume tee piece from Supelco. Then, for each one of the indicated ion source parameters, the optimal value was selected as the one providing the highest signal for most of the studied compounds (see instrumentation section). In contrast, a specific ESI tube lens offset voltage was selected for each compound, and the optimal values obtained are summarized in Table 1.

Full scan MS spectra (*m/z* 50 – 1000) of individual solutions of all the studied compounds in negative ionization mode were also registered. As an example, Figure S1a (Supporting information) shows the obtained MS spectra of syringaldehyde and ethyl gallate. As can be seen, the most abundant ion (base peak) in both spectra is the deprotonated molecule, $[M-H]^-$, at *m/z* 181.1 and 197.2 for syringaldehyde and ethyl gallate, respectively. Similar results were obtained for most of the studied phenolic and polyphenolic compounds, being the deprotonated molecule the spectrum base peak. Moreover, adduct formation with the mobile phase components was not observed. In general, no ion in-source fragmentation was obtained, excepting some particular compounds. For instance, in the case of polydatin the spectrum base peak was not the deprotonated molecule but the $[M-H-C_6H_{10}O_5]^-$ ion at *m/z* 227.0, although the $[M-H]^-$ was also very abundant. In the case of syringaldehyde (Figure S1a), and gentisic and 4-hydroxybenzoic acids, ion source fragmentations with relative intensities lower than 40% and 60%, respectively, were observed. Finally, it should be mention that in most of

the MS spectra obtained, a signal at *m/z* 91.2 was also observed due to the dimmer formation of the formic acid present in the mobile phase ($[HCOOH\text{-}HCOO]^-$). After the study of the MS spectra, the deprotonated ion was then proposed as the precursor ion for the further fragmentation studies (Table 1).

Fragmentation of the phenolic and polyphenolic compounds under study in the QqQ mass analyzer was also evaluated under tandem MS condition. As an example, Figure S1b and S1c (Supporting information) show the normalized collision energy curves and the product ion scan spectra, respectively, for syringaldehyde and ethyl gallate. The two most intense and characteristic product ions of each compound were selected for the quantifier and qualifier SRM transitions, and they are summarized in Table S1, together with the optimal NCE for each SRM transition and the quantifier/qualifier ion ratio. As can be seen in the table, all the compounds with partial or total co-elution in the chromatographic separation previously commented (Figure 1) showed different precursor-product ion transitions for both quantifier and qualifier ions.

In addition, ion-suppression effect in the ESI source for those co-eluting compounds was evaluated by comparing their signal when analyzed individually and under co-elution conditions at the same concentration level. In all cases, ion-suppression was lower than 10%, in agreement with previous reported studies.[31] Therefore, baseline chromatographic separation is not mandatory because these co-elutions can be selectively resolved by tandem MS using the appropriate SRM transitions.

Instrumental method performance

Method performance was evaluated from instrumental quality parameters such as limits of detection, limits of quantitation, linearity, run-to-run and day-to-day precision, and trueness. The obtained results for the 36 phenolic and polyphenolic compounds determined are summarized in Table 2.

Limits of detection (LODs), based on a signal-to-noise ratio of 3:1, were assessed by analyzing standard solutions at low concentration levels, obtaining values in a wide range depending on the compound (from 0.01 µg/L for D-(-)-quinic acid to 1.4 mg/L for kaempferol). Limits of quantitation (LOQs), based on a signal-to-noise ratio of 10:1, in the range 0.03 µg/L – 4.5 mg/L were then established. Of those, seven compounds showed LOQ values equal or below 1 µg/L, nine compounds in the range 1

– 50 µg/L, twelve compounds in the range 50 µg/L – 1 mg/L, and only 8 compounds provided LOQ values higher than 1 mg/L. Taking into account that these compounds are naturally occurring secondary metabolites in plant-based products and the huge variety of compounds and concentration levels that can be found (usually at the relatively low to high mg/L level), these values are acceptable for the quantitation of this family of compounds in paprika samples.

External calibration curves using phenolic and polyphenolic standards prepared in water and based on peak area at concentrations above LOQ to 15 mg/L were established. Very good linearities with correlation coefficients (R^2) higher than 0.995 were obtained.

Run-to-run and day-to-day precisions for compound quantification were also calculated at four concentration levels (5 µg/L, 50 µg/L, 500 µg/L and 10 mg/L) and the results are also given in Table 2. In the case of run-to-run precision, five replicate determinations for each concentration level were performed within the same day. For day-to-day precision, 15 replicate determinations at each concentration level were carried out within three non-consecutive days (five replicate determinations each day). In general, run-to-run precisions below 12.3%, expressed as percentage of relative standard deviations (%RSD), were obtained in all cases. As expected, better precisions were achieved at the highest concentration level evaluated (10 mg/L), with RSD values in the range 0.2 – 4.4% (for 33 compounds), and only asiatic and betulinic acids showed higher RSD values (6.7% and 7.3%, respectively). Precision slightly worsened at lower concentrations for those compounds that were still detected under the selected conditions, but the figures of merit were very acceptable, with values below 5.9%, 9.9% and 12.3% for the 500, 50 and 5 µg/L concentration levels, respectively. RSD values slightly increased when calculating day-to-day precisions, as expected. Nevertheless, RSD values below 13.2%, 8.6%, 15.9% and 19.2% for the 10 mg/L, 500 µg/L, 50 µg/L and 5 µg/L concentration levels, respectively, being quite acceptable taking into consideration the evaluated concentration levels and the methodology employed.

Method trueness was also evaluated at the four concentration levels by comparing the spiked concentrations with those calculated by external calibration using standards prepared in water. Relative errors (%) lower than 8.2, 12.6, 15.0 and 13.3% for the 10 mg/L, 500 µg/L, 50 µg/L and 5 µg/L concentration levels, respectively, were obtained.

The results showed that the proposed UHPLC-ESI-MS/MS method was very satisfactory in terms of sensitivity, precision and trueness for the determination of the 36 studied phenolic and polyphenolic compounds at the expected concentration levels.

Sample analysis

The applicability of the proposed UHPLC-ESI-MS/MS method for the determination of the 36 studied compounds in paprika was evaluated. Paprika samples were extracted by solid-liquid extraction with water:acetonitrile (20:80 *v/v*) as described in the experimental section. The obtained extracts were then analyzed in triplicate with the proposed analytical method and targeted compounds were quantified by external calibration. Quantitation results for all the 111 paprika samples analyzed are provided in the Supporting information (Phenolic and Polyphenolic concentration in Analysed Paprika Samples.xlxs). As an overview, Table 3 shows, for each compound, the concentration ranges and the mean values ± standard deviations found in the analyzed paprika samples depending on the production region (La Vera PDO, Murcia PDO and Czech Republic) and the paprika flavors. Gallic acid, quercetin and kaempferol were always detected below the LOQ value. 16 of the studied compounds (D-(-)-quinic acid, arbutin, 4-hydroxybenzoic acid, gentisic acid, (+)-catechin, syringic acid, (-)-epicatechin, ethyl gallate, (-)-epigallocatechin gallate, procyanidin C1, veratric acid, polydatin, procyanidin A2, fisetin, morin and asiatic acid) were not detected in any of the 111 paprika samples (these compounds were not included in Table 3). Anyway, these compounds were preliminarily selected for this study because of their presence in other similar matrices such as Serbian red spice paprika, Italian red sweet pepper, or in red pepper fruits and seed oils. [32–35]

Data was first analyzed with univariate methods trying to recognize some tentative biomarkers of the different paprika types. The average concentrations and boxplots comparing the three geographical origins and/or the flavor varieties suggested that some compounds were up- or down-expressed depending on the classes. Some representative examples are given in the boxplots with whiskers of Figure S2 (Supporting information) including model compounds much more abundant in one of the classes and others quite homogeneously distributed.

More in detail, some compounds were only found in some specific paprika samples depending on the production region so they could be considered as putative

markers with high selectivity with respect to origins. For example, homogentisic acid was only detected in Czech Republic samples, although always below the LOQ. Umbelliferone was only found, at low concentrations, in the spicy flavor paprika from Czech Republic, while betulinic acid was only found in La Vera PDO samples.

Other general patterns were extracted concerning non-selective compounds. For instance, homoplantaginin, rosmarinic acid and nepetin-7-glucoside exhibited concentrations 3- to 10-fold higher in Czech Republic samples compared to the other origins. A similar trend was found with hydroxycinnamic acids, also more abundant in Czech Republic paprika. For La Vera PDO, homovanillic acid and, especially, syringaldehyde, were quite characteristic. In contrast, no unique or featured molecules were encountered for Murcia samples which displayed, in general, intermediate concentration values between La Vera and Czech Republic. As an example, Figure S3 (Supporting information) depicts bar plots showing the distribution of three selected compounds (syringaldehyde, rutin, and nepetin-7-glucoside) in the analyzed paprika samples. It can be seen that rutin shows quite similar levels within all the paprika samples. In contrast, as commented above, syringaldehyde and nepetin-7-glucoside are more characteristic of La Vera PDO and Czech Republic samples, respectively. These clear differences in phenolic and polyphenolic distribution and concentrations depending on the region and flavor varieties may allow to propose polyphenols as good chemical descriptors to address paprika authentication.

The significance of the differences in the concentration values among classes was evaluated using statistical test. As a result, most of the previous considerations regarding the occurrence of quite featured compounds of the different classes could be confirmed. Results commented here have been limited to various illustrative cases since a comprehensive analysis dealing with all variables seems to be excessive. Data given as follows corresponds to the probability (p values) of t-student for the comparison of the means of two classes before a Fisher test of variances. We assume a confidence level of 0.99 so when $p < 0.01$ differences in the analyte concentrations among the classes are significant. Results reveal the existence of several compounds such as syringaldehyde (at least, $p < 0.0006$), caffeic acid (at least, $p < 0.0042$) and homoplantaginin (at least, $p < 0.0016$) with statistically relevant differences in the concentration levels depending on the origin. Other species such as ferulic acid and nepetin-7-glucoside show no significant differences among Murcia and Czech Republic ($p = 0.02$ and 0.048, respectively). Finally, compounds such as chlorogenic acid are

unspecific so its role in class description and discrimination is quite irrelevant (p = 0.04, 0.04 and 0.91 for La Vera/Murcia, Murcia/Czech Republic and La Vera/Czech Republic, respectively).

PDO authentication

Phenolic and polyphenolic concentration levels found in the analyzed paprika samples were evaluated as potential chemical descriptors to address sample classification and authentication. As a first approach a non-supervised exploratory PCA strategy was employed with the aim of studying the grouping trends among the analyzed samples. A matrix data was built including the 36 compound concentrations found in the 111 paprika samples and the QCs, and was subjected to PCA. Figure 2 shows the score plot of PC1 *vs* PC2 obtained. As can be seen, QCs appeared grouped and located close to the center area of the plot, showing the good performance and robustness of the proposed method and the chemometric results. QCs appeared distributed in the same area than La Vera Paprika PDO samples because QC composition is enhanced on La Vera Paprika due to the high number of samples belonging to this group (72 out of 111 paprika samples). Paprika samples were perfectly discriminated by PC1 in three separate groups: La Vera PDO at the left of the score plot, Murcia PDO at the top-right area, and Czech Republic samples at the bottom-right area of the plot. Therefore, concentration levels found with the proposed UHPLC-ESI-MS/MS method are excellent chemical descriptors to achieve sample discrimination regarding the paprika production region. In addition, paprika flavors from Murcia PDO (sweet *vs* spicy) and from Czech Republic (sweet *vs* smoked-sweet *vs* spicy) samples are also perfectly separated, being discriminated by PC2 and by PC1 in the case of Murcia PDO and Czech Republic samples, respectively. In contrast, no discrimination was observed among La Vera PDO paprika flavors (sweet, bittersweet and spicy), and all the samples appeared mixed. As previously commented in the introduction, phenolic and polyphenolic distribution and content in plant-based products may be related to multiple parameters such as climatic conditions, growing areas, water resources, agronomy conditions, etc.

The study of the PCA loadings plot allow to see which variables (concentration) are defining the separation observed in the score plot. Figure S4 (Supporting information) shows the obtained PCA loadings plot of PC1 *vs* PC2. Thus, the separation

of Czech Republic samples is achieved mainly by the presence of homoplantaginin, nepetin-7-glucoside, *p*-coumaric acid and kaempferol among other compounds. Chlorogenic acid, rutin and hesperidin are more discriminating compounds for the Murcia PDO samples. In contrast, vanillin, homovanillic acid, syringaldehyde and quercetin seem to be the more characteristic compounds to separate La Vera PDO samples from the other two groups. Although more studies will be necessary, *a priori* these compounds would be good candidates as potential biomarkers for the authentication of paprika.

A supervised pattern recognition technique such as PLS-DA was used to discriminate paprika according to their geographical and/or botanical origins for authentication purposes. In this case, the X-data matrix was again the concentration of the compounds determined in the studied samples, while the Y-data matrix was the sample class.

A first study was focused on the classification of paprika samples according to geographical origin into La Vera, Murcia and Czech Republic types. In this case, the calibration set was composed of 48 La Vera, 16 Murcia and 10 Czech Republic samples randomly selected, which approximately corresponded to 70% of the analyzed samples. The other ca. 30% of the samples was used as the test set for prediction purposes. The optimum number of LVs established by cross validation using Venetian blinds was 4, providing the minimum of the root mean square error of cross validation (RMSECV) function. The analysis of scores and loadings of LV1 *vs* LV2 (not shown here) revealed that the three classes were perfectly separated and relevant compounds for their discrimination were similar to those annotated for PCA.

Figure S5 (Supporting information) shows the plots of the qualitative parameters (regression vector, the variable importance in projection (VIP) and the selectivity ratio) for the previously obtained PLS-DA model. These parameters allow to predict which variables (compounds) are more discriminant to achieve the obtained PLS-DA distribution. As can be seen, homovanillic acid and syringaldehyde are the compounds appearing as the most important variables in the three qualitative parameters, therefore being the two most relevant compounds for the PLS-DA classification when dealing with the paprika production region.

Figure 3 shows the classification plots corresponding to (a) La Vera (rhombus symbols) *vs* the other samples, (b) Murcia (square symbols) *vs* the other samples, and (c) Czech Republic (triangle symbols) *vs* the other samples. The dashed line indicated

the classification boundary, so samples belonging to the targeted class were located to the top while those belonging to the other types were to the bottom. Samples to be used for calibration were to the left and those for prediction were to the right side. Results indicated that the classification rate was 100% so all the samples were correctly assigned to the corresponding classes in both calibration and prediction steps (confusion matrix was [24, 0, 0; 0, 8, 0; 0, 0, 5] for La Vera, Murcia and Czech Republic, respectively).

Table S2 (Supporting information) show the validation results for both calibration and prediction. The obtained validation results are satisfactory. Calibration sensitivity and specificity are 1, and the root mean square error of cross validation (RMSEC) and the bias showed values tending to zero, ensuring a good calibration model.

PLS-DA models were also applied to each paprika production region in order to study the classification of samples according to the flavor variety, and the obtained results are shown in Figure S6 (Supporting information). In order to build them, a total of 4, 2 and 2 LVs were needed for La Vera, Murcia and Czech Republic sample classification, respectively, As can be seen, again no discrimination was observed among the different La Vera PDO paprika samples, showing that the distribution and content of the targeted compounds found on La Vera samples is not enough to allow discrimination between sweet, bittersweet and spicy samples. In contrast, perfect discrimination among flavor varieties was obtained for both Murcia PDO and Czech Republic paprika samples. Based on the qualitative parameters (regression vector, the variable importance in projection (VIP) and the selectivity ratio) for the PLS-DA models applied to Murcia PDO and Czech Republic samples (Figure S7 in the Supporting information), compounds such as vanillin, kaempferol and *p*-coumaric acid seem to be important for the discrimination of Murcia DOP flavor varieties, and others such as rutin, hesperidin and chlorogenic acid are playing also an important role. In the case of Czech Republic samples, nepetin-7-glucoside seem to be the most important compound to discriminate among the three flavor varieties under study, together with other compounds such as rutin, herperidin or *p*-coumaric acid among others.

In this work, and for the first time, an important number of phenolic and polyphenolic compounds belonging to different families were determined in a high number of Spanish paprika samples with PDO attributes. This is very important to know

the distribution and levels of these chemicals, with antioxidant properties, in paprika samples with PDO, giving additional benefits and attributes to the agricultural practices and regions producing paprika. In addition, the results obtained in this work demonstrate that the phenolic and polyphenolic profiles and contents obtained by the proposed UHPLC-ESI-MS/MS method after a very simple sample extraction can be employed as good chemical descriptors for the characterization and classification of paprika samples These compounds resulted to be very useful also for the discrimination of flavor varieties in the case of Murcia PDO and Czech Republic paprika samples. Finally, several compounds resulted to be important factors to address sample classification by PCA and PLS-DA, and could be considered as potential biomarkers for paprika authentication.

Conflict of Interest

There are no conflicts of interest to declare.

Funding

The authors thank the financial support received from Spanish Ministry of Science and Innovation (project PGC2018-095013-B-I00) and from the Agency for Administration of University and Research Grants (Generalitat de Catalunya) (projects 2017SGR-171 and 2017SGR-310).

Acknowledgements

G. Campmajó thanks the University of Barcelona for the APIF grant.

Supporting Information description: The Supporting Information is available free of charge at **https://xxx.** The supporting information includes two tables, seven figures and an excel file with the concentration levels (mg/L) of the 36 determined compounds in the 111 paprika samples analysed.

References

(1) Serrano, N.; Cetó, X.; Núñez, O.; Aragó, M.; Gámez, A.; Ariño, C.; Díaz-Cruz, J. M. Characterization and Classification of Spanish Paprika (*Capsicum Annuum* L.) by Liquid Chromatography Coupled to Electrochemical Detection with

Screen-Printed Carbon-Based Nanomaterials Electrodes. *Talanta* **2018**, *189* (June), 296–301.

(2) Barboza, G. E.; García, C. C.; González, S. L.; Scaldaferro, M.; Reyes, X. Four New Species of Capsicum (Solanaceae) from the Tropical Andes and an Update on the Phylogeny of the Genus. *PLoS One* **2019**, *14* (1), 1–26.

(3) Carrizo García, C.; Barfuss, M. H. J.; Sehr, E. M.; Barboza, G. E.; Samuel, R.; Moscone, E. A.; Ehrendorfer, F. Phylogenetic Relationships, Diversification and Expansion of Chili Peppers (Capsicum, Solanaceae). *Ann. Bot.* **2016**, *118* (1), 35–51.

(4) Bata-Vidács, I.; Baka, E.; Tóth, Á.; Csernus, O.; Luzics, S.; Adányi, N.; Székács, A.; Kukolya, J. Investigation of Regional Differences of the Dominant Microflora of Spice Paprika by Molecular Methods. *Food Control* **2018**, *83*, 109–117.

(5) Martín, A.; Hernández, A.; Aranda, E.; Casquete, R.; Velázquez, R.; Bartolomé, T.; Córdoba, M. G. Impact of Volatile Composition on the Sensorial Attributes of Dried Paprikas. *Food Res. Int.* **2017**, *100* (May), 691–697.

(6) Wu, N.; Gao, W.; Lian, Y.; Du, J.; Tie, X. The Transfer of Natural Rhodamine B Contamination from Raw Paprika Fruit to Capsicum Oleoresin during the Extraction Process. *Food Chem.* **2017**, *237*, 786–792.

(7) Topuz, A.; Dincer, C.; Özdemir, K. S.; Feng, H.; Kushad, M. Influence of Different Drying Methods on Carotenoids and Capsaicinoids of Paprika (Cv.; Jalapeno). *Food Chem.* **2011**, *129* (3), 860–865.

(8) Jackson, S.; Swiner, D. J.; Capone, P. C.; Badu-Tawiah, A. K. Thread Spray Mass Spectrometry for Direct Analysis of Capsaicinoids in Pepper Products. *Anal. Chim. Acta* **2018**, *1023*, 81–88.

(9) Tunali Akar, S.; Gorgulu, A.; Akar, T.; Celik, S. Decolorization of Reactive Blue 49 Contaminated Solutions by *Capsicum Annuum* Seeds: Batch and Continuous Mode Biosorption Applications. *Chem. Eng. J.* **2011**, *168* (1), 125–133.

(10) Wahyuni, Y.; Ballester, A. R.; Sudarmonowati, E.; Bino, R. J.; Bovy, A. G. Metabolite Biodiversity in Pepper (Capsicum) Fruits of Thirty-Two Diverse Accessions: Variation in Health-Related Compounds and Implications for Breeding. *Phytochemistry* **2011**, *72* (11–12), 1358–1370.

(11) Mudric, S. Z.; Gašic, U. M.; Dramicanin, A. M.; Ciric, I. Z.; Milojkovic-Opsenica, D. M.; Popovic-Dordevic, J. B.; Momirovic, N. M.; Tešic, Z. L. The

555 Polyphenolics and Carbohydrates as Indicators of Botanical and Geographical
556 Origin of Serbian Autochthonous Clones of Red Spice Paprika. *Food Chem.*
557 **2017**, *217*, 705–715.
558 (12) Han, X.; Shen, T.; Lou, H. Dietary Polyphenols and Their Biological
559 Significance. *Int. J. Mol. Sci.* **2007**, *8* (9), 950–988.
560 (13) Monago-Maraña, O.; Muñoz De La Peña, A.; Galeano-Díaz, T. Isocratic LC-
561 DAD-FLD Method for the Determination of Flavonoids in Paprika Samples by
562 Using a Rapid Resolution Column and Post-Column PH Change. *Talanta* **2016**,
563 *152*, 15–22.
564 (14) Monago-Maraña, O.; Durán-Merás, I.; Galeano-Díaz, T.; Muñoz De La Peña, A.
565 Fluorescence Properties of Flavonoid Compounds. Quantification in Paprika
566 Samples Using Spectrofluorimetry Coupled to Second Order Chemometric
567 Tools. *Food Chem.* **2016**, *196*, 1058–1065.
568 (15) Vallejos, S.; Moreno, D.; Ibeas, S.; Muñoz, A.; García, F. C.; García, J. M.
569 Polymeric Chemosensor for the Colorimetric Determination of the Total
570 Polyphenol Index (TPI) in Wines. *Food Control* **2019**, *106* (April), 106684.
571 (16) Fernández de Córdova, M. L.; Ruiz Medina, M. Chapter 29 - Analytical Methods
572 for Determination of Polyphenols in Beer. In *Processing and Impact on*
573 *Antioxidants in Beverages*; 2014; p Pages 289-299.
574 (17) Pardo-Mates, N.; Vera, A.; Barbosa, S.; Hidalgo-Serrano, M.; Núñez, O.;
575 Saurina, J.; Hernández-Cassou, S.; Puignou, L. Characterization, Classification
576 and Authentication of Fruit-Based Extracts by Means of HPLC-UV
577 Chromatographic Fingerprints, Polyphenolic Profiles and Chemometric Methods.
578 *Food Chem.* **2017**, *221*.
579 (18) Lucci, P.; Saurina, J.; Núñez, O. Trends in LC-MS and LC-HRMS Analysis and
580 Characterization of Polyphenols in Food. *TrAC - Trends Anal. Chem.* **2017**, *88*,
581 1–24.
582 (19) Morales-Soto, A.; Gómez-Caravaca, A. M.; García-Salas, P.; Segura-Carretero,
583 A.; Fernández-Gutiérrez, A. High-Performance Liquid Chromatography Coupled
584 to Diode Array and Electrospray Time-of-Flight Mass Spectrometry Detectors
585 for a Comprehensive Characterization of Phenolic and Other Polar Compounds
586 in Three Pepper (*Capsicum Annuum* L.) Samples. *Food Res. Int.* **2013**, *51* (2),
587 977–984.
588 (20) Chen, L.; Kang, Y. H. Anti-Inflammatory and Antioxidant Activities of Red

Pepper (Capsicum Annuum L.) Stalk Extracts: Comparison of Pericarp and Placenta Extracts. *J. Funct. Foods* **2013**, *5* (4), 1724–1731.

(21) Consejo de la union europea. RCOUNCIL REGULATION (EC) No 510/2006 of 20 March 2006 on the Protection of Geographical Indications and Designations of Origin for Agricultural Products and Foodstuffs, Off. J. European Union L93 12-25., 2006.

(22) Kamiloglu, S. Authenticity and Traceability in Beverages. *Food Chem.* **2019**, *277* (June 2018), 12–24.

(23) Zapata, M.; Bañon, S.; Cabrera, P. Zapata, M.; Bañón, S.; Cabrera, P. El Pimiento Para Pimentón, Mundi Prensa Libros, S.A., Madrid, Spain. 1992. ISBN: 978-8-47114-372-3, 1992.

(24) COMMISSION REGULATION (EC) No 982/2007 of 21 August 2007 Registering Certain Names in the Register of Protected Designations of Origin and Protected Geographical Indications (Pimentón de La Vera (PDO) — Karlovarský Suchar (PGI) — Riso Di Baraggia Biellese.

(25) Palacios-Morillo, A.; Jurado, J. M.; Alcázar, Á.; De Pablos, F. Geographical Characterization of Spanish PDO Paprika by Multivariate Analysis of Multielemental Content. *Talanta* **2014**, *128*, 15–22.

(26) COUNCIL REGULATION (EEC) No 2081/92 of 14 July 1992 on the Protection of Geographical Indications and Designations of Origin for Agricultural Products and Foodstuffs, Off. J. European Communities L208 1-8.

(27) Horn, B.; Esslinger, S.; Pfister, M.; Fauhl-Hassek, C.; Riedl, J. Non-Targeted Detection of Paprika Adulteration Using Mid-Infrared Spectroscopy and One-Class Classification – Is It Data Preprocessing That Makes the Performance? *Food Chem.* **2018**, *257* (November 2017), 112–119.

(28) Cetó, X.; Serrano, N.; Aragó, M.; Gámez, A.; Esteban, M.; Díaz-Cruz, J. M.; Núñez, O. Determination of HPLC-UV Fingerprints of Spanish Paprika (*Capsicum Annuum* L.) for Its Classification by Linear Discriminant Analysis. *Sensors (Switzerland)* **2018**, *18* (12).

(29) SOLO Software, Eigenvecto Research, Http://Www.Eigenvector.Com/Software/Solo.Htm (Accessed on 12th December 2019).

(30) Massart, J.; Vandeginste, D.L.; Buydens, B.G.M.; de Jong, L.M.C.; Lewi, S.; Smeyers-Verbeke, P.J. Handbook of Chemometrics and Qualimetrics, J. Chem.

Inf. Comput. Sci., 1997.

(31) Puigventós, L.; Navarro, M.; Alechaga, É.; Núñez, O.; Saurina, J.; Hernández-Cassou, S.; Puignou, L. Determination of Polyphenolic Profiles by Liquid Chromatography-Electrospray-Tandem Mass Spectrometry for the Authentication of Fruit Extracts. *Anal. Bioanal. Chem.* **2015**, *407* (2), 597–608.

(32) Mudrić, S.; Gašić, U. M.; Dramićanin, A. M.; Ćirić, I.; Milojković-Opsenica, D. M.; Popović-Đorđević, J. B.; Momirović, N. M.; Tešić, Ž. L. The Polyphenolics and Carbohydrates as Indicators of Botanical and Geographical Origin of Serbian Autochthonous Clones of Red Spice Paprika. *Food Chem.* **2017**, *217*, 705–715.

(33) Liu, Y.; Lv, J.; Liu, Z.; Wang, J.; Yang, B.; Chen, W.; Ou, L.; Dai, X.; Zhang, Z.; Zou, X. Integrative Analysis of Metabolome and Transcriptome Reveals the Mechanism of Color Formation in Pepper Fruit (Capsicum Annuum L.). *Food Chem.* **2020**, *306* (October 2019), 125629.

(34) Sobolev, A. P.; Mannina, L.; Capitani, D.; Sanzò, G.; Ingallina, C.; Botta, B.; Fornarini, S.; Crestoni, M. E.; Chiavarino, B.; Carradori, S.; et al. A Multi-Methodological Approach in the Study of Italian PDO "Cornetto Di Pontecorvo" Red Sweet Pepper. *Food Chem.* **2018**, *255* (February), 120–131.

(35) Chouaibi, M.; Rezig, L.; Hamdi, S.; Ferrari, G. Chemical Characteristics and Compositions of Red Pepper Seed Oils Extracted by Different Methods. *Ind. Crops Prod.* **2019**, *128* (July 2018), 363–370.

Figure captions

Figure 1. UHPLC-ESI-MS chromatographic separation of the 36 studied compounds (standard of 500 µg/L in water) under the proposed elution program (see instrumental section). MS acquisition was performed in secondary ion monitoring (SIM) mode by following the $[M-H]^-$ ion for each compound. Peak identification: 1, D-(-)-Quinic acid; 2, Arbutin; 3, Gallic acid; 4, Homogentisic acid; 5, Protocatechuic aldehyde; 6, 4-hydroxybenzoic acid; 7, Gentisic acid; 8, Chlorogenic acid; 9, (+)-catechin; 10, Caffeic acid; 11, Homovanillic acid; 12, Syringic acid; 13, Vanillin; 14, (-)-Epicatechin; 15, Ethyl gallate; 16, *p*-coumaric acid; 17, (-)-Epigallocatechin gallate; 18, Syringaldehyde; 19, Umbelliferone; 20, Procyanidin C1; 21, Veratric acid; 22, Ferulic acid; 23, Sinapic acid; 24, Polydatin; 25, Rutin; 26, Procyanidin A2; 2,7 Nepetin-7-glucoside; 28, Hesperidin; 29, Homoplantaginin; 30, Fisetin; 31, Rosmarinic acid; 32, Morin; 33, Quercetin; 34, Kaempferol; 35, Asiatic acid; 36, Betulinic acid.

Figure 2. PCA score plot of PC1 *vs* PC2 when using the 36 compound concentrations found in the analyzed paprika samples as chemical descriptors.

Figure 3. PLS-DA classification plots according to the production region. (a) La Vera versus other classes; (b) Murcia versus other classes; (c) Czech Republic versus other classes. Sample assignation: Rhombus = La Vera PDO, square = Murcia, triangle = Czech Republic. Dashed line means the classification boundary.

Table 1. Instrumental MRM acquisition parameters.

Segment	Time (min)		Compound	Tube lens offset voltage (V)	Precursor ion		Quantifier product ion			Qualifier product ion			*Quantifier/Qualifier ion ratio*
					m/z	Assignment	*m/z*	NCE (eV)	Assignment	*m/z*	NCE (eV)	Assignment	
1	0.7	1	D-(−)-Quinic acid	-50	190.9	$[M-H]^-$	85.2	20	$[M-H-C_3H_6O_4]^-$	93.2	20	$[M-H-C_4H_2O_3]^-$	2.6
1	1.0	2	Arbutin	-124	271.0	$[M-H]^-$	161.3	10	$[M-H-C_6H_{10}O_5]^-$	108.7	25	$[M-H-C_6H_6O_2]^-$	10.0
1	1.4	3	Gallic acid	-95	169.0	$[M-H]^-$	125.1	15	$[M-H-CO_2]^-$	79.0	23	$[M-H-C_2H_2O_4]^-$	17.5
1	2.3	4	Homogentisic acid	-94	167.2	$[M-H]^-$	123.0	13	$[M-H-CO_2]^-$	121.9	23	$[M-H-CHO_2]^{-\bullet}$	2.0
1	4.9	5	Protocatechuic aldehyde	-72	137.0	$[M-H]^-$	135.9	20	$[M-H-H]^{-\bullet}$	92.1	25	$[M-H-CHO_2]^{-\bullet}$	3.9
1	5.3	6	4-Hydroxybenzoic acid	-90	136.9	$[M-H]^-$	93.2	20	$[M-H-CO_2]^-$	65.0	35	$[M-H-C_2O_3]^-$	20.2
1	5.7	7	Gentisic acid	-87	153.0	$[M-H]^-$	109.0	20	$[M-H-CO_2]^-$	81.4	20	$[M-H-C_2O_3]^-$	15.6
1-2	8.4	8	Chlorogenic acid	-148	353.0	$[M-H]^-$	190.9	21	$[M-H-C_9H_6O_3]^-$	85.1	44	$[M-H-C_{12}H_{12}O_7]^-$	23.7
2	8.5	9	(+)-Catechin	-73	288.9	$[M-H]^-$	244.8	15	$[M-H-C_2H_4O]^-$	203.2	20	$[M-H-C_4H_6O_2]^-$	2.0
2	8.8	10	Caffeic acid	-63	179.0	$[M-H]^-$	134.9	16	$[M-H-CO_2]^-$	133.8	25	$[M-H-CHO_2]^{-\bullet}$	6.2
2	9.2	11	Homovanillic acid	-97	181.1	$[M-H]^-$	137.2	10	$[M-H-CO_2]^-$	122.0	16	$[M-H-C_2H_3O_2]^{-\bullet}$	6.3
2	9.4	12	Syringic acid	-83	196.9	$[M-H]^-$	182.0	14	$[M-H-CH_3]^{-\bullet}$	123.1	24	$[M-H-C_2H_2O_3]^-$	2.8
2	10.3	13	Vanillin	-77	151.2	$[M-H]^-$	136.0	15	$[M-H-CH_3]^{-\bullet}$	91.9	20	$[M-H-C_2H_3O_2]^{-\bullet}$	7.4
2-3	11.6	14	(-)-Epicatechin	-95	289.1	$[M-H]^-$	244.9	16	$[M-H-C_2H_4O]^-$	203.0	20	$[M-H-C_4H_6O_2]^-$	1.9
2-3	11.9	15	Ethyl gallate	-97	197.2	$[M-H]^-$	123.9	22	$[M-H-C_3H_5O_2]^{-\bullet}$	169.0	15	$[M-H-C_2H_4]^-$	1.6
2-3	12.2	16	p-Coumaric acid	-85	163.1	$[M-H]^-$	118.8	17	$[M-H-CO_2]^-$	93.1	35	$[M-H-C_3H_2O_2]^-$	14.8
2-3	12.3	17	(-)-Epigallocatechin gallate	-125	457.0	$[M-H]^-$	169.0	19	$[M-H-C_{15}H_{12}O_6]^-$	125.2	39	$[M-H-C_{16}H_{12}O_8]^-$	3.1
3	12.5	18	Syringaldehyde	-57	181.1	$[M-H]^-$	166.2	13	$[M-H-CH_3]^{-\bullet}$	150.9	21	$[M-H-CH_2O]^-$	1.3
3	12.9	19	Umbelliferone	-94	160.9	$[M-H]^-$	133.1	20	$[M-H-CO]^-$	105.1	23	$[M-H-C_3H_4O]^-$	2.9
3	14.2	20	Procyanidin C1	-151	864.8	$[M-H]^-$	407.0	40	$[M-H-C_{23}H_{22}O_{10}]^-$	286.9	30	$[M-H-C_{30}H_{26}O_{12}]^-$	1.4
3	14.4	21	Veratric acid	-86	181.0	$[M-H]^-$	136.9	15	$[M-H-CO_2]^-$	107.0	25	$[M-H-C_2H_2O_3]^-$	5.9

3	14.4	22	Ferulic acid	-91	193.2	$[M-H]^-$	134.1	18	$[M-H-C_2H_3O_2]^{-\bullet}$	178.1	14	$[M-H-CH_3]^{-\bullet}$	1.3
3	14.7	23	Sinapic acid	-91	223.0	$[M-H]^-$	207.9	15	$[M-H-CH_3]^{-\bullet}$	163.8	18	$[M-H-C_2H_3O_2]^{-\bullet}$	1.9
3	14.9	24	Polydatin	-144	389.1	$[M-H]^-$	227.0	20	$[M-H-C_6H_{10}O_5]^-$	185.2	38	$[M-H-C_8H_{12}O_6]^-$	9.8
3	15.0	25	Rutin	-139	609.0	$[M-H]^-$	300.1	35	$[M-H-C_{12}H_{21}O_9]^{-\bullet}$	270.9	60	$[M-H-C_{13}H_{22}O_{10}]^-$	1.8
3	15.7	26	Procyanidin A2	-155	575.1	$[M-H]^-$	284.9	22	$[M-H-C_{15}H_{14}O_6]^-$	449.0	23	$[M-H-C_6H_6O_3]^-$	1.3
3-4	15.7	27	Nepetin-7-glucoside	-135	477.1	$[M-H]^-$	315.2	25	$[M-H-C_6H_{10}O_5]^-$	299.7	35	$[M-H-C_7H_{13}O_5]^{-\bullet}$	1.4
3-4	16.8	28	Hesperidin	-139	608.8	$[M-H]^-$	301.0	20	$[M-H-C_{12}H_{20}O_9]^-$	325.1	35	$[M-H-C_{13}H_{16}O_7]$-	21.8
3-4	17.2	29	Homoplantaginin	-163	461.0	$[M-H]^-$	283.1	34	$[M-H-C_7H_{14}O_5]^-$	297.1	35	$[M-H-C_9H_8O_3]^-$	4.2
3-4	17.7	30	Fisetin	-108	285.1	$[M-H]^-$	135.1	23	$[M-H-C_8H_6O_3]^-$	120.9	27	$[M-H-C_8H_4O_4]^-$	1.9
3-4	17.8	31	Rosmarinic acid	-115	358.7	$[M-H]^-$	161.0	18	$[M-H-C_9H_{10}O_5]^-$	133.2	40	$[M-H-C_{10}H_{10}O_6]^-$	4.3
4	19.8	32	Morin	-91	301.0	$[M-H]^-$	151.1	21	$[M-H-C_8H_6O_3]^-$	148.9	29	$[M-H-C_7H_4O_4]^-$	1.8
4	20.2	33	Quercetin	-121	300.9	$[M-H]^-$	151.1	25	$[M-H-C_7H_2O_4]^-$	179.1	20	$[M-H-C_6H_2O_3]^-$	1.8
4	20.5	34	Kaempferol	-107	285.0	$[M-H]^-$	185.2	25	$[M-H-C_4H_4O_3]^-$	117.2	43	$[M-H-C_7H_4O_5]^-$	1.1
4	21.6	35	Asiatic acid	-126	487.3	$[M-H]^-$	409.1	35	$[M-H-C_2H_6O_3]^-$	379.4	45	$[M-H-C_3H_8O_4]^-$	1.3
4	24.2	36	Betulinic acid	-123	455.0	$[M-H]^-$	-	-	-	-	-	-	-

Table 2. Method performance. Instrumental quality parameters.

	Compound	LOD (µg/L)	LOQ (µg/L)	Linearity (R^2)	Run-to-run precision (RSD, %)				Day-to-day precision (RSD, %)				Trueness (relative error, %)			
					5 µg/L	50 µg/L	500 µg/L	10 mg/L	5 µg/L	50 µg/L	500 µg/L	10 mg/L	5 µg/L	50 µg/L	500 µg/L	10 mg/L
1	D-(−)-Quinic acid	0.01	0.03	0.998	7.0	4.9	4.7	2.2	14.6	6.3	6.0	4.2	5.0	0.1	5.2	3.6
2	Arbutin	0.22	0.73	0.996	9.5	8.2	4.5	1.9	16.5	9.7	5.2	4.9	4.6	0.9	2.3	2.1
3	Gallic acid	964	3214	0.995	-	-	-	3.0	-	-	-	3.7	-	-	-	4.1
4	Homogentisic acid	770	2566	0.996	-	-	-	2.5	-	-	-	3.0	-	-	-	3.8
5	Protocatechuic aldehyde	2	7	0.995	6.4	5.3	3.2	2.8	8.9	6.1	5.6	4.4	5.6	2.7	1.5	3.0
6	4-Hydroxybenzoic acid	21	71	0.999	-	8.0	3.8	1.4	-	9.5	4.8	3.9	-	5.3	4.1	3.1
7	Gentisic acid	11	35	0.998	-	1.5	0.4	0.2	-	9.4	7.9	5.0	-	15.0	6.8	3.3
8	Chlorogenic acid	0.87	3	0.999	-	4.0	2.9	2.1	-	6.7	4.8	3.7	-	7.9	0.3	4.7
9	(+)-Catechin	161	537	0.999	-	-	-	2.2	-	-	-	5.0	-	-	-	1.2
10	Caffeic acid	18	60	0.995	-	5.1	2.6	2.5	-	8.9	6.1	2.9	-	9.7	6.0	3.1
11	Homovanillic acid	425	1417	0.995	-	-	-	2.3	-	-	-	4.4	-	-	-	1.5
12	Syringic acid	18	59	0.998	-	7.9	4.0	1.9	-	15.9	8.6	3.3	-	14.0	12.6	0.8
13	Vanillin	10	33	0.998	-	5.4	1.5	1.1	-	6.6	2.3	1.7	-	3.4	4.5	1.7
14	(-)-Epicatechin	1282	4272	0.999	-	-	-	2.8	-	-	-	4.3	-	-	-	3.6
15	Ethyl gallate	262	872	0.998	-	-	3.1	1.3	-	-	6.7	2.6	-	-	9.4	6.9
16	*p*-Coumaric acid	4	12	0.995	-	5.8	2.9	2.0	-	14.1	4.3	4.1	-	13.7	5.5	8.2

17	(-)-Epigallocatechin gallate	770	2565	0.999	-	-	-	1.1	-	-	-	2.0	-	-	-	0.9
18	Syringaldehyde	2	8	0.999	-	9.9	5.0	0.7	-	10.6	5.5	1.9	-	4.5	0.3	1.5
19	Umbelliferone	0.37	1	0.998	12.3	6.4	3.4	2.3	15.4	8.3	5.6	5.4	9.4	5.3	3.4	4.0
20	Procyanidin C1	359	1196	0.998	-	-	-	1.0	-	-	-	2.6	-	-	-	3.9
21	Veratric acid	281	936	0.999	-	-	-	3.9	-	-	-	8.5	-	-	-	1.4
22	Ferulic acid	2	6	0.997	6.2	4.4	5.5	4.0	18.5	9.3	7.8	6.0	13.3	0.2	0.9	0.4
23	Sinapic acid	25	84	0.995	-	-	5.9	2.5	-	-	6.5	5.3	-	-	1.4	0.7
24	Polydatin	0.14	0.48	0.999	5.9	2.3	2.6	2.2	19.2	13.6	5.8	2.8	11.1	4.4	2.1	4.3
25	Rutin	3	9	0.996	-	5.0	4.8	4.4	-	15.7	7.3	6.4	-	2.6	3.0	4.4
26	Procyanidin A2	170	566	0.998	-	-	-	1.1	-	-	-	2.5	-	-	-	2.1
27	Nepetin-7-glucoside	0.06	0.21	0.998	6.2	5.0	3.0	1.9	12.6	7.7	5.8	5.1	12.3	8.0	4.0	3.8
28	Hesperidin	0.31	1	0.999	7.9	4.1	2.7	2.1	16.3	5.3	5.2	3.8	10.5	0.7	2.3	1.6
29	Homoplantaginin	0.19	0.63	0.998	9.1	3.5	1.9	1.3	11.4	4.7	5.7	3.3	11.8	1.2	4.0	1.9
30	Fisetin	759	2529	0.995	-	-	-	1.0	-	-	-	1.4	-	-	-	4.4
31	Rosmarinic acid	12	41	0.999	-	2.7	1.2	0.8	-	5.2	4.4	2.0	-	7.3	0.6	1.9
32	Morin	209	696	0.999	-	4.5	2.1	1.2	-	9.6	6.2	3.5	-	8.6	2.6	2.2
33	Quercetin	89	296	0.996	-	-	1.8	1.2	-	-	7.1	2.0	-	-	1.4	7.0
34	Kaempferol	1357	4522	0.998	-	-	-	1.6	-	-	-	2.7	-	-	-	3.4
35	Asiatic acid	210	700	0.995	-	-	-	6.7	-	-	-	13.2	-	-	-	0.9
36	Betulinic acid	265	885	0.998	-	-	-	7.3	-	-	-	10.3	-	-	-	5.6

Table 3. Concentrations of studied compounds in the analyzed paprika samples.

	Compound	La Vera PDO paprika (mg/L)						Murcia PDO paprika (mg/L)				Czech Republic paprika (mg/L)					
		spicy		bittersweet		sweet		spicy		sweet		spicy		Sweet-smoked		sweet	
		Range	Mean ± SD	Range	Mean ± SD	Range	Mean ± SD	Range	Mean ± SD	Range	Mean ± SD	Range	Mean ± SD	Range	Mean ± SD	Range	Mean ± SD
3	Gallic acid	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ
4	Homogentisic acid	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ
5	Protocatechuic aldehyde	0.1 – 0.6	0.3 ± 0.1	0.2 – 0.7	0.3 ± 0.1	0.1 – 0.6	0.3 ± 0.1	0.1 – 0.5	0.4 ± 0.1	0.4 – 0.8	0.7 ± 0.1	0.6 – 0.8	0.7 ± 0.1	0.5 – 0.6	0.5 ± 0.1	0.6 – 0.8	0.7 ± 0.1
8	Chlorogenic acid	5.0 –10.1	7.4 ± 1.5	5.3 – 10.4	6.9 ± 1.9	5.1 – 9.7	6.8 ± 1.2	5.2 – 6.8	6.0 ± 0.6	7.1 –12.4	10.4 ± 1.5	6.6 – 7.5	7.1 ± 0.4	5.6 – 8.3	7.0 ± 1.0	5.8 – 8.4	7.0 ± 1.0
10	Caffeic acid	1.4 – 2.8	1.8 ± 0.3	1.2 – 2.0	1.7 ± 0.2	1.0 – 2.4	1.6 ± 0.3	1.8 – 2.6	2.2 ± 0.3	1.6 – 2.0	1.8 ± 0.1	2.3 – 5.2	3.2 ± 1.2	2.3 – 7.4	4.0 ± 2.2	1.9 – 5.5	3.0 ± 1.5
11	Homovanillic acid	2.8 – 8.9	6.1 ± 1.7	1.6 – 10.3	5.6 ± 2.3	0.5 – 12.9	5.9 ± 2.8	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	4.23 – 5.0	4.6 ± 0.3	n.d.	n.d.
13	Vanillin	1.9 –11.2	3.9 ± 2.0	0.9 – 6.4	3.3 ± 1.4	1.4 – 7.2	3.5 ± 1.5	3.2 – 4.4	3.7 ± 0.3	2.3 – 3.3	2.7 ± 0.3	2.1 – 3.3	2.7 ± 0.5	1.9 – 2.7	2.3 ± 0.3	1.7 –2.8	2.2 ± 0.5
16	*p*–Coumaric acid	1.6 –5.6	3.3 ± 1.0	2.1 – 4.5	3.3 ± 0.7	1.4 – 5.0	3.1 ± 0.9	3.9 –10.3	8.2 ± 1.8	1.3 – 3.7	2.8 ± 0.8	9.7 –10.7	10.1 ± 0.4	7.4 – 8.7	8.0 ± 0.6	6.2 – 8.6	7.3 ± 1.0
18	Syringaldehyde	4.5 – 10.8	7.7 ± 1.7	4.0 – 14.3	8.5 ± 2.9	1.2 – 14.8	8.0 ± 3.0	0.8 – 1.8	1.2 ± 0.3	0.8 – 1.8	1.2 ± 0.4	1.4 – 1.8	1.6 ± 0.2	3.0 – 4.1	3.6 ± 0.5	1.6 – 1.9	1.7 ± 0.1
19	Umbelliferone	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.1 – 0.2	0.1 ± 0.1	n.d.	n.d.	n.d.	n.d..
22	Ferulic acid	1.4 – 16.3	7.1 ± 3.4	2.6 – 8.1	5.2 ± 1.6	1.3 – 9.9	5.7 ± 2.4	2.8 –19.1	10.3 ± 5.2	4.7 –16.1	9.6 ± 3.6	11.8 – 18.4	14.9 ± 2.7	12.8 – 19.2	15.2 ± 2.8	5.4 –12.1	9.5 ± 3.2
23	Sinapic acid	2.0 – 5.7	3.4 ± 1.0	1.9 – 5.0	3.2 ± 1.0	1.6 – 4.8	3.1 ± 1.0	1.2 – 6.8	2.6 ± 1.5	1.7 – 4.5	3.2 ± 0.9	1.9 – 4.6	3.7 ± 1.1	2.1 – 5.5	3.6 ± 1.4	3.1 – 4.5	4.1 ± 0.6
25	Rutin	1.3 – 6.1	3.0 ± 1.2	1.4 – 5.0	2.9 ± 0.8	1.3 – 4.5	2.8 ± 0.9	1.0 – 1.8	1.4 ± 0.3	3.8 – 5.4	4.5 ± 0.5	3.4 – 3.8	3.6 ± 0.2	2.8 – 3.3	3.0 ± 0.2	2.5 – 3.0	2.6 ± 0.2
27	Nepetin–7–glucoside	n.d. – 0.2	0.1 ± 0.1	n.d. –0.2	0.1 ± 0.1	n.d. –0.1	0.1 ± 0.0	0.1 – 0.2	0.2 ± 0.0	0.1 –0.2	0.1 ± 0.0	0.3 – 0.4	0.3 ± 0.0	0.2 – 0.2	0.2 ± 0.0	0.6 – 0.7	0.6 ± 0.0
28	Hesperidin	<LOQ – 0.4	0.4 ± 0.0	<LOQ – 5.1	2.6 ± 3.6	<LOQ – 5.8	5.8 ± 0.0	0.1 –0.5	0.2 ± 0.1	3.8 – 6.1	4.6 ± 0.7	7.1 – 9.2	7.9 ± 1.0	1.2 – 2.4	1.9 ± 0.5	<LOQ –0.2	0.1 ± 0.10
29	Homoplantaginin	0.001 – 0.008	0.004 ± 0.002	0.002 – 0.009	0.005 ± 0.002	0.001 – 0.008	0.004 ± 0.002	0.004 – 0.011	0.008 ± 0.002	0.008 – 0.017	0.011 ± 0.003	0.011 – 0.040	0.031 ± 0.012	0.004 – 0.024	0.012 ± 0.010	0.011 – 0.038	0.027 ± 0.012
31	Rosmarinic acid	1.1 – 5.5	2.3 ± 1.8	1.4 – 8.3	3.8 ± 2.5	1.1 – 5.6	2.5 ± 1.6	1.5 – 2.1	1.7 ± 0.2	1.2 – 4.4	3.6 ± 0.9	5.7 – 10.4	7.4 ± 1.8	3.7 – 7.5	4.7 ± 1.6	2.9 – 4.8	3.7 ± 0.7
33	Quercetin	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ
34	Kaempferol	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ
36	Betulinic acid	1.2 – 6.4	3.2 ± 1.6	1.4 – 9.1	3.6 ± 1.5	1.0 – 8.3	3.0 ± 1.6	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.

n.d.: not detected; S.D.: standard deviation

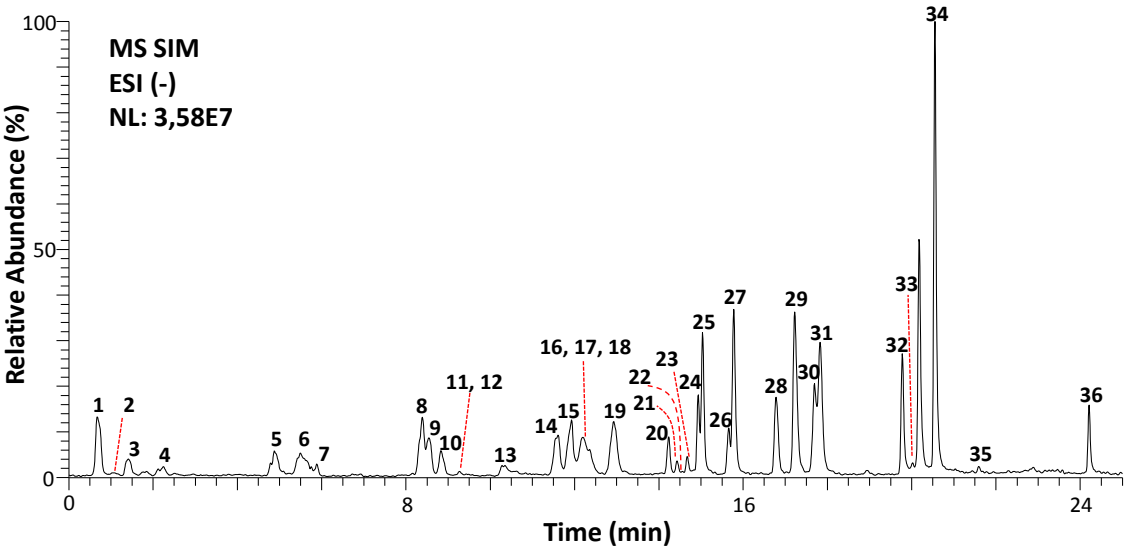


Figure 1. UHPLC-ESI-MS chromatographic separation of the 36 studied compounds (standard of 500 µg/L in water) under the proposed elution program (see instrumental section). MS acquisition was performed in secondary ion monitoring (SIM) mode by following the $[M-H]^-$ ion for each compound. Peak identification: 1, D-(-)-Quinic acid; 2, Arbutin; 3, Gallic acid; 4, Homogentisic acid; 5, Protocatechuic aldehyde; 6, 4-hydroxybenzoic acid; 7, Gentisic acid; 8, Chlorogenic acid; 9, (+)-catechin; 10, Caffeic acid; 11, Homovanillic acid; 12, Syringic acid; 13, Vanillin; 14, (-)-Epicatechin; 15, Ethyl gallate; 16, *p*-coumaric acid; 17, (-)-Epigallocatechin gallate; 18, Syringaldehyde; 19, Umbelliferone; 20, Procyanidin C1; 21, Veratric acid; 22, Ferulic acid; 23, Sinapic acid; 24, Polydatin; 25, Rutin; 26, Procyanidin A2; 2,7 Nepetin-7-glucoside; 28, Hesperidin; 29, Homoplantaginin; 30, Fisetin; 31, Rosmarinic acid; 32, Morin; 33, Quercetin; 34, Kaempferol; 35, Asiatic acid; 36, Betulinic acid.

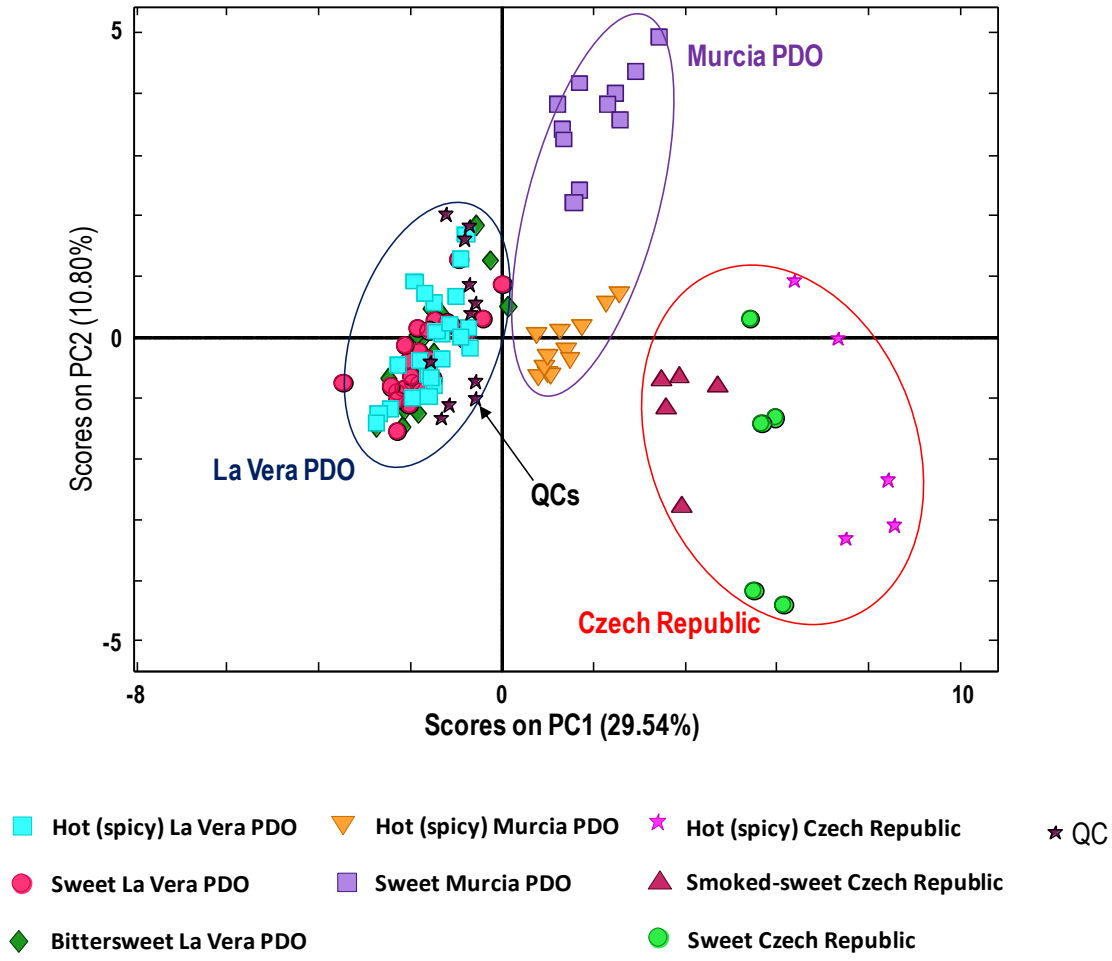


Figure 2. PCA score plot of PC1 *vs* PC2 when using the 36 compound concentrations found in the analyzed paprika samples as chemical descriptors.

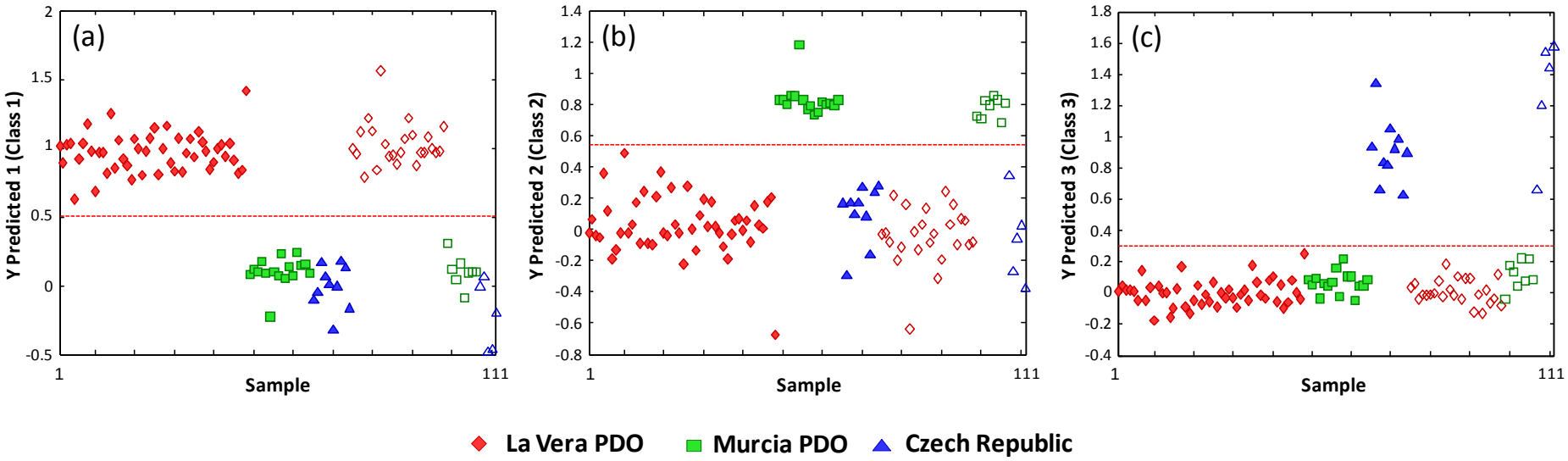
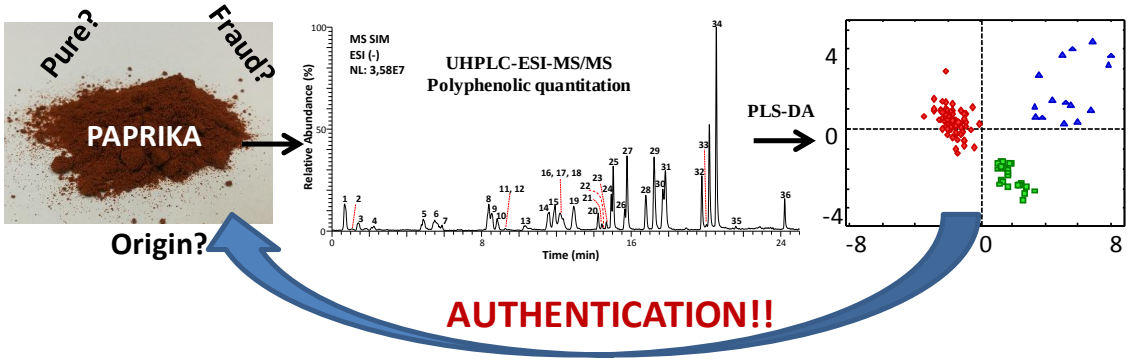


Figure 3. PLS-DA classification plots according to the production region. (a) La Vera versus other classes; (b) Murcia versus other classes; (c) Czech Republic versus other classes. Sample assignation: Rhombus = La Vera PDO, square = Murcia, triangle = Czech Republic. Dashed line means the classification boundary.


Pure?
Fraud?
PAPRIKA
Origin?
MS SIM
ESI (-)
NL: 3,58E7
UHPLC-ESI-MS/MS
Polyphenolic quantitation
Relative Abundance (%)
Time (min)
PLS-DA
AUTHENTICATION!!

TOC FIGURE

Supporting Information for:

Determination of Phenolic Compounds in Paprika by Ultra-High Performance Liquid Chromatography-Tandem Mass Spectrometry. Application to Product Designation of Origin Authentication by Chemometrics.

Sergio Barbosa[a,#], Guillem Campmajó[a,b,#], Javier Saurina[a,b], Lluis Puignou[a,b] and Oscar Núñez[a,b,c,*]

[a] Department of Chemical Engineering and Analytical Chemistry, University of Barcelona. Martí i Franquès 1-11, E08028 Barcelona, Spain.

[b] Research Institute in Food Nutrition and Food Safety, University of Barcelona. Av. Prat de la Riba 171, Edifici Recerca (Gaudí), E-08901 Santa Coloma de Gramanet, Barcelona, Spain.

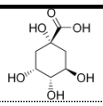
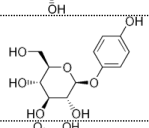
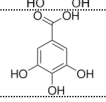
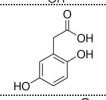
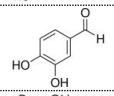
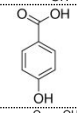
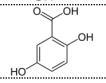
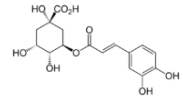
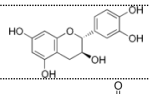
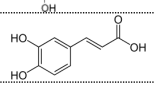
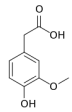
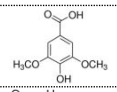
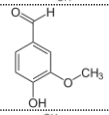
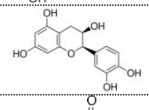
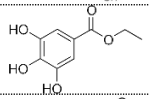
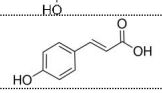
[c] Serra Hunter Fellow. Generalitat de Catalunya (Spain).

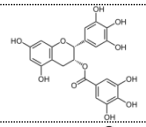
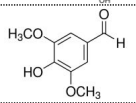
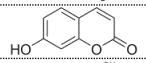
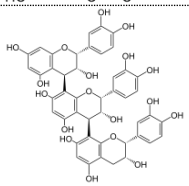
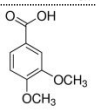
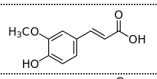
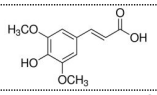
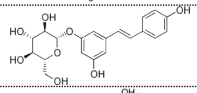
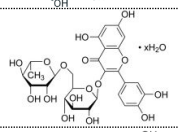
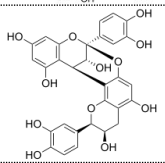
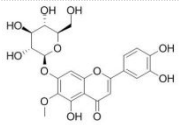
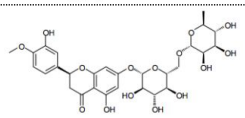
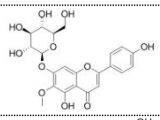
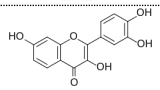
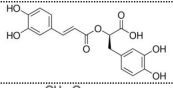
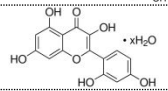
* Corresponding author: Oscar Núñez

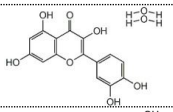
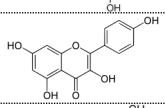
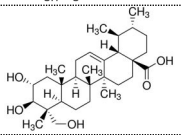
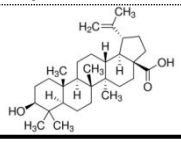
(Phone: 34-93-403-3706, Fax: 34-93-402-1233, e-mail: oscar.nunez@ub.edu)

[#] These authors contributed equally in the research of this work.

Table S1. Structures, family group, CAS number and supplier of the 36 polyphenolic compounds under study.

	Compound	Family	Structure	CAS number	Supplier
1	D-(−)-Quinic acid	Phenolic acid		77-95-2	Sigma-Aldrich
2	Arbutin	Other phenolics		497-76-7	Sigma-Aldrich
3	Gallic acid	Phenolic acid		149-91-7	Fluka
4	Homogentisic acid	Phenolic acid		451-13-8	Fluka
5	Protocatechuic aldehyde	Phenolic aldehyde		139-85-5	Sigma-Aldrich
6	4-Hydroxybenzoic acid	Phenolic acid		99-96-7	Sigma-Aldrich
7	Gentisic acid	Phenolic acid		490-79-9	Sigma-Aldrich
8	Chlorogenic acid	Phenolic acid		327-97-9	HWI Analytik GMBH
9	(+)-Catechin	Flavonoid		154-23-4	Fluka
10	Caffeic acid	Phenolic acid		331-39-5	Sigma-Aldrich
11	Homovanillic acid	Phenolic acid		306-08-1	Sigma-Aldrich
12	Syringic acid	Phenolic acid		530-57-4	Fluka
13	Vanillin	Phenolic aldehyde		121-33-5	Fluka
14	(-)-Epicatechin	Flavonoid		490-46-0	Sigma-Aldrich
15	Ethyl gallate	Other phenolics		831-61-8	Sigma-Aldrich
16	p-Coumaric acid	Phenolic acid		501-98-4	Sigma-Aldrich

17	(-)-Epigallocatechin gallate	Flavonoid		989-51-5	Sigma-Aldrich
18	Syringaldehyde	Phenolic aldehyde		134-96-3	Sigma-Aldrich
19	Umbelliferon	Other phenolics		93-35-6	Sigma-Aldrich
20	Procyanidin C1	Flavonoid		37064-30-5	Fluka
21	Veratric acid	Phenolic acid		93-07-2	Fluka
22	Ferulic acid	Phenolic acid		1135-24-6	Fluka
23	Sinapic acid	Phenolic acid		530-59-6	Sigma-Aldrich
24	Polydatin	Estilben		65914-17-2	Sigma-Aldrich
25	Rutin	Flavonoid		207671-50-9	Sigma-Aldrich
26	Procyanidin A2	Flavonoid		41743-41-3	Fluka
27	Nepetin-7-glucoside	Flavonoid		569-90-4	PhytoLab
28	Hesperidin	Flavonoid		520-26-3	Sigma-Aldrich
29	Homoplantaginin	Flavonoid		17680-84-1	PhytoLab
30	Fisetin	Flavonoid		345909-34-4	Sigma-Aldrich
31	Rosmarinic acid	Phenolic acid		2083-92-5	Sigma-Aldrich
32	Morin	Flavonoid		654055-01-3	Sigma-Aldrich

33	Quercetin	Flavonoid		6151-25-3	Riedel-de-Haën
34	Kaempferol	Flavonoid		520-18-3	Sigma-Aldrich
35	Asiatic acid	Phenolic acid		464-92-6	Sigma-Aldrich
36	Betulinic acid	Phenolic acid		472-15-1	Sigma-Aldrich

Sigma Aldrich (St. Louis, MO, USA); Fluka (Steinheim, Germany); HWI Analytic GMBH (Rülzheim, Germany); PhytoLab (Vestenbergsgreuth, Germany); Riedel-de-Haën (Seelze, Germany)

Table S2. Quality indicators for the calibration and prediction of the proposed PLS-DA model when dealing with the classification of paprika samples regarding the production region.

	Calibration model		
	La Vera PDO	**Murcia PDO**	**Czech Republic**
Calibration sensitivity	1	1	1
Calibration specificity	1	1	1
Calibration R^2	0.907	0.786	0.892
RSMEC	0.145	0.190	0.113
Calibration Bias	$-4 \cdot 10^{-16}$	$-3 \cdot 10^{-17}$	0
	Prediction model		
	La Vera PDO	**Murcia PDO**	**Czech Republic**
Prediction sensitivity	1	1	1
Prediction specificity	1	1	1
Prediction R^2	0.883	0.788	0.890
RMSEP	0.185	0.207	0.185
Prediction Bias	0.015	-0.082	0.067

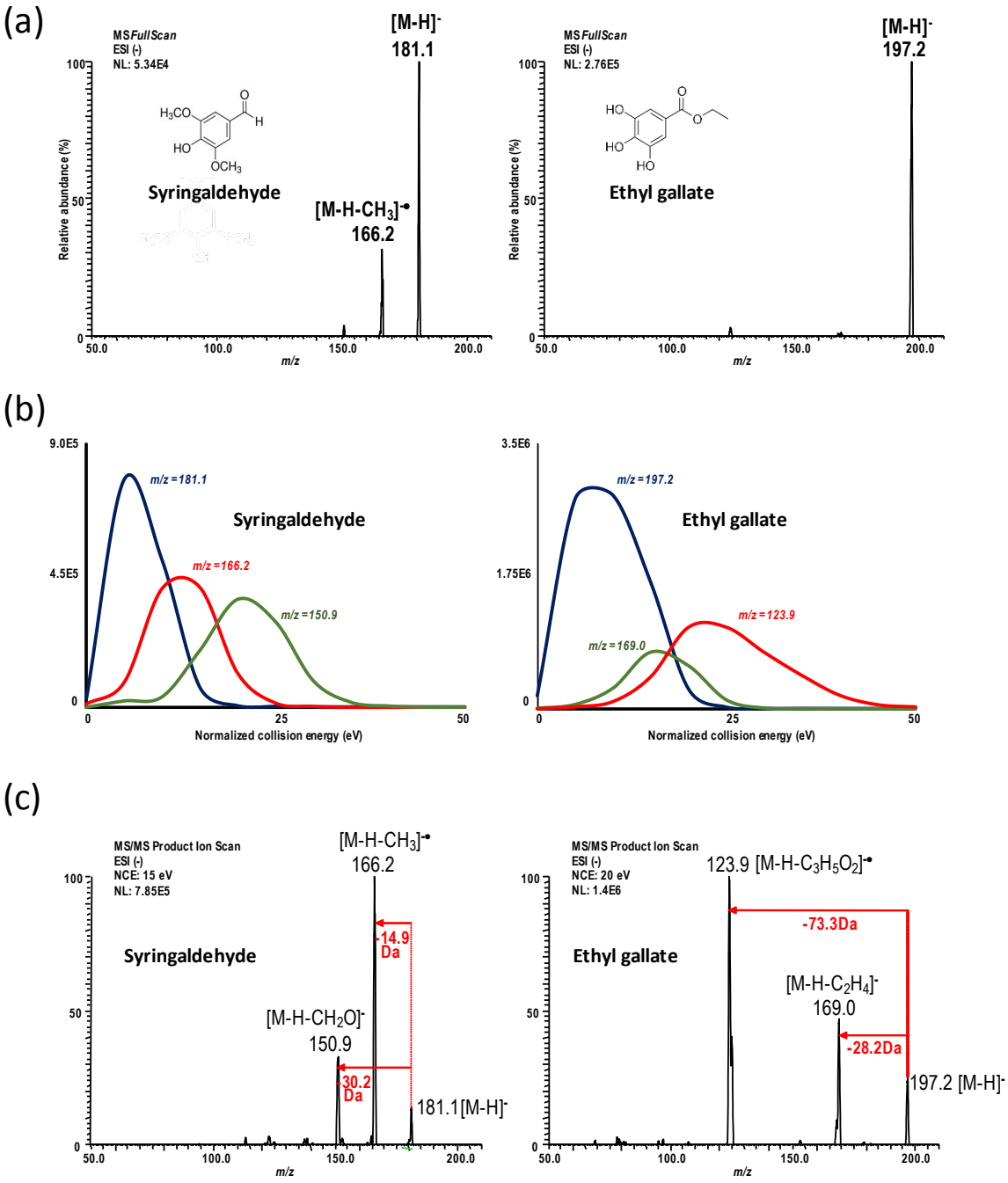


Figure S1. (a) MS full scan spectra, (b) collision energy curves, and (c) MS/MS product ion scan spectra for Syringaldehyde and Ethyl gallate.

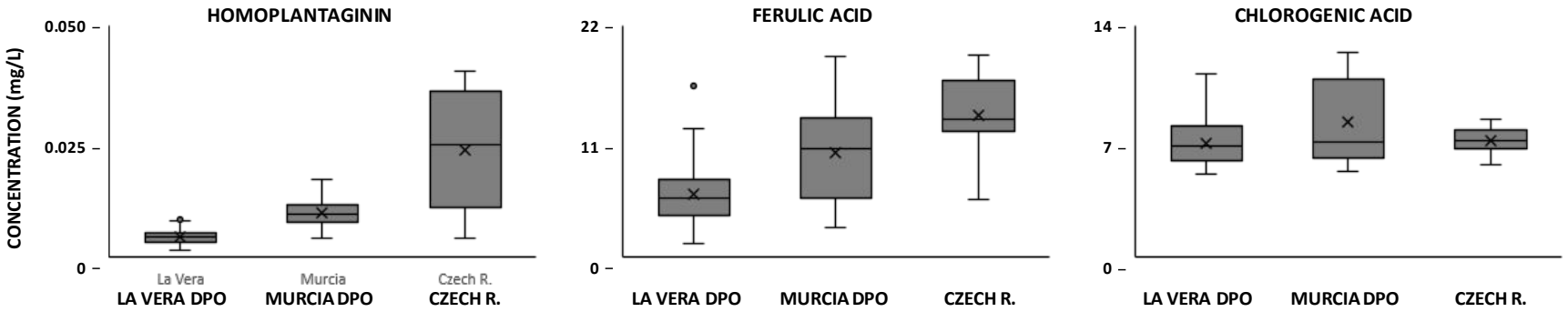


Figure S2. Boxplots with whiskers representing the concentration of selected compounds in the set of La Vera, Murcia and Czech Republic samples. From the left to the right: homoplantaginin, ferulic acid and chlorogenic acid.

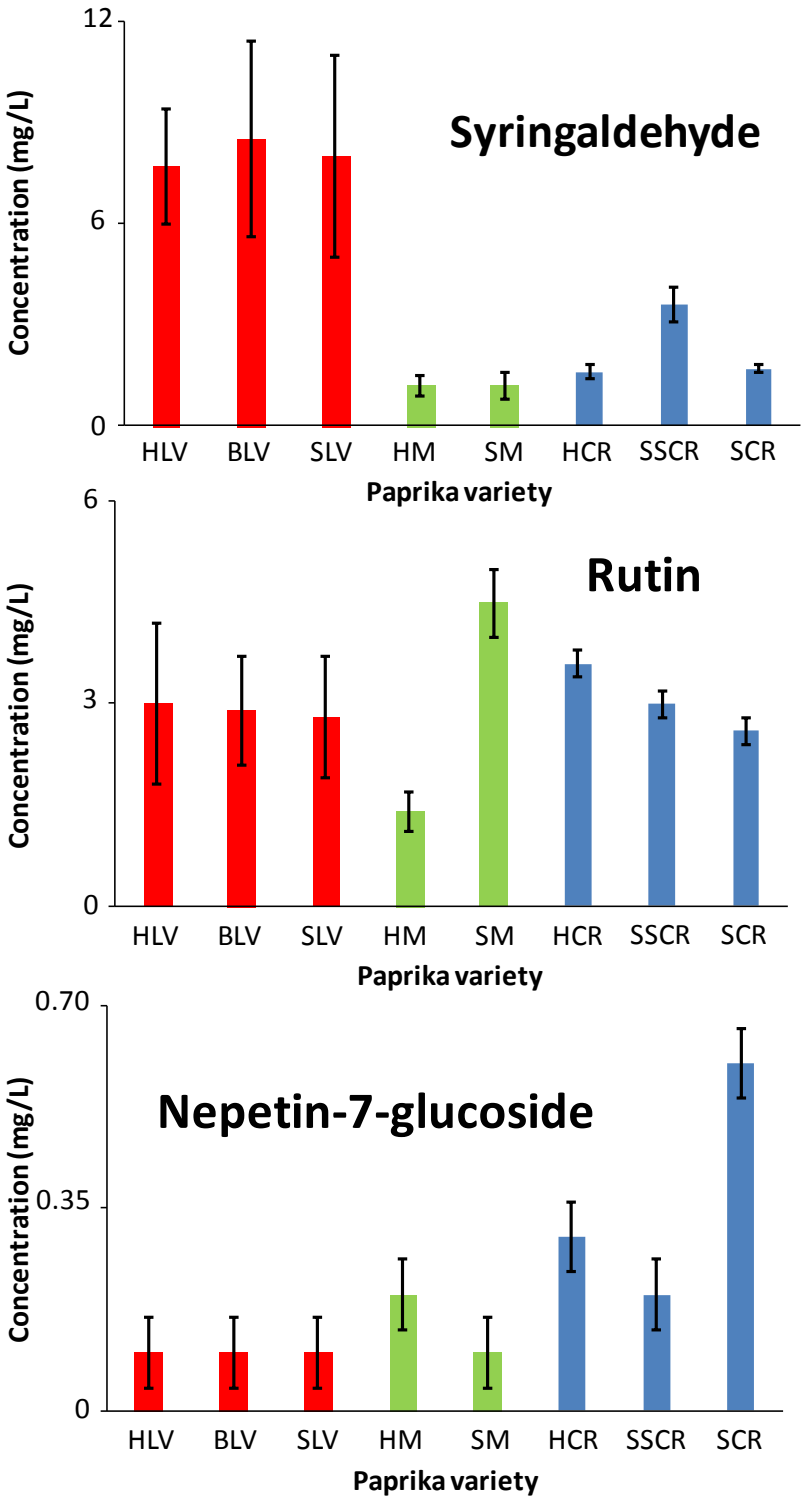


Figure S3. Bar plots showing concentration levels found in the analyzed paprika samples for syringaldehyde, rutin and nepetin-7-glucoside. HLV: hot (spicy) La Vera PDO; BLV: Bittersweet La Vera PDO; SLV: Sweet La Vera PDO; HM: Hot (spicy) Murcia PDO; SM: Sweet Murcia PDO; HCR: Hot (spicy) Czech Republic; SSCR: Smoked-sweet Czech Republic; SCR: Sweet Czech Republic.

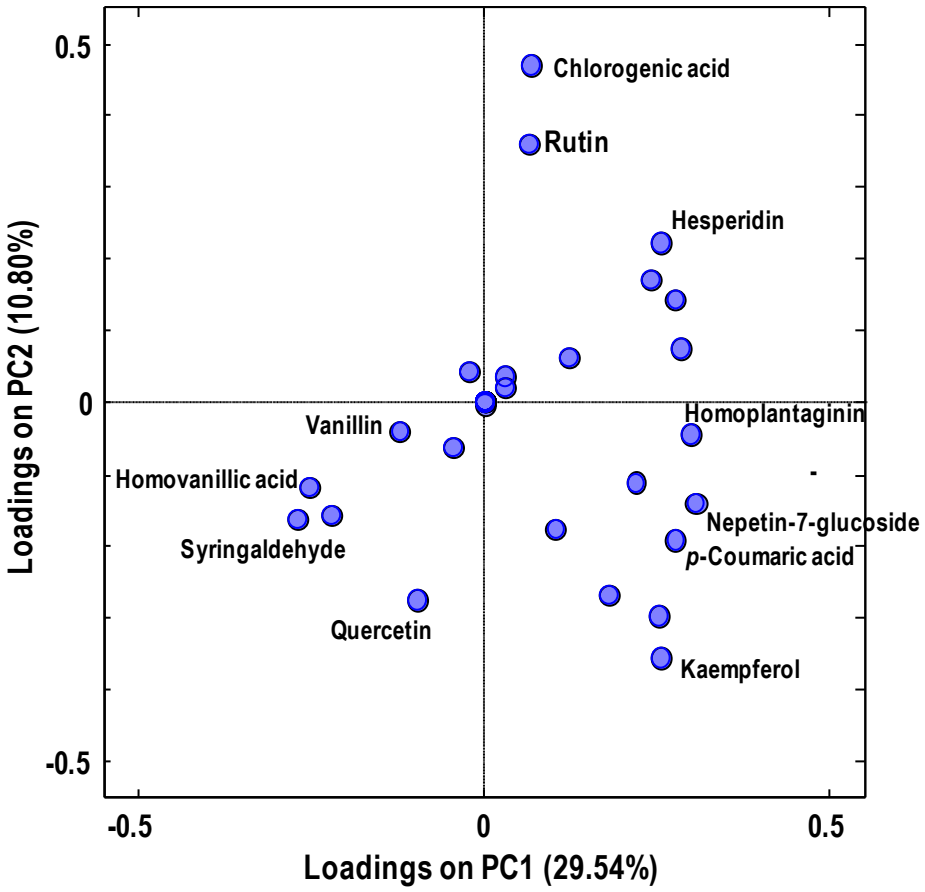


Figure S4. PCA loadings plot of PC1 *vs* PC2 when using the 36 compound concentrations found in the analysed paprika samples as chemical descriptors.

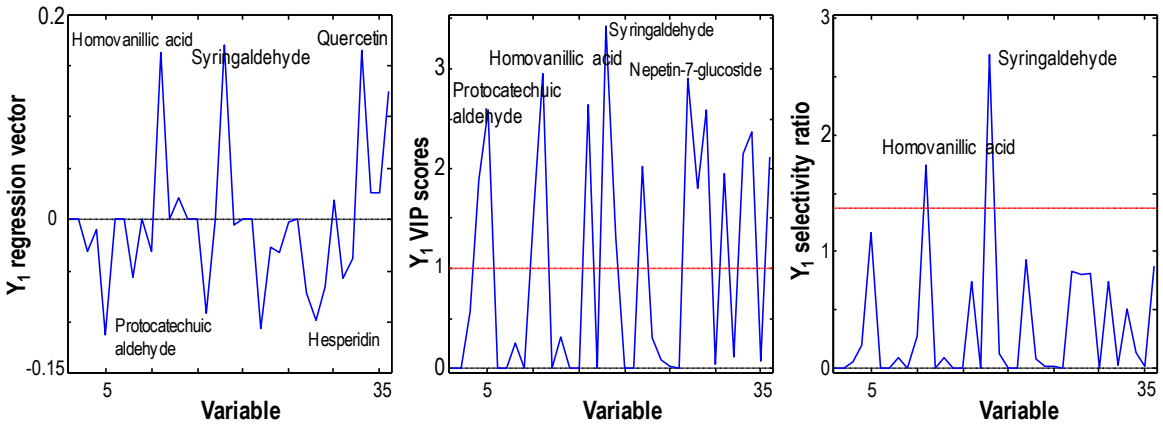


Figure S5. Plots of La Vera PDO class qualitative parameters (regression vector, the variable importance in projection (VIP) and the selectivity ratio) for the PLS-DA model obtained for the classification of paprika samples according to the production region (La Vera PDO, Murcia PDO and Czech Republic).

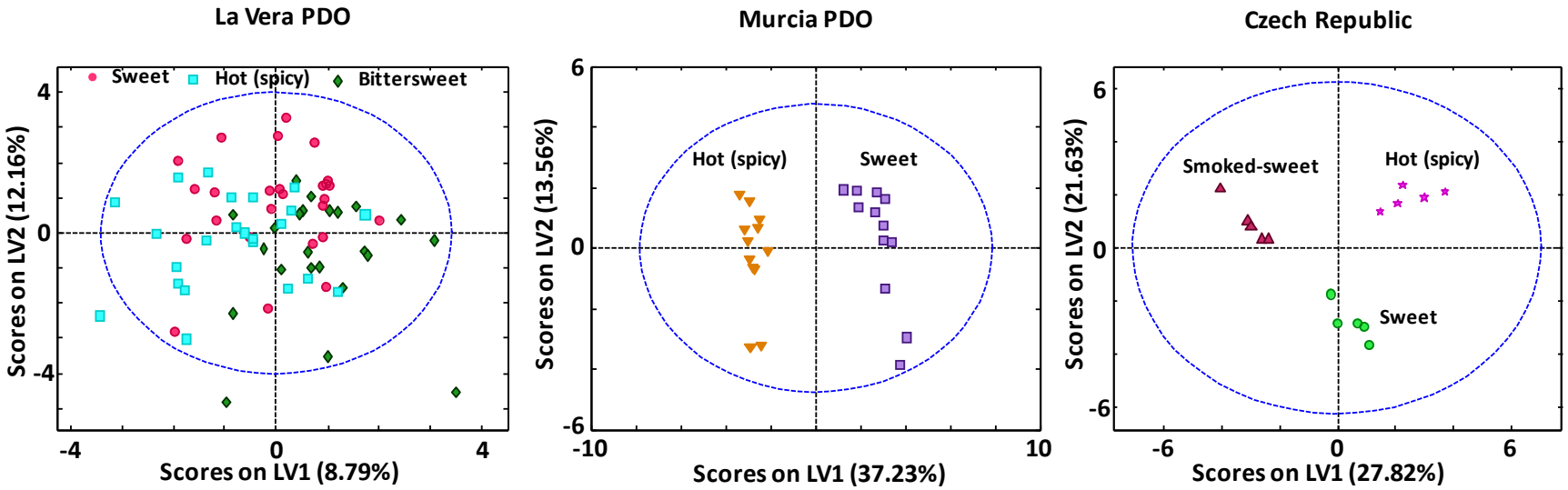


Figure S6. PLS-DA score plots of LV1 vs LV2 when using the 36 compound concentrations as chemical descriptors for the classification of each production regions (La Vera PDO, Murcia PDO and Czech Republic samples) according to their different flavor varieties.

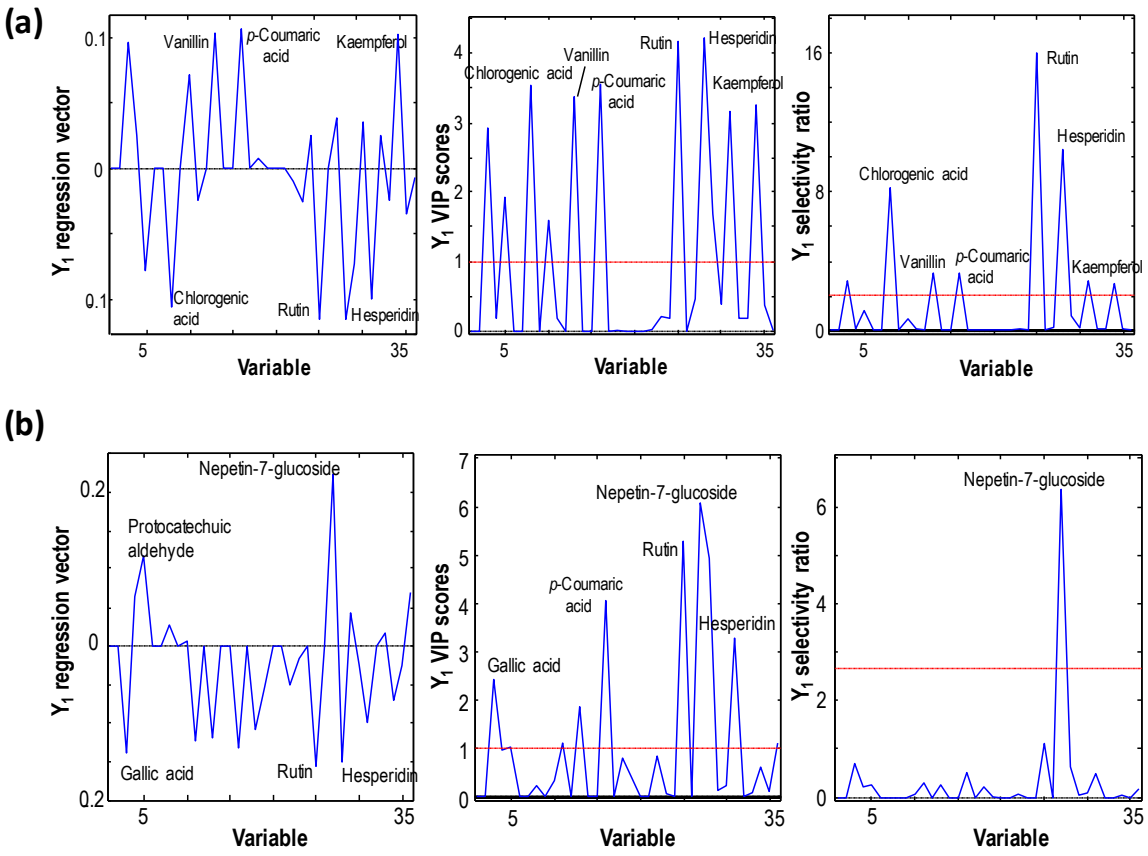


Figure S7. Plots of a) hot Murcia PDO and b) sweet Czech Republic paprika qualitative parameters (regression vector, the variable importance in projection (VIP) and the selectivity ratio) for the PLS-DA models obtained for the classification according to the different flavor varieties.