Determination of capsaicinoids and carotenoids for the characterization and geographical origin authentication of paprika by UHPLC–APCI–HRMS

Ane Arrizabalaga-Larrañaga[a,#], Guillem Campmajó[a,b,#], Javier Saurina[a,b], Oscar Núñez[a,b,c], Francisco Javier Santos[a], Encarnación Moyano[a,*]

[a] Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, Barcelona, Spain

[b] Research Institute in Food Nutrition and Food Safety, University of Barcelona, Santa Coloma de Gramenet, Spain

[c] Serra Húnter Program, Generalitat de Catalunya, Barcelona, Spain

* Corresponding author: Encarnación Moyano Morcillo

Department of Chemical Engineering and Analytical Chemistry
University of Barcelona

Av. Diagonal 645, E-08028, Barcelona, Spain

Tel.: +34-93-403-9277

E-mail: encarna.moyano@ub.edu

ORCID:
Ane Arrizabalaga-Larrañaga (0000−0002−9379−5775)
Guillem Campmajó (0000−0002−9252−0032)
Javier Saurina (0000−0002−4892−9535)
Oscar Núñez (0000−0001−5850−8972)
Francisco Javier Santos (0000−0002−8959−0363)
Encarnación Moyano (0000−0002−1233−8864)

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

Abstract

The production area mislabeling of a food product is considered a fraudulent practice worldwide. In this work, a method that uses ultra-high-performance liquid chromatography coupled to high-resolution mass spectrometry using atmospheric pressure chemical ionization (UHPLC–APCI–HRMS) was used for the geographical origin authentication of paprika based on the determination of capsaicinoids and carotenoids. Satisfactory instrumental method performance was obtained, providing good linearity (R^2 > 0.998), run-to-run and day-to-day precisions (%RSD < 15 and 10%, respectively), and trueness (relative errors < 10%), while method limits of quantification were between 0.21 and 51 $mg \cdot kg^{-1}$. Capsaicinoids and carotenoids were determined in 136 paprika samples, from different origins (*La Vera*, *Murcia*, Hungary, and the Czech Republic) and types (hot, sweet, and bittersweet). The composition of capsaicinoids and carotenoids was used as chemical descriptors to achieve paprika authentication through a classification decision tree built by partial least squares regression−discriminant analysis (PLS-DA) models and reaching a rate of 80.9%.

Keywords: Paprika; Capsaicinoids; Carotenoids; UHPLC-HRMS; Food authentication.

1. Introduction

Food authentication has become a concern for consumers, manufacturers, researchers, and international government administrations, due to the recent increase of food fraud, which implies illegal manipulation practices of foodstuff (e.g., adulteration, ingredient substitution, mislabeling, and dilution) with an economic gain purpose. It aims to certify intrinsic food properties, usually related to quality and safety, geographical origin, and production systems (Medina, Perestrelo, Silva, Pereira, & Câmara, 2019). Among food products, spices are at extremely high risk of food fraud ("Food Fraud Risk Information," 2020; Hong et al., 2017) because of their high cost and demand, as well as their complex supply chain. Other vulnerabilities, such as availability of the crops or weather events, also influence (Galvin-King, Haughey, & Elliott, 2018).

Paprika is a dried and ground spice obtained from different varieties of red pepper (genus *Capsicum* that belongs to the Solanaceae family). Its distinctive organoleptic properties, such as intense red color, characteristic aroma, and sometimes, a pungent flavor, make it widely used in international cuisines, although it is also employed in the cosmetic and pharmaceutical fields. Some of these properties are mainly related to bioactive substances named capsaicinoids and carotenoids. Moreover, these compounds have been found to gather human health beneficial aspects, being both anticarcinogenic substances, among others (de Sá Mendes & Branco de Andrade Gonçalves, 2020).

The worldwide production of paprika was estimated to be around four million tons in 2018, with Asia being the main producer ("Food and Agriculture Organization of the United Nations," 2019). Its production in Europe is mainly located in Spain and certain countries in Eastern Europe such as Hungary and the Czech Republic. Moreover, the European Commission on Agriculture and Rural Development distinguishes six European paprika products with the Protected Designation of Origin (PDO) ("European

Commission. eAmbrosia - the EU geographical indications register," 2020)]: *Pimentón de La Vera* (Spain), *Pimentón de Murcia* (Spain), *Kalocsai fűszerpaprika-őrlemény* (Hungary), *Szegedi fűszerpaprika-őrlemény* (Hungary), *Piment d'Espelette* (France), and *Paprika Žitava* (Slovakia). The presence of the PDO label ensures the geographical origin as well as the inherent qualities of the product. However, it is also related to higher prices, making them more vulnerable to fraudulent practices such as the mislabeling of the agricultural origin of paprika. Therefore, analytical methodologies to detect and prevent these frauds are needed.

In the last years, a large variety of analytical strategies combined with chemometrics —mostly using principal component analysis (PCA), linear discriminant analysis (LDA), and partial least squares regression−discriminant analysis (PLS−DA)— have been developed to address the authenticity of paprika origin. For instance, some authors have proposed multi-elemental content profiling, determined by both inductively coupled plasma optical emission spectroscopy (IPC−OES) or mass spectrometry (ICP−MS), for the authentication of *Szegedi fűszerpaprika* PDO (Brunner, Katona, Stefánka, & Prohaska, 2010), the comparison of hot and sweet Hungarian paprika (Ördög et al., 2018), and the discrimination between *La Vera* and *Murcia* denominations (Ana Palacios-Morillo, Jurado, Alcázar, & De Pablos, 2014). Instead, other techniques such as spectrophotometric measurements(A. Palacios-Morillo, Jurado, Alcázar, & Pablos, 2016) or the combination of different parameters (e.g., sample moisture, elemental analysis, and total ash, lipids, nitrogen, saccharides content)(Václav Štursa , Pavel Diviš, 2018) have also been evaluated. Alternatively, several chromatographic fingerprinting approaches using high-performance liquid chromatography with electrochemical detection (HPLC/ECD)(Serrano et al., 2018) or ultraviolet detection (HPLC/UV) (Cetó, Sánchez, Serrano, Díaz-Cruz, & Núñez, 2020; Cetó et al., 2018), and ultra-high-performance liquid

chromatography coupled to high-resolution mass spectrometry (UHPLC−HRMS) (Barbosa, Saurina, Puignou, & Núñez, 2020), have recently focused on *La Vera* and *Murcia* PDO discrimination and adulteration detection.

Chemical profiling based on the determination of targeted compounds by liquid chromatography-mass spectrometry (LC–MS) has also been exploited to authenticate paprika according to its agricultural origin. The presence, distribution, and content of bioactive substances is directly related to many food features, such as the production area. Thus, they are commonly used as chemical descriptors for classificatory purposes through a semi-quantification (Campmajó, Núñez, & Núñez, 2019). To date, ultra-high-performance liquid chromatography coupled to tandem mass spectrometry (UHPLC–MS/MS) for targeted polyphenols and UHPLC–HRMS for polyphenols and capsaicinoids (Barbosa, Saurina, Puignou, & Nuñez, 2020), and polyphenols and carbohydrates (Mudrić et al., 2017), have also been evaluated for paprika classification. Thereby, although capsaicinoid and carotenoid content has been extensively studied in red pepper and its derived products (Giuffrida et al., 2013; Nagy, Daood, Koncsek, Molnár, & Helyes, 2017), their simultaneous analysis has not yet been used to deal with the classification of paprika. Therefore, this study aimed to develop an UHPLC–HRMS method for the determination of capsaicinoids and carotenoids in European paprika, and the subsequent use of target compound composition for the geographical origin authentication by multivariate chemometric methodologies.

2. Experimental

2.1. Reagents and materials

Chemical formula, acronyms, and chemical structures of target capsaicinoids and carotenoids are summarized in Fig. 1 and they were purchased from Sigma-Aldrich (Steinheim, Germany) with purities higher than 90%.

Individual stock standard solutions of capsaicinoids (1,000 mg·L^{-1}) were prepared in LC–MS grade methanol, except capsaicin and dihydrocapsaicin that were prepared in ethanol, while carotenoid were prepared in acetonitrile (500 mg·L^{-1}). Intermediate mixture containing all target compounds (50 mg·L^{-1}) was weekly prepared from stock solutions by appropriate dilution in acetonitrile:acetone (1:1, *v:v*) and was subsequently used to obtain calibration solutions (0.001 to 10 mg·L^{-1}) for quantification. All stock solutions were stored at -20 ºC until their use.

Acetone for pesticide residue analysis (≥ 99.8%), LC–MS grade water, methanol, and acetonitrile were purchased from Sigma-Aldrich, whereas absolute ethanol for analysis was obtained from Panreac (Barcelona, Spain). Moreover, a 0.22 µm pore size Nylon membrane (Whatman, Clifton, NJ, USA) was employed to filter mobile phase components before their use.

2.2. Instrumentation

An UHPLC system equipped with an Accela 1250 quaternary pump, an Accela autosampler, and a column oven (Thermo Fisher Scientific, San Jose, CA, USA) was used for the chromatographic separation. Accucore C_{18} analytical column (100 mm × 2.1 mm id., 2.6 µm particle size) and guard column (10 mm × 2.1 mm id., 2.6 µm particle size), both packed with superficially porous particles, were employed for the chromatographic separation of both carotenoid and capsaicinoid families. The developed chromatographic method used a quaternary gradient elution program with water, methanol, acetonitrile, and acetone as solvent A, B, C, and D, respectively. After optimization of the chromatographic separation (see Section 3.2) the gradient elution program used in this study started with a 3 min isocratic step at 60% solvent A and 40% solvent C and followed by a linear gradient elution up to 80% solvent C in 0.5 min, and an isocratic step at these last conditions for 2.5 min. Later, solvent B was introduced, and the mobile

phase was linearly changed to 10% solvent B and 90% solvent C in 1.25 min, keeping in these conditions for 3 min. Afterward, another linear gradient elution changed the composition in 1 min up to 50% solvent C and D and kept at isocratic conditions for 1.5 min. Finally, solvent D was linearly increased up to 80% in 3 min, and this last percentage was used in an isocratic step for 2 min, before turning back to the initial conditions. The mobile phase flow rate was 600 $\mu L \cdot min^{-1}$, the injection volume was 10 µL, and the column oven temperature was 25°C.

The UHPLC system was coupled to a hybrid quadrupole-Orbitrap mass spectrometer (Q-Exactive Orbitrap, Thermo Fisher Scientific) equipped with an atmospheric pressure chemical ionization (APCI) source (positive-ion mode). Nitrogen was purchased from Linde (Barcelona, Spain) and used as a sheath, sweep, and auxiliary gas at flow rates of 60, 0, and 40 a.u. (arbitrary units), respectively. Both vaporizer and capillary temperatures were set at 350 °C, corona discharge current at +6 kV and SLens RF level at 70 V. The Q-Exactive Orbitrap system was tuned and calibrated every three days, using a calibration solution for positive-ion mode. The HRMS instrument operated in full scan MS mode (*m/z* 50 – 700) at a mass resolution of 70,000 full width at half maximum (FWHM) at *m/z* 200. Moreover, an automatic gain control of 3.0×10^{-6} and a maximum injection time of 200 ms was used. For the analysis of samples, two-events acquisition mode was used: an MS full scan and an "all-ion fragmentation" (AIF) (*m/z* 50 – 700, in both events) with stepped normalized collision energies (NCE) of 20, 30, 40 eV for ion fragmentation. The Xcalibur software *v* 4.1 (Thermo Fisher Scientific) was used to control the LC–MS system and to acquire and process data.

2.3. Sample analysis

A total of 136 paprika samples from different origins and types were purchased and analyzed in this work. They were produced in Spain (*La Vera* and *Murcia*), Hungary

147 and the Czech Republic; regarding types, hot, bittersweet, and sweet paprika were
148 considered. Table 1 summarizes sample details such as the acronyms used for each region
149 and the number of samples analyzed for each type of sample.

150 A simple solid-liquid extraction of target analytes from paprika samples was carried
151 out as follows: 0.05 g of paprika were extracted with 4 mL of methanol:acetone (1:1, *v/v*)
152 solution in a 15 mL PTFE tube. Subsequently, the sample was stirred in a Stuart Vortex
153 for 0.5 min (Staffordshire, United Kingdom) and sonicated for 10 min (5510 Branson
154 ultrasonic bath, Hampton, NH, USA). Afterward, the extract was centrifuged for 15 min
155 at 4,500 rpm (ROTANTA 460 HR Centrifuge, Hettich, Germany). Finally, the
156 supernatant was filtered through 0.22 µm Nylon membrane filters and stored at 4 ºC in 2
157 mL glass injection vials until the analysis by UHPLC-HRMS.

158 *2.4. Instrumental and quality parameters*

159 Instrumental and method limits of detection (ILODs, MLODs) were estimated as
160 the smallest analyte concentration, providing a well-defined chromatographic peak with
161 a good peak shape. This criterion was used because of the absence of baseline noise in
162 the extracted ion chromatograms using a narrow mass tolerance window (<5 ppm) under
163 high-resolution mass spectrometry conditions (FWHM 70,000 at *m/z* 200) on the Orbitrap
164 mass analyzer. Instead, instrumental and method limits of quantification (ILOQs,
165 MLOQs) were calculated from LOD values and considering the established ratio of three
166 to ten between LODs and LOQs. In this way, ILODs have been determined using standard
167 solutions in solvent injected directly into the UHPLC–HRMS system, whereas MLODs
168 were calculated considering the sample treatment recovery and the matrix effect. Besides,
169 both precision and trueness were studied by analyzing in triplicate two standard solutions
170 at low and medium level concentrations, being near and around ten times higher than the
171 LOQs, respectively. Precision (run-to-run and day-to-day) was expressed as the relative

172 standard deviation (RSD, %), whereas trueness was defined as the relative error (RE, %),
173 both calculated according to the obtained concentrations.

174 Due to the lack of a blank paprika (free of target analytes), matrix effect (ME, %) in
175 the UHPLC–APCI–HRMS method was evaluated by spiking a sweet paprika from the
176 Czech Republic (which presented the lowest concentration of target compounds) at 1
177 $mg \cdot kg^{-1}$. This concentration was three times higher than the endogenous one determined
178 previously in the same sample. Thus, the ME in the ionization process was evaluated by
179 estimating the relative difference between the chromatographic peak area obtained in the
180 analysis of the spiked extract and that obtained from the analysis of standard mixtures at
181 the same concentration level.

182 To ensure the quality of the results and check the reproducibility of the LC separation
183 and sensitivity of the UHPLC−APCI−HRMS system, a solution of a mixture of standards
184 and procedural blanks were included within the sample batch when analyzing calibration
185 curves and samples.

186 *2.5. Data analysis*

187 Solo 8.6 chemometrics software from Eigenvector Research (Manson, WA, USA)
188 was used to perform data PCA and PLS-DA and employ the hierarchical model builder
189 (HMB).

190 PCA relies on the concentration of the dataset's relevant information, originally
191 contained in the compositional profiles of capsaicinoids and carotenoids, into a reduced
192 number of principal components (PCs). Such concentration values are arranged in the X-
193 matrix, which is mathematically decomposed into the submatrices of scores T
194 (coordinates of the samples) and loadings PT (eigenvectors), providing information on
195 the distribution of samples and variables, respectively. Moreover, the detection of

potential outlier samples bases on the distance to the center of the model calculated from the Hoteling T^2 and Q statistical parameters, being T^2 the sum of the normalized squared scores and Q the sum of squares of residuals of a given sample.

In this study, PLS-DA has been used as the classification method. The PLS-DA model is built from a training set composed of well-known paprika samples belonging to the different classes to be assessed. At this stage, PLS-DA assigns each sample into a class (numerically encoded depending on the origin and type), following rules based on the distance to the center of each class, calculated from T^2 and Q. The classification model is established to reach the minimum prediction error in assigning these calibration samples into their actual classes.

More details of the theoretical background of these chemometric techniques are addressed elsewhere (Massart, D. L., Vandeginste, B. G. M., Buydens, L. M. C., de Jong, S., Lewi, P.J., & Smeyers-Verbeke, 1997).

PCA and PLS-DA X-data matrices consisted of the target compounds' concentration levels as a function of the paprika samples under study, while PLS-DA Y-data matrices defined the membership of each sample in a class. Before building the chemometric model, data was autoscaled to provide the same weight to each variable by suppressing differences in their magnitude and amplitude scales. Moreover, the most suitable number of latent variables (LVs) in PLS-DA was established at the first significant minimum point of the cross-validation (CV) error. Venetian blinds were set by default as the CV method, except for small data matrices (less than twenty paprika samples), where the leave-one-out method were employed. Moreover, considering the complexity of the studied issue, where several sample origins and types were presented, the classification has not been obtained from the segregation of all the classes at once but sequentially using HMB. Therefore, different PLS-DA models were consecutively combined, breaking

down the classification aim into sub-groups. The applicability of the built chemometric method was evaluated through external validation: 70% of a sample group was used as the training set (data set used for model generation and optimization), and the remaining 30% as the test set.

A quality control (QC) sample, consisting of a mix prepared with 50 µL of each paprika sample extract, was used to control the repeatability and robustness of the chemometric results as well as to detect systematic errors. In this line, samples were also randomly injected to minimize the influence of instrumental drifts in the models.

3. Results and discussion

3.1. HRMS and AIF (HRMS) characterization of targeted capsaicinoids and carotenoids

In the present work, four capsaicinoids (nordihydrocapsaicin, NDC; capsaicin, CAP; dihydrocapsaicin, DC; nordihydrocapsiate, NDCT) and six carotenoids (capsanthin, CT; capsorubin, CR; violaxanthin, VIO; lutein, LUT; β–cryptoxanthin, ß–CRYPT; β–carotene, ß–CAR) were determined by UHPLC−APCI−HRMS in paprika samples. These compounds are commonly found in red pepper-derived products (Arimboor, Natarajan, Menon, Chandrasekhar, & Moorkoth, 2015; Schweiggert, Carle, & Schieber, 2006) and their structures are depicted in Fig. 1.

The ions generated by APCI for targeted compounds were studied using a hybrid high-resolution mass spectrometer (quadrupole-Orbitrap) equipped with a high-energy collision dissociation (HCD) cell. This instrument allows monitoring ions at HRMS and fragmenting them to provide more specific chemical structural information useful for confirmatory purposes. Thus, the mass spectral information of ions generated in APCI (positive-ion mode) are summarized in Table 2. The mass spectra of CAP, DC, and NDC

showed the protonated molecule $[M+H]^+$ as base peak, and they did not show any adduct ion. Nevertheless, an intense signal at *m/z* 137.0597 (Rel Ab. 20–70%) always appeared due to the in-source CID fragmentation of the protonated molecule because of the β-cleavage at the N-R bond. (Reilly et al., 2003). In addition, ions at *m/z* 170.1536 (CAP), *m/z* 172.1693 (DC), and *m/z* 158.1537 (NDC), were assigned to a common loss (136.0518 Da) from the protonated molecule $[M+H-C_8H_8O_2]^+$, which corresponded to the fraction of the acyl chain that results from removing the aromatic ring (Schweiggert et al., 2006). Instead, the mass spectrum of NDCT showed the in-source collision-induced dissociation (CID) fragment ion at *m/z* 137.0597 as base peak because, after the above mentioned β-cleavage, the charge remained in the common fragment $[C_8H_9O_2]^+$. Nevertheless, although most of the carotenoids also showed the $[M+H]^+$ as the base peak, a significant in-source CID fragmentation where a water molecule is lost $[M+H–H_2O]^+$ was observed in some cases (CR, *m/z* 583.4137; VIO, *m/z* 583.4137; CT, *m/z* 567.4186; ß–CRYPT, *m/z* 535.4291; LUT, *m/z* 551.4239). Moreover, this in-source CID fragment ion was the base peak of LUT and CR, as displayed in other studies (Arrizabalaga-Larrañaga, Rodríguez, Medina, Santos, & Moyano, 2020).

The UHPLC–APCI–HRMS method was carried out using independent data analysis based on two scanning events – HRMS full scan and all ion fragmentation (AIF) – to improve detectability and obtain structural information of target analytes. Regardless of the compound fragmentation, to obtain a rich AIF mass spectrum within the whole *m/z* range studied, the full scan of fragment ions was performed by employing stepped normalized collision energies (NCE: 20, 30, 40 eV). In this way, it provided the average of AIF (HRMS) mass spectra at the different collision energies. Fig. 2 shows the HRMS spectrum and AIF (HRMS) spectrum of (A) DC and (B) CT.

The AIF (HRMS) spectrum was obtained for all targeted compounds and the diagnostic fragment ions, the corresponding ion assignments, and the accurate mass errors are summarized in Table 2. Each family of compounds showed a distinctive fragmentation pathway. For instance, all capsaicinoids showed common fragment ions *m/z* 137.0597, *m/z* 122.0362, *m/z* 94.0413 and *m/z* 66.0464 (Fig. 2). The fragment ion at *m/z* 122.0362 $[C_7H_6O_2]^{+\bullet}$ was produced by the α–cleavage of the C-O bond, generating the dissociation of the methylene moiety from the fragment ion at *m/z* 137.0597 $[C_8H_9O_2]^+$ (Wolf, Huschka, Raith, Wohlrab, & Neubert, 1999). Moreover, the ion at *m/z* 122.0362 $[C_7H_6O_2]^{+\bullet}$ can be further fragmented through neutral losses of CO (27.9943 Da) to form both fragment ions at *m/z* 94.0413 $[C_7H_6O_2\text{-}CO]^{+\bullet}$ and *m/z* 66.0464 $[C_7H_6O_2\text{-}C_2O_2]^{+\bullet}$. These fragmentation steps may involve the opening of the aromatic ring, yielding into these linear polyunsaturated chain ions. On the other hand, carotenoids presented other characteristic common fragment ions such as $[C_{11}H_{13}]^+$ (*m/z* 145.1012), $[C_9H_{11}]^+$ (*m/z* 119.0855), and $[C_8H_9]^+$ (*m/z* 105.0699), which were generated because of the fragmentation of the high polyene conjugation. In addition, CR and VIO isomers showed the same fragment ion $[C_{15}H_{21}O_2]^+$ (*m/z* 221.1536) corresponding to the oxo-cycle fused to the 3-hydroxy-β-ring and produced by the cleavage between carbons 10 and 11 (Wolf et al., 1999). Moreover, the fragment ion $[C_8H_{13}]^+$ (*m/z* 109.1011) presented in both AIF (HRMS) spectrum of CR and CT (Fig. 2) corresponded to the dehydrated five-membered ring (Breemen, Dong, & Pajkovic, 2012).

3.2. UHPLC–HRMS method development

The chromatographic separation of all target compounds was performed in a reversed-phase UHPLC Accucore C_{18} column, under a quaternary gradient elution with water, methanol, acetonitrile, and acetone as the mobile phase components. The gradient elution was based on a chromatographic method previously developed for the separation

of chlorophylls and carotenoids (Arrizabalaga-Larrañaga, Rodríguez, Medina, Santos, & Moyano, 2019). However, some modifications were required to deal with the simultaneous determination of capsaicinoids and carotenoids. Hence, given the differences in polarity among both families of compounds, the water content of the mobile phase at the beginning of the gradient elution was increased to ensure an effective separation of the most polar capsaicinoids (Daood et al., 2015). Thus, an isocratic step of water:acetonitrile (60:40, *v/v*) was included as starting elution conditions followed by a linear gradient up to 20:80 to retain capsaicinoids and allow their elution after four-fold the hold-up time (t_M), which corresponded to 0.97 min, and before carotenoids. The inclusion of CR and CT among the carotenoid compounds made necessary to lengthen the isocratic step of methanol:acetonitrile (10:90, *v/v*). Moreover, the mobile phase eluotropic strength had to be increased at the end of the chromatographic run using acetonitrile:acetone (50:50, *v/v*) to allow the elution of β-CAR, the most hydrophobic carotenoid. Under the final gradient elution (*see section 2.2.*), a baseline separation of all target compounds was achieved in less than 15 minutes, except for CAP and NDC, which partially co-eluted. However, the isotope cluster of their ions did not overlap; thus, they could be isolated in individual extracted chromatograms according to *m/z*. Besides, the study of ion suppression or ion enhancement for these co-eluting compounds was carried out by injecting individual standard solutions and a mixture of the co-eluting target compounds (1 $mg \cdot L^{-1}$) in the UHPLC–APCI–HRMS. The difference of the obtained chromatographic peak areas was lower than 10%, similarly to the RSD% observed between successive injections, which indicated that the co-elution of these compounds did not affect their responses.

The performance of the developed UHPLC–APCI–HRMS method was evaluated by determining the linearity, ILODs, ILOQs, precision, and trueness. The linearity within

319 the concentration range, 0.001-10 mg·kg^{-1} for most of the compounds and 0.1-10 mg·kg^{-1}
320 for ß–CRYPT and LUT, was satisfactory and showed correlation coefficients (R^2)
321 higher than 0.998. ILODs ranged from 0.001 to 0.025 mg·kg^{-1} for most of the target
322 compounds, although for ß–CRYPT and LUT values were slightly higher (0.1 and 0.25
323 mg·kg^{-1}, respectively). In terms of RSD and based on concentration values, run-to-run
324 and day-to-day precision were always lower than 15% and 10%, respectively. Moreover,
325 the trueness, based on the same concentration values, showed relative errors below 10%.
326 These results demonstrated the good instrumental performance of the developed
327 UHPLC–APCI–HRMS method for the determination of capsaicinoids and carotenoids.

328 Besides, before the determination of capsaicinoids and carotenoids by UHPLC–APCI–
329 HRMS in paprika, samples were submitted to a solid-liquid extraction. Because of the
330 differences in the physicochemical properties of both families of compounds, three
331 commonly used solvents, methanol, acetonitrile, and acetone, as well as mixtures of them,
332 were evaluated to achieve the most effective simultaneous extraction of target
333 compounds. It was found that acetonitrile had less effectiveness in extracting carotenoids
334 than both pure acetone and the mixture methanol:acetone. Moreover, pure methanol
335 extracted more efficiently capsaicinoids, than pure acetonitrile or acetone. Nevertheless,
336 the combination of both methanol and acetone seemed to improve the solubility of these
337 compounds, and thus, as a compromise, a mixture methanol:acetone (1:1, *v/v*) was chosen
338 as the most effective solvent for the simultaneous extraction of both capsaicinoids and
339 carotenoids (*section 2.3.*) in agreement with Nagy et al. who proposed a similar solvent
340 mixture (Nagy et al., 2017). Using the proposed extraction procedure, estimated MLODs
341 ranged from 0.06 to 1.5 mg·kg^{-1} for most of the analytes, except for ß–CRYPT and LUT,
342 which were 6.1 and 15.3 mg·kg^{-1}, respectively. While, MLOQs were comprised between
343 0.21 and 51 mg·kg^{-1}.

3.3. Analysis of paprika samples

In this work, to test the potential of the UHPLC–APCI–HRMS method to determine capsaicinoids and carotenoids for authentication purposes, a total of 136 paprika samples from different regions were analyzed. Samples from countries such as Spain (*La Vera* and *Murcia*), Hungary, and the Czech Republic, as well as distinct flavor types (hot, sweet, and bittersweet), were evaluated.

Matrix-effect in the ionization of target compounds was evaluated as described in section 2.4 and the results showed ME% values from 10 to 50%. These results indicated that analytical correction strategies for accurate quantitative results should be performed. In this line, matrix-matched calibration cannot be applied to the determination of endogenous bioactive compounds because of the lack of blank samples. Instead, although standard addition calibration and isotope dilution mass spectrometry (IDMS) allow the correction of the matrix effect, they are not suitable for this study since standard addition calibration is time-consuming for the analysis of large sample batches, and IDMS requires expensive internal labeled standards, which are not available for all the target compounds. Therefore, these drawbacks make it difficult to apply these strategies to obtain an accurate quantitative analysis of capsaicinoids and carotenoids in paprika samples. Instead, some published studies have proposed to extract the targeted compounds from the food matrix to obtain blank samples that are proposed to be used in matrix-matched calibration. However, this strategy completely modifies the original food matrix, and thus, its application was not considered in this study. Therefore, external calibration methods are commonly proposed in most of the published studies dealing with the determination of these families of compounds in food and natural samples. For instance, capsaicinoids and carotenoids in paprika have been determined by some authors using only one or two available standards because of the chemical structural similarities (Barbero, Liazid,

Ferreiro-González, Palma, & Barroso, 2016; Bijttebier et al., 2014; Stipcovich, Barbero, Ferreiro-González, Palma, & Barroso, 2018). Moreover, since the present study aimed to determine capsaicinoids and carotenoids for their use as chemical descriptors for paprika authentication, and the matrix influence could contribute as a potential source of discrimination between samples, external standard calibration method by employing ten standards was performed for the analysis of paprika samples. Thereby, the results obtained for the presence of both capsaicinoids and carotenoids in the 136 paprika samples analyzed are summarized in Table S1 (Supporting Material).

The qualitative capsaicinoid and carotenoid patterns (UHPLC–APCI–HRMS chromatograms) observed for all paprika samples were similar in terms of compounds detected, but they showed differences in the corresponding abundances. As an example, the diversity of the capsaicinoid and carotenoid profile is shown in Fig. 3, depicting the extracted UHPLC–APCI–HRMS chromatograms obtained from the analysis of a sweet paprika sample from (A) *Murcia* "MS9" and (B) Hungary "HS5". To better study the relationship between their concentration and the type and production country of the samples, the total capsaicinoid and carotenoid contents, as well as the capsaicinoid/carotenoid ratio were evaluated. (Table S2 and Fig. S1).

For instance, independently of the geographical origin, hot paprika showed a higher total capsaicinoid content, 656 ± 453 $mg \cdot kg^{-1}$, and hence a higher capsaicinoid ratio (40-90%), than sweet and bittersweet samples, 9 ± 5 and 31 ± 32 $mg \cdot kg^{-1}$, respectively. This result was expected since these target compounds are responsible for the characteristic hot taste (de Sá Mendes & Branco de Andrade Gonçalves, 2020). Besides, within a specific flavor type, the capsaicinoid/carotenoid ratios between non-smoked and smoked samples showed similar behavior (Table S1). Thus, they were jointly considered in the subsequent studies. Regarding individual target compounds, among capsaicinoids, DC

and CAP were found in major concentrations within all hot, sweet, and bittersweet samples, whereas NDCT was not detected in any sample above its MLOD.

The carotenoid content usually did not significantly differ when comparing the different types (hot, sweet, and bittersweet) of samples from the same region (Table S2). Hungarian samples had the highest total content of carotenoids, independently of the flavor type. For instance, the total carotenoid amounts of hot *La Vera*, *Murcia*, and the Czech Republic paprika samples were 106 ± 51, 118 ± 69, and 75 ± 24 $mg \cdot kg^{-1}$, respectively; whereas hot Hungary samples contained 719 ± 192 $mg \cdot kg^{-1}$. Besides, in accordance to Giuffrida *et. al*. (Giuffrida et al., 2013), ß–CAR was found to be the most predominant carotenoid (15-510 $mg \cdot kg^{-1}$) in all samples, followed by ß–CRYPT (25-360 $mg \cdot kg^{-1}$), and CT (6-270 $mg \cdot kg^{-1}$). Intead, VIO and CR occurred at lower concentrations (4.2-42 $mg \cdot kg^{-1}$). Moreover, although it seemed that LUT was detected in samples from Hungary, this signal may be due to zeaxanthin (ZEA), which is a lutein isomer that cannot be separated from LUT using a C18 column (Kim, Geon, Park, Pyo, & Kim, 2016) and whose presence has been reported previously in red paprika (Deli, Molnár, Matus, & Tóth, 2001; Hassan, Yusof, Yahaya, Rozali, & Othman, 2019). Because of the structural similarities between ZEA and LUT, which may lead to comparable ionization efficiency, ZEA was quantified using LUT standard. Furthermore, VIO could not be quantified in samples from the Czech Republic and Murcia, since its concentration was below its MLOQ. Therefore, because of the observed differences in the presence of capsaicinoid and carotenoid, they were proposed as chemical descriptors to address paprika authentication based on chemometrics.

3.4. Multivariate data analysis

In views of the qualitative and quantitative differences between paprika samples of different geographical origins and types, the concentrations of carotenoids and

419 capsaicinoids were proposed as chemical descriptors to address their authentication by
420 multivariate data analysis. PCA was preliminarily applied to check the behavior of
421 paprika and QC samples. Hence, the data matrix of 151 × 10 (samples × variables)
422 dimension, containing the calculated carotenoid and capsaicinoid content for the analyzed
423 paprika and QC samples (15), was studied. The scores plot of PC1 *vs.* PC2 depicted in
424 Fig. S2A (PC1 and PC2 explained variance of 50.23 and 31.18%, respectively) showed
425 that QC samples appeared in the middle of the plot, meaning the absence of systematic
426 errors in the data acquisition and validating the chemometric results. Moreover, high
427 Hotelling T^2 and Q residual values were not observed (Fig. S2B), suggesting the absence
428 of outlier samples.

429 PLS-DA was chosen as the chemometric technique to conduct the classificatory
430 analysis. A first PLS-DA model was built, which included all the paprika samples under
431 study, according to both origins, and type. Thus, a 136 × 10 X-data matrix and a Y-data
432 matrix, assigning samples to nine classes, were used. Fig. 4 shows the corresponding
433 scores plot of LV1 *vs.* LV2 (two LVs, explaining the 18.29% Y-variance, were chosen
434 for constructing the PLS-DA model), where remarkable discrimination between types
435 could be seen. In this line, sweet samples were located on the upper side of the plot,
436 whereas the hot ones on the bottom. Variable importance in projection (VIP) values
437 indicated that this separation was mainly because of CAP, NDC, and DC contents.
438 However, bittersweet *La Vera* samples did not present significant differences with *La*
439 *Vera* sweet ones, so they were considered both as sweet in the following chemometric
440 studies. Regarding the production area, Hungary paprika samples were clearly
441 distinguished in the right part of the plot (displaying positive LV1 scores values) from the
442 other samples, whose classification was not achieved with this PLS-DA model.

Therefore, considering the complexity of the classification due to the wide range of classes, the design of a classification decision tree formed by smaller PLS-DA models was proposed. The followed path to achieve sample classification is shown in Fig. 5 and consisted of four main steps in the PLS-DA model: firstly, hot *vs.* sweet; secondly, Hungary *vs.* others; thirdly, *La Vera vs.* others; and finally, *Murcia vs.* the Czech Republic. Calibration model details such as data matrices dimensions, CV approach, LVs for their construction, X and Y-variance explained, and calibration sensitivity and specificity, are also given in Fig. 5. These PLS-DA calibration models, whose classification scores plots of some of them are depicted in Fig. S3, were built with 70% of the analyzed paprika samples as the training set (89 × 10, dimension data matrix), while the external validation was carried out with the remaining 30% (47 × 10, dimension data matrix). Satisfactory results regarding the geographical origin classification of paprika samples by the determination of carotenoid and capsaicinoid were obtained with a rate of 80.9%. When evaluating the results by origins, 87.5, 60.0, 90.0, and 100.0% rates were reached for Hungary, *La Vera*, *Murcia*, and the Czech Republic paprika samples, respectively. Most of *La Vera* misclassified samples were assigned as *Murcia* samples and backward, which could indicate that specific external conditions related to the country of origin (e.g., climate or farmland) are related to the capsaicinoid and carotenoid profile.

4. Conclusions

In this work, the UHPLC–APCI–HRMS capsaicinoid and carotenoid profile have proved to be an adequate chemical descriptor to classify and authenticate paprika samples from different geographical origins (*La Vera*, *Murcia*, Hungary, and the Czech Republic) and types (hot, sweet and bittersweet). One of the main advantages of the proposed UHPLC–APCI–HRMS methods is the efficient ionization of both capsaicinoids and carotenoid under APCI conditions and the greater selectivity achieved by HRMS.

Besides, a total classification rate of 80.9% was led by building a classification decision tree based on consecutive PLS-DA models and performing an external validation. The breaking down of this result by origin reached 87.5, 60.0, 90.0, and 100.0% rates for Hungary, *La Vera*, *Murcia*, and the Czech Republic samples, respectively. The capsaicinoid content was strongly related to the flavor paprika type, while the carotenoid content could be associated with the country of origin by external conditions since most *La Vera* misclassified samples were assigned as *Murcia* samples and backward.

In future estudies, other geographical origin paprika samples could be also tested to further demonstrate the wide applicability of the proposed UHPLC–APCI–HRMS method. Additionally, other carotenoids, capsaicinoids or derivative compounds (*e.g.*, antheraxanthin, cryptocapsin, or capsanthin-3,6-epoxide) could also be included as target compounds to provide UHPLC–APCI–HRMS profiles with richer information. Finally, the use of data fusion strategies combining the capsaicinoid and carotenoid profile with the polyphenolic profile, as well as other supervised classificatory chemometric techniques such as orthogonal projections to latent structures-discriminant analysis (OPLS-DA) or soft independent modeling of class analogy (SIMCA) could also be explored in future works to further improve the classification of paprika samples.

Acknowledgements

Authors acknowledge the financial support from the Spanish Ministry of Science, Innovation and Universities (PGC2018-095013-B-I00), and from the Agency for Management of University and Research Grants, Government of Catalonia (2017SGR–310). Ane Arrizabalaga-Larrañaga thanks the Agency for Management of University and Research Grants and the European Social Fund for the PhD FI–DGR fellowship and Guillem Campmajó thanks the University of Barcelona for the PhD APIF fellowship.

References

Arimboor, R., Natarajan, R. B., Menon, K. R., Chandrasekhar, L. P., & Moorkoth, V. (2015). Red pepper (Capsicum annuum) carotenoids as a source of natural food colors: analysis and stability—a review. *Journal of Food Science and Technology*, *52*(3), 1258–1271. https://doi.org/10.1007/s13197-014-1260-7

Arrizabalaga-Larrañaga, A., Rodríguez, P., Medina, M., Santos, F. J., & Moyano, E. (2019). Simultaneous analysis of natural pigments and E-141i in olive oils by liquid chromatography–tandem mass spectrometry. *Analytical and Bioanalytical Chemistry*, *411*(21), 5577–5591. https://doi.org/10.1007/s00216-019-01939-3

Arrizabalaga-Larrañaga, A., Rodríguez, P., Medina, M., Santos, F. J., & Moyano, E. (2020). Pigment profiles of Spanish extra virgin olive oils by ultra-high-performance liquid chromatography coupled to high-resolution mass spectrometry. *Food Additives and Contaminants - Part A Chemistry, Analysis, Control, Exposure and Risk Assessment*, *37*(7), 1075–1086. https://doi.org/10.1080/19440049.2020.1753891

Barbero, G. F., Liazid, A., Ferreiro-González, M., Palma, M., & Barroso, C. G. (2016). Fast separation of capsaicinoids from peppers by reversed phase ultra-performance liquid chromatography: Comparation with traditional high-performance liquid chromatography methods. *International Journal of Food Properties*, *19*(5), 984–992. https://doi.org/10.1080/10942912.2015.1050673

Barbosa, S., Saurina, J., Puignou, L., & Nuñez, O. (2020). Targeted UHPLC-HRMS (Orbitrap) Polyphenolic and Capsaicinoid Profiling for the Chemometric Characterization and Classification of Paprika with Protected Designation of Origin (PDO) Attributes. *Molecules*, *25*, 1–16. https://doi.org/10.3390/molecules25071623

Barbosa, S., Saurina, J., Puignou, L., & Núñez, O. (2020). Classification and authentication of paprika by UHPLC-HRMS fingerprinting and multivariate calibration methods (PCA and PLS-DA). *Foods*, *9*(4). https://doi.org/10.3390/foods9040486

Bijttebier, S., Zhani, K., D'Hondt, E., Noten, B., Hermans, N., Apers, S., & Voorspoels, S. (2014). Generic characterization of apolar metabolites in red chili peppers (Capsicum frutescens L.) by orbitrap mass spectrometry. *Journal of Agricultural and Food Chemistry*, *62*(20), 4812–4831. https://doi.org/10.1021/jf500285g

Breemen, R. B. van, Dong, L., & Pajkovic, N. D. (2012). Atmospheric Pressure chemical Ionization Tandem Mass Spectrometry of Carotenoids. *International Journal of Mass Spectrometry*, *312*, 163–172. https://doi.org/10.1016/j.ijms.2011.07.030

Brunner, M., Katona, R., Stefánka, Z., & Prohaska, T. (2010). Determination of the geographical origin of processed spice using multielement and isotopic pattern on the example of Szegedi paprika. *European Food Research and Technology*, *231*(4), 623–634. https://doi.org/10.1007/s00217-010-1314-7

Campmajó, G., Núñez, N., & Núñez, O. (2019). The Role of Liquid Chromatography-Mass Spectrometry in Food Integrity and Authenticity. In G. S. Kamble (Ed.), *Mass Spectrometry - Future Perceptions and Applications* (pp. 3–20). https://doi.org/http://dx.doi.org/10.5772/57353

Cetó, X., Sánchez, C., Serrano, N., Díaz-Cruz, J. M., & Núñez, O. (2020). Authentication of paprika using HPLC-UV fingerprints. *Lwt*, *124*(December 2019). https://doi.org/10.1016/j.lwt.2020.109153

Cetó, X., Serrano, N., Aragó, M., Gámez, A., Esteban, M., Díaz-Cruz, J. M., & Núñez, O. (2018). Determination of HPLC-UV fingerprints of spanish paprika (Capsicum

annuum L.) for its classification by linear discriminant analysis. *Sensors (Switzerland)*, *18*(12). https://doi.org/10.3390/s18124479

Daood, H. G., Halasz, G., Palotás, G., Palotás, G., Bodai, Z., & Helyes, L. (2015). HPLC determination of capsaicinoids with cross-linked C18 column and buffer-free eluent. *Journal of Chromatographic Science*, *53*(1), 135–143. https://doi.org/10.1093/chromsci/bmu030

de Sá Mendes, N., & Branco de Andrade Gonçalves, É. C. (2020). The role of bioactive components found in peppers. *Trends in Food Science and Technology*, *99*(February), 229–243. https://doi.org/10.1016/j.tifs.2020.02.032

Deli, J., Molnár, P., Matus, Z., & Tóth, G. (2001). Carotenoid composition in the fruits of red paprika (Capsicum annuum var. lycopersiciforme rubrum) during ripening; biosynthesis of carotenoids in red paprika. *Journal of Agricultural and Food Chemistry*, *49*(3), 1517–1523. https://doi.org/10.1021/jf000958d

European Commission. eAmbrosia - the EU geographical indications register. (2020). Retrieved May 8, 2020, from https://ec.europa.eu/info/food-farming-fisheries/food-safety-and-quality/certification/quality-labels/geographical-indications-register/

Food and Agriculture Organization of the United Nations. (2019). Retrieved May 8, 2020, from Food and agriculture data website: http://www.fao.org/faostat/en/#data/QC/visualize/

Food Fraud Risk Information. (2020). Retrieved May 8, 2020, from Food fraud risk information database. website: https://trello.com/b/aoFO1UEf/food-fraud-risk-information/

Galvin-King, P., Haughey, S. A., & Elliott, C. T. (2018). Herb and spice fraud; the drivers, challenges and detection. *Food Control*, *88*, 85–97. https://doi.org/10.1016/j.foodcont.2017.12.031

Giuffrida, D., Dugo, P., Torre, G., Bignardi, C., Cavazza, A., Corradini, C., & Dugo, G. (2013). Characterization of 12 Capsicum varieties by evaluation of their carotenoid profile and pungency determination. *Food Chemistry*, *140*(4), 794–802. https://doi.org/10.1016/j.foodchem.2012.09.060

Hassan, N. M., Yusof, N. A., Yahaya, A. F., Rozali, N. N. M., & Othman, R. (2019). Carotenoids of capsicum fruits: Pigment profile and health-promoting functional attributes. *Antioxidants*, *8*(10), 1–25. https://doi.org/10.3390/antiox8100469

Hong, E., Lee, S. Y., Jeong, J. Y., Park, J. M., Kim, B. H., Kwon, K., & Chun, H. S. (2017). Modern analytical methods for the detection of food fraud and adulteration by food category. *Journal of the Science of Food and Agriculture*, *97*(12), 3877–3896. https://doi.org/10.1002/jsfa.8364

Kim, J., Geon, C., Park, J., Pyo, Y., & Kim, S. (2016). Carotenoid profiling from 27 types of paprika (Capsicum annuum L .) with different colors , shapes , and cultivation methods. *FOOD CHEMISTRY*, *201*, 64–71. https://doi.org/10.1016/j.foodchem.2016.01.041

Massart, D. L., Vandeginste, B. G. M., Buydens, L. M. C., de Jong, S., Lewi, P.J., & Smeyers-Verbeke, J. (1997). . In Elsevier (Ed.), *Handbook of chemometrics and qualimetrics* (1st ed.). Amsterdam.

Medina, S., Perestrelo, R., Silva, P., Pereira, J. A. M., & Câmara, J. S. (2019). Current trends and recent advances on food authenticity technologies and chemometric approaches. *Trends in Food Science and Technology*, *85*(December 2018), 163–176. https://doi.org/10.1016/j.tifs.2019.01.017

Mudrić, S., Gašić, U. M., Dramićanin, A. M., Ćirić, I., Milojković-Opsenica, D. M., Popović-Đorđević, J. B., … Tešić, Ž. L. (2017). The polyphenolics and carbohydrates as indicators of botanical and geographical origin of Serbian

autochthonous clones of red spice paprika. *Food Chemistry*, *217*, 705–715. https://doi.org/10.1016/j.foodchem.2016.09.038

Nagy, Z., Daood, H., Koncsek, A., Molnár, H., & Helyes, L. (2017). The simultaneous determination of capsaicinoids, tocopherols, and carotenoids in pungent pepper powder. *Journal of Liquid Chromatography and Related Technologies*, *40*(4), 199–209. https://doi.org/10.1080/10826076.2017.1297722

Ördög, A., Poór, P., Štajner, D., Popović, B., Bátori, Z., & Tari, I. (2018). Comparison of the mineral content of processed spice samples of sweet and hot paprika from the szeged region*. *Journal of Elementology*, *23*(2), 521–530. https://doi.org/10.5601/jelem.2017.22.4.1497

Palacios-Morillo, A., Jurado, J. M., Alcázar, A., & Pablos, F. (2016). Differentiation of Spanish paprika from Protected Designation of Origin based on color measurements and pattern recognition. *Food Control*, *62*, 243–249. https://doi.org/10.1016/j.foodcont.2015.10.045

Palacios-Morillo, Ana, Jurado, J. M., Alcázar, Á., & De Pablos, F. (2014). Geographical characterization of Spanish PDO paprika by multivariate analysis of multielemental content. *Talanta*, *128*, 15–22. https://doi.org/10.1016/j.talanta.2014.04.025

Reilly, C. A., Ehlhardt, W. J., Jackson, D. A., Kulanthaivel, P., Mutlib, A. E., Espina, R. J., … Yost, G. S. (2003). Metabolism of capsaicin by cytochrome P450 produces novel dehydrogenated metabolites and decreases cytotoxicity to lung and liver cells. *Chemical Research in Toxicology*, *16*(3), 336–349. https://doi.org/10.1021/tx025599q

Schweiggert, U., Carle, R., & Schieber, A. (2006). Characterization of major and minor capsaicinoids and related compounds in chili pods (Capsicum frutescens L.) by

high-performance liquid chromatography/atmospheric pressure chemical ionization mass spectrometry. *Analytica Chimica Acta*, *557*(1–2), 236–244. https://doi.org/10.1016/j.aca.2005.10.032

Serrano, N., Cetó, X., Núñez, O., Aragó, M., Gámez, A., Ariño, C., & Díaz-Cruz, J. M. (2018). Characterization and classification of Spanish paprika (Capsicum annuum L.) by liquid chromatography coupled to electrochemical detection with screen-printed carbon-based nanomaterials electrodes. *Talanta*, *189*(June), 296–301. https://doi.org/10.1016/j.talanta.2018.06.085

Stipcovich, T., Barbero, G. F., Ferreiro-González, M., Palma, M., & Barroso, C. G. (2018). Fast analysis of capsaicinoids in Naga Jolokia extracts (Capsicum chinense) by high-performance liquid chromatography using fused core columns. *Food Chemistry*, *239*, 217–224. https://doi.org/10.1016/j.foodchem.2017.06.098

Václav Štursa , Pavel Diviš, J. P. (2018). CHARACTERISTICS OF PAPRIKA SAMPLES OF DIFFERENT GEOGRAPHICAL ORIGIN. *Journal, Potravinarstvo Slovak Sciences, Food*, *12*(1), 254–261. https://doi.org/10.5219/902

Wolf, R., Huschka, C., Raith, K., Wohlrab, W., & Neubert, R. H. H. (1999). Rapid quantification of capsaicin and dihydrocapsaicin in human skin extracts after dermal administration using HPLC-ESI-MS. *Journal of Liquid Chromatography and Related Technologies*, *22*(4), 531–539. https://doi.org/10.1081/JLC-100101678

641 ***Figure Captions***

642 **Fig. 1** Chemical structures, acronyms, and chemical formula of the studied
643 capsaicinoids and carotenoids.

644 **Fig. 2** HRMS spectrum and AIF (HRMS) spectrum of (A) DC and (B) CT.

645 **Fig. 3** UHPLC–APCI–HRMS capsaicinoid and carotenoid profile chromatograms
646 of sweet paprika samples from (A) *Murcia*, sample MS9, and (B) Hungary,
647 sample HS5.

648 **Fig. 4** PLS-DA Scores plot of LV1 *vs*. LV2, using the UHPLC-HRMS capsaicinoid
649 and carotenoid profiling for the classification of all the paprika samples
650 tested.

651 **Fig. 5** Classification decision tree built by HMB for paprika geographical origin
652 authentication by means of PLS-DA models. Dimensions, CV used method,
653 LVs, and sensitivity and specificity of the model are detailed.

654

655

Table 1. Description of the samples analysed in the paprika classification study.

Country	Region	Abbreviation	Number of samples			PDO	Production year
			Hot (H)	Sweet (S)	Bittersweet (BS)		
Spain	La Vera	V	15[a]	15 [a]	15 [a]	Yes	2017
	Murcia	M	15	15	–	Yes	2017
Hungary	Kalocsa	H	18 + 5[a]	18 + 5[a]	–	No	2018
Czech Republic	–	CR	5	5 + 5[a]	–	No	2017

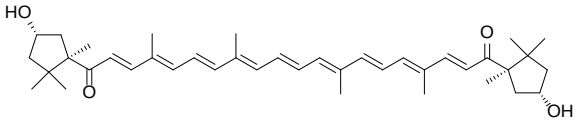
[a] Smoked paprika simples

Table 2. Retention time, ion assignment and accurate mass error of target compounds obtained from the UHPLC–HRMS and AIF (HRMS) data.

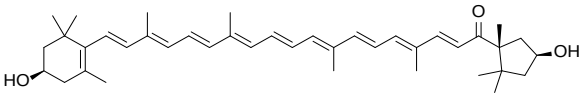
Compound	LC	HRMS			MS/HRMS		
	t_R (min)	Experimental *m/z* (Rel. Ab. %)	Ion Assignment	Accurate mass error (ppm)	Fragment ion (*m/z*)	Ion Assignment	Accurate mass error (ppm)
NDC	4.30	294.2060 (100)	$[M+H]^+$	-1.0	158.1536	$[M+H-C_8H_8O_2]^+$	-1.9
		158.1537 (85)	$[M+H-C_8H_8O_2]^+$	-1.3	137.0595	$[C_8H_9O_2]^+$	-1.5
		137.0598 (25)	$[C_8H_9O_2]^+$	0.7	122.0362	$[C_7H_6O_2]^{+\bullet}$	0.0
					94.0417	$[C_7H_6O_2\text{-}CO]^{+\bullet}$	4.2
					66.0465	$[C_7H_6O_2\text{-}C_2O_2]^{+\bullet}$	1.5
CAP	4.33	306.2056 (100)	$[M+H]^+$	-2.3	137.0594	$[C_8H_9O_2]^+$	-2.2
		170.1536 (15)	$[M+H-C_8H_8O_2]^+$	0.0	122.0362	$[C_7H_6O_2]^{+\bullet}$	0.0
		137.0595 (75)	$[C_8H_9O_2]^+$	-1.4	94.0417	$[C_7H_6O_2\text{-}CO]^{+\bullet}$	4.2
					66.0465	$[C_7H_6O_2\text{-}C_2O_2]^{+\bullet}$	3.0
DC	4.50	308.2214 (100)	$[M+H]^+$	-1.9	172.1692	$[M+H-C_8H_8O_2]^+$	-2.3
		172.1693 (30)	$[M+H-C_8H_8O_2]^+$	-1.7	137.0595	$[C_8H_9O_2]^+$	-1.4
		137.0596 (35)	$[C_8H_9O_2]^+$	-0.7	122.0362	$[C_7H_6O_2]^{+\bullet}$	0.0
					94.0417	$[C_7H_6O_2\text{-}CO]^{+\bullet}$	4.2
					66.0465	$[C_7H_6O_2\text{-}C_2O_2]^{+\bullet}$	1.5
NDCT	5.32	137.0596 (100)	$[C_8H_9O_2]^+$	-0.7	137.0595	$[C_8H_9O_2]^+$	-1.5
					122.0362	$[C_7H_6O_2]^{+\bullet}$	0.8
					94.0417	$[C_7H_6O_2\text{-}CO]^{+\bullet}$	4.2
					66.0465	$[C_7H_6O_2\text{-}C_2O_2]^{+\bullet}$	1.5
CR	7.03	601.4241 (30)	$[M+H]^+$	-1.7	221.1531	$[C_{14}H_{21}O_2]^+$	-2.3
		583.4137 (100)	$[M+H-H_2O]^+$	-1.4	109.1013	$[C_8H_{13}]^+$	1.8
VIO	7.45	601.424 (100)	$[M+H]^+$	-1.8	583.4132	$[M+H-H_2O]^+$	-2.2
		583.4137 (45)	$[M+H-H_2O]^+$	-1.4	221.153	$[C_{14}H_{21}O_2]^+$	-2.7
					165.0907	$[C_{10}H_{13}O_2]^+$	-1.9

					119.0853	$[C_9H_{11}]^+$	-1.9
CT	7.28	585.4291 (100)	$[M+H]^+$	-1.9	567.4183	$[M+H-H_2O]^+$	-2.3
		567.4186 (45)	$[M+H-H_2O]^+$	-1.8	119.0856	$[C_9H_{11}]^+$	0.6
					109.1013	$[C_8H_{13}]^+$	1.8
LUT	8.23	569.4349 (20)	$[M+H]^+$	-0.7	145.101	$[C_{11}H_{13}]^+$	-1.2
		551.4239 (100)	$[M+H-H_2O]^+$	-1.4	119.0856	$[C_9H_{11}]^+$	0.6
					105.0701	$[C_8H_9]^+$	2.2
β–CRYPT	11.60	553.4394 (100)	$[M+H]^+$	-1.8	535.4294	$[M+H-H_2O]^+$	-0.7
		535.4291 (25)	$[M+H-H_2O]^+$	-1.3	145.101	$[C_{11}H_{13}]^+$	-1.2
					119.0856	$[C_9H_{11}]^+$	0.6
					105.0701	$[C_8H_9]^+$	2.2
β–CAR	12.30	537.4445 (100)	$[M+H]^+$	-1.9	177.1634	$[C_{13}H_{21}]^+$	-1.7
					119.0856	$[C_9H_{11}]^+$	0.8
					105.0700	$[C_8H_9]^+$	1.2

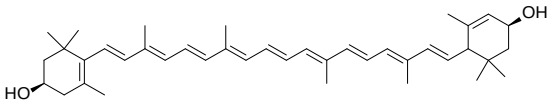
Figure 1



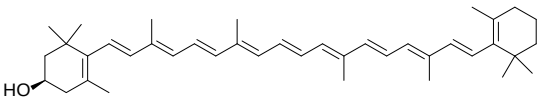
Capsorubin (CR) $C_{40}H_{56}O_4$



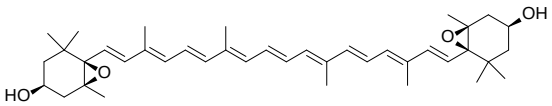
Capsanthin (CT) $C_{40}H_{56}O_3$



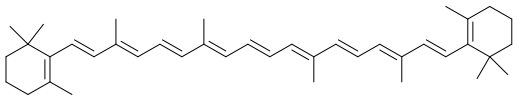
Lutein (LUT) $C_{40}H_{56}O_2$



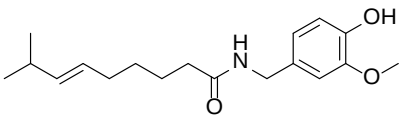
β–Cryptoxanthin (β–CRYPT) $C_{40}H_{56}O$



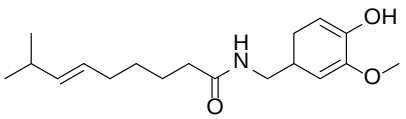
Violaxanthin (VIO) $C_{40}H_{56}O_4$



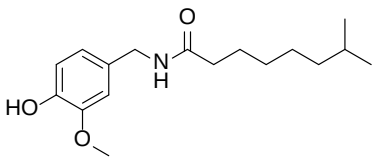
β–Carotene (β–CAR) $C_{18}H_{27}NO_3$

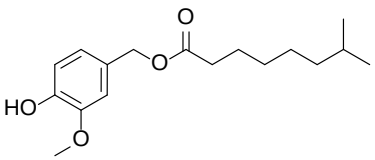


Capsaicin (CAP) $C_{18}H_{27}NO_3$



Dihydrocapsaicin (DC) $C_{18}H_{29}NO_3$





Nordihydro

Figure 2

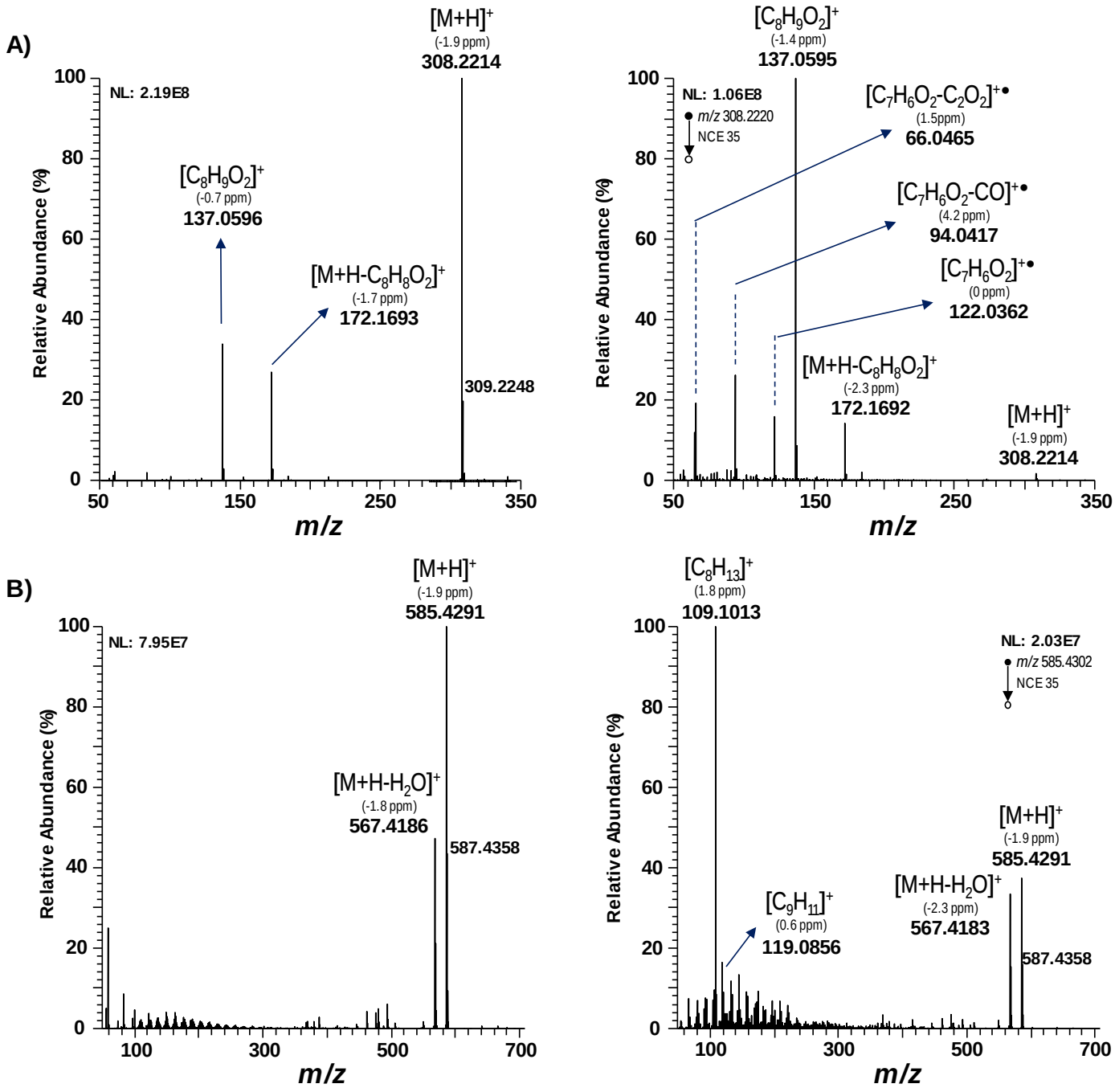

A)
[M+H]+
(-1.9 ppm)
308.2214
NL: 2.19E8
[C8H9O2]+
(-0.7 ppm)
137.0596
[M+H-C8H8O2]+
(-1.7 ppm)
172.1693
309.2248
Relative Abundance (%)
m/z
[C8H9O2]+
(-1.4 ppm)
137.0595
NL: 1.06E8
m/z 308.2220
NCE 35
[C7H6O2-C2O2]+•
(1.5ppm)
66.0465
[C7H6O2-CO]+•
(4.2 ppm)
94.0417
[C7H6O2]+•
(0 ppm)
122.0362
[M+H-C8H8O2]+
(-2.3 ppm)
172.1692
[M+H]+
(-1.9 ppm)
308.2214
B)
[M+H]+
(-1.9 ppm)
585.4291
NL: 7.95E7
[M+H-H2O]+
(-1.8 ppm)
567.4186
587.4358
[C8H13]+
(1.8 ppm)
109.1013
NL: 2.03E7
m/z 585.4302
NCE 35
[C9H11]+
(0.6 ppm)
119.0856
[M+H-H2O]+
(-2.3 ppm)
567.4183

Figure 3

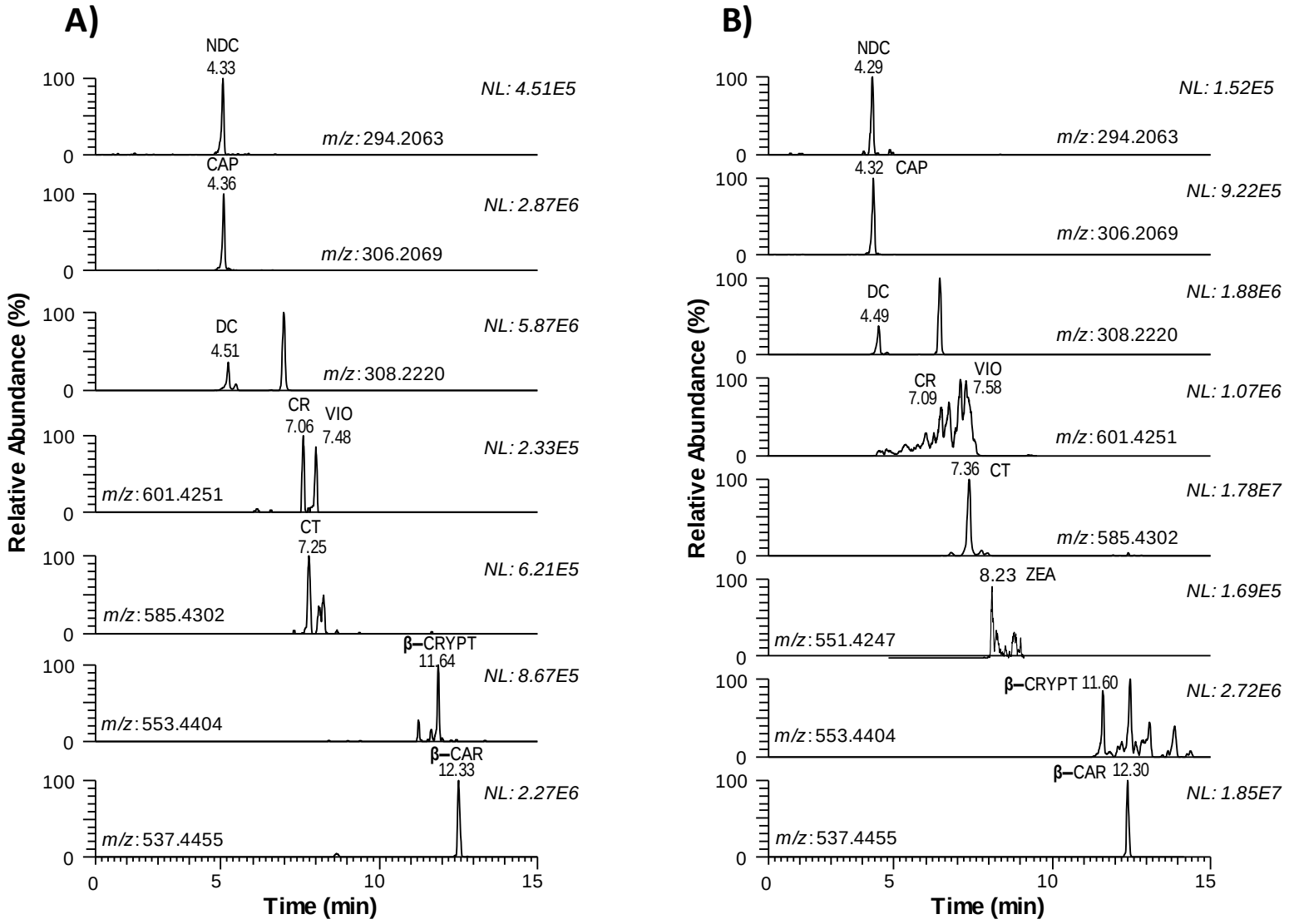

A)
NDC
4.33
NL: 4.51E5
m/z: 294.2063
CAP
4.36
NL: 2.87E6
m/z: 306.2069
DC
4.51
NL: 5.87E6
m/z: 308.2220
CR
7.06
VIO
7.48
NL: 2.33E5
m/z: 601.4251
CT
7.25
NL: 6.21E5
m/z: 585.4302
β–CRYPT
11.64
NL: 8.67E5
m/z: 553.4404
β–CAR
12.33
NL: 2.27E6
m/z: 537.4455
Relative Abundance (%)
Time (min)
B)
NDC
4.29
NL: 1.52E5
m/z: 294.2063
4.32 CAP
NL: 9.22E5
m/z: 306.2069
DC
4.49
NL: 1.88E6
m/z: 308.2220
CR
7.09
VIO
7.58
NL: 1.07E6
m/z: 601.4251
7.36 CT
NL: 1.78E7
m/z: 585.4302
8.23 ZEA
NL: 1.69E5
m/z: 551.4247
β–CRYPT 11.60
NL: 2.72E6
m/z: 553.4404
β–CAR 12.30
NL: 1.85E7
m/z: 537.4455
Relative Abundance (%)
Time (min)

Figure 4

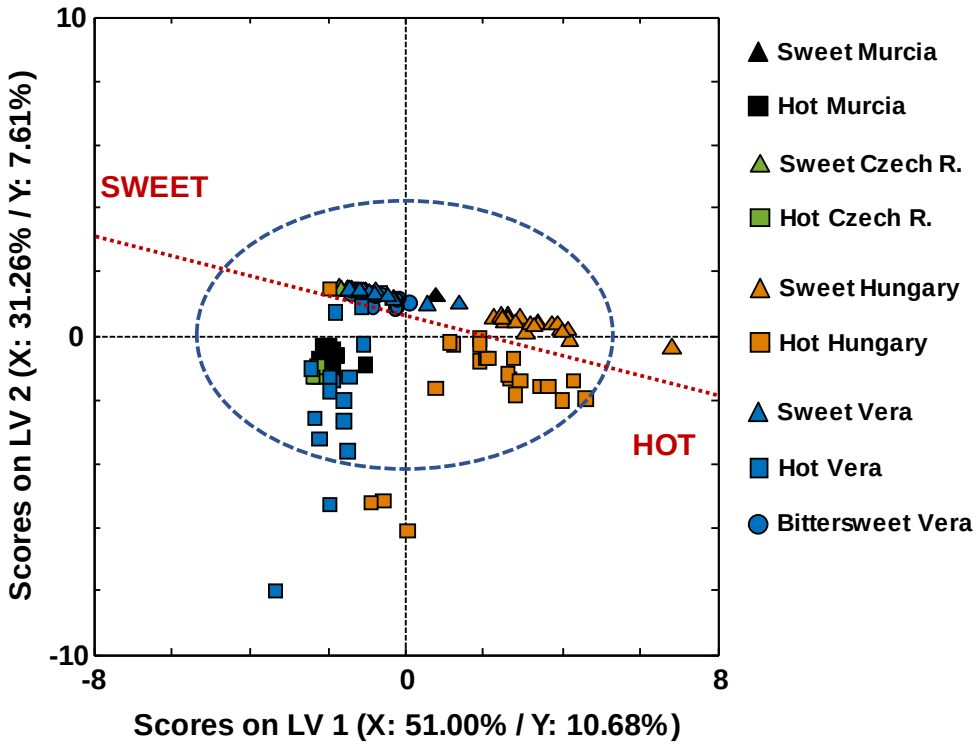
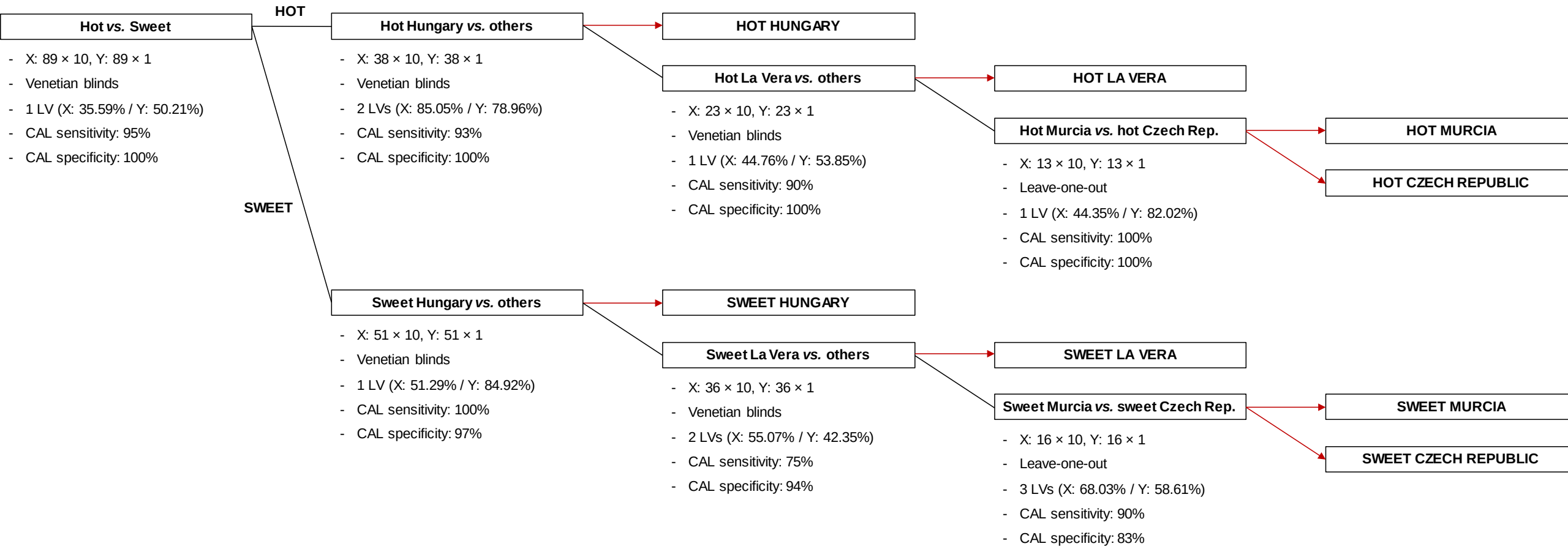

10
0
-10
-8
0
8
SWEET
HOT
Scores on LV 2 (X: 31.26% / Y: 7.61%)
Scores on LV 1 (X: 51.00% / Y: 10.68%)
Sweet Murcia
Hot Murcia
Sweet Czech R.
Hot Czech R.
Sweet Hungary
Hot Hungary
Sweet Vera
Hot Vera
Bittersweet Vera

Figure 5


Hot *vs.* Sweet
- X: 89 × 10, Y: 89 × 1
- Venetian blinds
- 1 LV (X: 35.59% / Y: 50.21%)
- CAL sensitivity: 95%
- CAL specificity: 100%
HOT
SWEET
Hot Hungary *vs.* others
- X: 38 × 10, Y: 38 × 1
- Venetian blinds
- 2 LVs (X: 85.05% / Y: 78.96%)
- CAL sensitivity: 93%
- CAL specificity: 100%
Sweet Hungary *vs.* others
- X: 51 × 10, Y: 51 × 1
- Venetian blinds
- 1 LV (X: 51.29% / Y: 84.92%)
- CAL sensitivity: 100%
- CAL specificity: 97%
HOT HUNGARY
Hot La Vera *vs.* others
- X: 23 × 10, Y: 23 × 1
- Venetian blinds
- 1 LV (X: 44.76% / Y: 53.85%)
- CAL sensitivity: 90%
- CAL specificity: 100%
SWEET HUNGARY
Sweet La Vera *vs.* others
- X: 36 × 10, Y: 36 × 1
- Venetian blinds
- 2 LVs (X: 55.07% / Y: 42.35%)
- CAL sensitivity: 75%
- CAL specificity: 94%
HOT LA VERA
Hot Murcia *vs.* hot Czech Rep.
- X: 13 × 10, Y: 13 × 1
- Leave-one-out
- 1 LV (X: 44.35% / Y: 82.02%)
- CAL sensitivity: 100%
- CAL specificity: 100%
SWEET LA VERA
Sweet Murcia *vs.* sweet Czech Rep.
- X: 16 × 10, Y: 16 × 1
- Leave-one-out
- 3 LVs (X: 68.03% / Y: 58.61%)
- CAL sensitivity: 90%
- CAL specificity: 83%
HOT MURCIA
HOT CZECH REPUBLIC
SWEET MURCIA
SWEET CZECH REPUBLIC

Supplementary Material

Determination of capsaicinoids and carotenoids for the characterization and geographical origin authentication of paprika by UHPLC–APCI–HRMS

Ane Arrizabalaga-Larrañaga[#(1)], Guillem Campmajó[#(1,2)], Javier Saurina[(1,2)], Oscar Núñez[(1,2,3)], Francisco Javier Santos[(1)], Encarnación Moyano[(1)]*

[(1)] Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, Barcelona, Spain
[(2)] Research Institute in Food Nutrition and Food Safety, University of Barcelona, Santa Coloma de Gramenet, Spain
[(3)] Serra Húnter Program, Generalitat de Catalunya, Barcelona, Spain

* Corresponding author: E. Moyano

Table of Contents

Supplementary Tables

Table S1. Concentrations ($mg \cdot kg^{-1}$) of capsaicinoids and carotenoids determined in paprika samples.

Sample	NDC	CAP	DC	NDCT	CR	VIO	CT	ZEA	β–CRYPT	β–CAR
[a]VH1	17	180	242	nd	9.1	6.0	12	nd	64	70
[a]VH2	31	288	373	nd	<LOQ	<LOQ	9.5	nd	<LOQ	15
[a]VH3	62	507	692	nd	11	7.5	16	<LOQ	68	59
[a]VH4	52	409	594	nd	<LOQ	<LOQ	16	<LOQ	17	19
[a]VH5	33	341	375	nd	7.4	<LOQ	6.4	nd	46	28
[a]VH6	45	478	554	nd	10	7.0	12	nd	55	55
[a]VH7	137	1020	1133	nd	9.0	5.4	<LOQ	nd	<LOQ	15
[a]VH8	31	381	478	nd	7.1	6.0	7.7	nd	<LOQ	33
[a]VH9	11	78	104	nd	<LOQ	<LOQ	<LOQ	nd	28	37
[a]VH10	57	510	644	nd	9.5	5.8	<LOQ	nd	<LOQ	18
[a]VH11	34	279	373	nd	7.9	<LOQ	14	<LOQ	63	95
[a]VH12	26	349	643	nd	8.7	8.5	13	nd	25	47
[a]VH13	6.8	42.4	66.5	nd	7.2	<LOQ	8.3	nd	54	57
[a]VH14	29.2	361.6	368.4	nd	6.0	<LOQ	7.5	nd	49	38
[a]VH15	85.4	691.8	919.4	nd	14	6.6	15	nd	60	57
[a]VS1	0.5	3.1	5.5	nd	8.6	<LOQ	9.8	nd	54	77
[a]VS2	0.8	4.9	7.6	nd	13	6.9	<LOQ	nd	79	119
[a]VS3	0.5	1.1	3.1	nd	13	13	220	nd	126	119
[a]VS4	1.4	6.3	12	nd	7.5	<LOQ	11	nd	69	82
[a]VS5	0.3	1.4	2.6	nd	7.4	6.0	112	nd	<LOQ	35
[a]VS6	1.6	7.6	12	nd	11	6.3	35	<LOQ	82	57
[a]VS7	0.2	1.2	3.1	nd	7.4	<LOQ	13	nd	43	83
[a]VS8	<LOQ	0.8	1.7	nd	7.3	5.6	19	31	39	25
[a]VS9	1.5	7.5	13	nd	<LOQ	<LOQ	<LOQ	nd	73	52

[a]smoked paprika sample; nd: not detected (<MLOD)

Table S1. (Cont) Concentrations ($mg \cdot kg^{-1}$) of capsaicinoids and carotenoids determined in paprika samples.

Sample	NDC	CAP	DC	NDCT	CR	VIO	CT	ZEA	β–CRYPT	β–CAR
[a]VS10	0.7	2.7	5.7	nd	8.3	6.3	16	nd	62	55
[a]VS11	1.1	5.2	8.0	nd	16	11	37	<LOQ	121	81
[a]VS12	0.6	3.0	4.6	nd	<LOQ	<LOQ	11	nd	73	37
[a]VS13	0.4	1.4	3.3	nd	6.5	<LOQ	7.4	nd	<LOQ	50
[a]VS14	<LOQ	0.8	1.7	nd	12	8.3	13	nd	64	70
[a]VS15	0.4	1.0	2.5	nd	7.4	<LOQ	12	nd	25	49
[a]VBS	4.6	36	47	nd	13.0	7.3	18	<LOQ	78	79
[a]VBS	1.3	11	15	nd	12.5	<LOQ	15	nd	49	56
[a]VBS	2.1	14	21	nd	6.1	<LOQ	5.5	nd	24	34
[a]VBS	6.7	39	58	nd	9.1	6.9	24	<LOQ	48	22
[a]VBS	4.4	26	41	nd	15	7.8	15	nd	69	93
[a]VBS	0.8	4.8	6.3	nd	5.3	<LOQ	19	nd	64	48
[a]VBS	0.6	3.7	6.2	nd	17	9.3	25	nd	52	40
[a]VBS	0.4	1.6	2.6	nd	9.0	5.6	16	nd	<LOQ	55
[a]VBS	1.3	4.9	7.8	nd	<LOQ	<LOQ	6.7	nd	35	85
[a]VBS	0.4	1.3	3.1	nd	8.2	<LOQ	12	nd	54	101
[a]VBS	1.7	18	19	nd	17	8.9	19	nd	85	116
[a]VBS	0.6	4.2	5.7	nd	11	5.7	16	nd	69	63
[a]VBS	0.6	2.0	3.9	nd	14	7.2	14	nd	79	117
[a]VBS	1.9	9.8	17	nd	13	7.8	16	nd	78	105
[a]VBS	0.8	4.9	7.3	nd	9.4	5.5	11	nd	55	79
MH1	25	272	257	nd	<LOQ	<LOQ	<LOQ	nd	17	68
MH2	27	292	269	nd	<LOQ	<LOQ	<LOQ	nd	40	59
MH3	25	251	271	nd	<LOQ	<LOQ	5.3	nd	25	61
MH4	24	240	254	nd	<LOQ	<LOQ	<LOQ	nd	21	65
MH5	20	238	244	nd	<LOQ	<LOQ	<LOQ	nd	<LOQ	42
MH6	22	270	278	nd	<LOQ	<LOQ	<LOQ	nd	20	57

[a]smoked paprika sample; nd: not detected (<MLOD)

Table S1. (Cont) Concentrations ($mg \cdot kg^{-1}$) of capsaicinoids and carotenoids determined in paprika samples.

Sample	NDC	CAP	DC	NDCT	CR	VIO	CT	ZEA	β–CRYPT	β–CAR
MH7	22	235	240	nd	<LOQ	<LOQ	<LOQ	nd	41	56
MH8	28	303	300	nd	<LOQ	<LOQ	261	nd	27	62
MH9	25	257	252	nd	<LOQ	<LOQ	<LOQ	nd	53	73
MH10	29	317	317	nd	<LOQ	<LOQ	6.1	<LOQ	33	64
MH11	22	234	214	nd	<LOQ	<LOQ	<LOQ	nd	25	52
MH12	24	302	266	nd	<LOQ	<LOQ	<LOQ	nd	<LOQ	49
MH13	24	258	247	nd	<LOQ	<LOQ	<LOQ	nd	44	64
MH14	25	255	271	nd	<LOQ	<LOQ	<LOQ	nd	52	61
MH15	24	234	232	nd	<LOQ	<LOQ	<LOQ	nd	48	32
MS1	0.8	3.8	7.5	nd	<LOQ	<LOQ	<LOQ	nd	<LOQ	18
MS2	0.7	3.0	5.6	nd	5.6	<LOQ	<LOQ	nd	<LOQ	64
MS3	0.7	3.3	6.6	nd	<LOQ	<LOQ	5.5	nd	42	60
MS4	0.8	3.4	5.7	nd	6.3	<LOQ	5.7	nd	47	60
MS5	0.8	3.3	5.7	nd	6.5	<LOQ	7.7	nd	50	62
MS6	0.8	3.3	6.5	nd	5.3	<LOQ	6.3	nd	26	69
MS7	0.8	4.0	7.5	nd	<LOQ	<LOQ	73	nd	32	71
MS8	0.6	3.0	5.2	nd	<LOQ	<LOQ	7.1	nd	46	48
MS9	1.1	4.0	8.1	nd	5.3	<LOQ	384	nd	124	73
MS10	0.8	3.1	5.9	nd	5.6	<LOQ	<LOQ	nd	27	72
MS11	0.8	3.0	5.9	nd	5.9	<LOQ	<LOQ	nd	61	80
MS12	0.7	3.6	6.0	nd	7.4	<LOQ	9.7	nd	50	68
MS13	0.6	3.2	5.2	nd	5.9	<LOQ	6.2	nd	53	63
MS14	0.6	3.1	6.0	nd	5.7	<LOQ	7.4	nd	62	80
MS15	0.8	3.8	7.0	nd	<LOQ	<LOQ	<LOQ	nd	47	75
CRH1	40	280	364	nd	<LOQ	<LOQ	<LOQ	nd	<LOQ	49
CRH2	41	244	331	nd	<LOQ	<LOQ	<LOQ	nd	<LOQ	44
CRH3	43	238	367	nd	<LOQ	<LOQ	<LOQ	nd	21	83

[a]smoked paprika sample; nd: not detected (<MLOD)

Table S1. (Cont) Concentrations ($mg \cdot kg^{-1}$) of capsaicinoids and carotenoids determined in paprika samples.

Sample	NDC	CAP	DC	NDCT	CR	VIO	CT	ZEA	β–CRYPT	β–CAR
CRH4	36	259	323	nd	<LOQ	<LOQ	<LOQ	nd	<LOQ	45
CRH5	48	262	338	nd	<LOQ	<LOQ	<LOQ	nd	<LOQ	28
CRS1	1.0	3.6	6.6	nd	<LOQ	<LOQ	5.7	nd	<LOQ	50
CRS2	0.9	3.6	6.6	nd	<LOQ	<LOQ	5.7	nd	<LOQ	46
CRS3	1.0	4.0	8.7	nd	<LOQ	<LOQ	4.9	nd	26	47
CRS4	0.8	3.3	6.0	nd	<LOQ	<LOQ	4.8	nd	24	41
CRS5	1.1	4.2	9.0	nd	<LOQ	<LOQ	6.7	nd	<LOQ	44
[a]CRS1	1.0	3.1	6.9	nd	<LOQ	<LOQ	<LOQ	nd	30	116
[a]CRS2	1.0	3.2	5.9	nd	<LOQ	<LOQ	31.7	nd	81	86
[a]CRS3	1.0	3.1	8.7	nd	<LOQ	<LOQ	<LOQ	nd	31	78
[a]CRS4	0.8	3.1	6.6	nd	<LOQ	<LOQ	5.7	nd	43	88
[a]CRS5	1.0	3.1	7.8	nd	<LOQ	<LOQ	<LOQ	nd	24	123
HH1	21	151	166	nd	17	9.3	109	155	177	212
HH2	19	138	144	nd	18	9.1	103	155	188	230
HH3	19	134	154	nd	16	9.0	104	150	194	283
HH4	19	59	164	nd	24	10	123	187	123	353
HH5	27	93	204	nd	29	12	63	343	104	97
HH6	24	80	179	nd	22	16	5.3	284	321	388
HH7	90	616	805	nd	7.8	6.6	34	47	160	255
HH8	91	684	986	nd	8.5	10	10	97	189	293
HH9	85	624	884	nd	10	5.6	34	<LOQ	115	263
HH10	17	61	142	nd	21	12	236	240	207	260
HH11	17	68	153	nd	24	9.1	<LOQ	89	104	237
HH12	14	47	125	nd	21	10	71	83	167	282
HH13	27	98	258	nd	24	15	135	143	140	269
HH14	31	114	267	nd	24	12	127	158	195	313
HH15	35	114	293	nd	27	19	45	242	221	362

[a]smoked paprika sample; nd: not detected (<MLOD)

Table S1. (Cont) Concentrations ($mg \cdot kg^{-1}$) of capsaicinoids and carotenoids determined in paprika samples.

Sample	NDC	CAP	DC	NDCT	CR	VIO	CT	ZEA	β–CRYPT	β–CAR
HH16	16	74	133	nd	17	6.6	91	99	226	277
HH17	15	64	133	nd	13	9.1	66	61	152	195
HH18	17	72	141	nd	19	9.5	<LOQ	70	154	213
HS1	0.3	1.6	3.2	nd	26	8.6	121	70	202	317
HS2	0.4	1.2	3.0	nd	27	5.6	113	105	172	334
HS3	0.5	1.7	3.5	nd	25	17	<LOQ	312	131	151
HS4	0.2	1.1	2.1	nd	30	11	112	82	188	446
HS5	0.3	1.1	2.2	nd	4.2	26	39	83	219	500
HS6	0.2	1.1	2.4	nd	36	12	163	77	168	424
HS7	0.5	1.4	3.1	nd	23	5.3	269	124	181	295
HS8	0.4	1.2	2.6	nd	25	5.2	238	112	148	290
HS9	0.5	1.4	2.8	nd	31	13	180	186	211	290
HS10	1.3	4.3	7.9	nd	11	13	41	339	237	504
HS11	1.2	4.1	7.6	nd	14	14	48	51	201	486
HS12	1.3	4.2	7.2	nd	11	10	41	94	191	467
HS13	0.4	1.7	3.2	nd	27	8.6	<LOQ	161	232	218
HS14	0.4	1.2	2.4	nd	24	15	160	78	300	354
HS15	0.4	1.7	3.6	nd	42	22	213	190	359	496
HS16	0.4	1.4	2.6	nd	22	11	122	148	177	229
HS17	0.5	1.6	3.9	nd	27	13	6.7	110	209	303
HS18	0.3	1.3	2.7	nd	22	13	101	64	182	286
[a]HH1	32	111	329	nd	33	22	143	140	233	432
[a]HH2	33	107	298	nd	24	18	131	104	215	385
[a]HH3	30	104	279	nd	27	16	130	169	206	365
[a]HH4	40	142	376	nd	15	13	21	124	123	117
[a]HH5	38	130	358	nd	28	10	101	108	146	509

[a]smoked paprika sample; nd: not detected (<MLOD)

Table S1. (Cont) Concentrations ($mg \cdot kg^{-1}$) of capsaicinoids and carotenoids determined in paprika samples.

Sample	NDC	CAP	DC	NDCT	CR	VIO	CT	ZEA	β–CRYPT	β–CAR
[a]HS1	1.0	3.7	9.5	nd	26	18	118	101	155	310
[a]HS2	1.4	4.2	11	nd	28	16	129	110	249	447
[a]HS3	1.4	4.2	9.7	nd	23	15	129	165	261	382
[a]HS4	1.5	4.6	10	nd	14	8.6	74	117	252	342
[a]HS5	1.3	3.9	10	nd	29	13	6.3	95	233	435

[a]smoked paprika sample; nd: not detected (<MLOD)

Table S2. Total capsaicinoid content (Σ CAPS), total carotenoid content (Σ CAR), and their respective sum (Σ CAPS + Σ CAR), expressed as mean ± standard deviation, obtained for the analyzed paprika samples according to their geographical origin and flavor variety.

	Hot			Sweet			Bittersweet		
	Σ CAPS	Σ CAR	Σ CAPS + CAR	Σ CAPS	Σ CAR	Σ CAPS + CAR	Σ CAPS	Σ CAR	Σ CAPS + CAR
La Vera	942 ± 554	106 ± 50	1048 ± 547	9 ± 6	185 ± 99	194 ± 100	31 ± 31	165 ± 49	196 ± 61
Murcia	549 ± 53	118 ± 69	667 ± 107	10 ± 1	154 ± 125	164 ± 125			
Czech Republic	642 ± 27	75 ± 24	717 ± 38	12 ± 1	117 ± 47	128 ± 46			
Hungary	504 ± 455	719 ± 192	1224 ± 432	8 ± 4	844 ± 160	851 ± 161			

Supplementary Figures

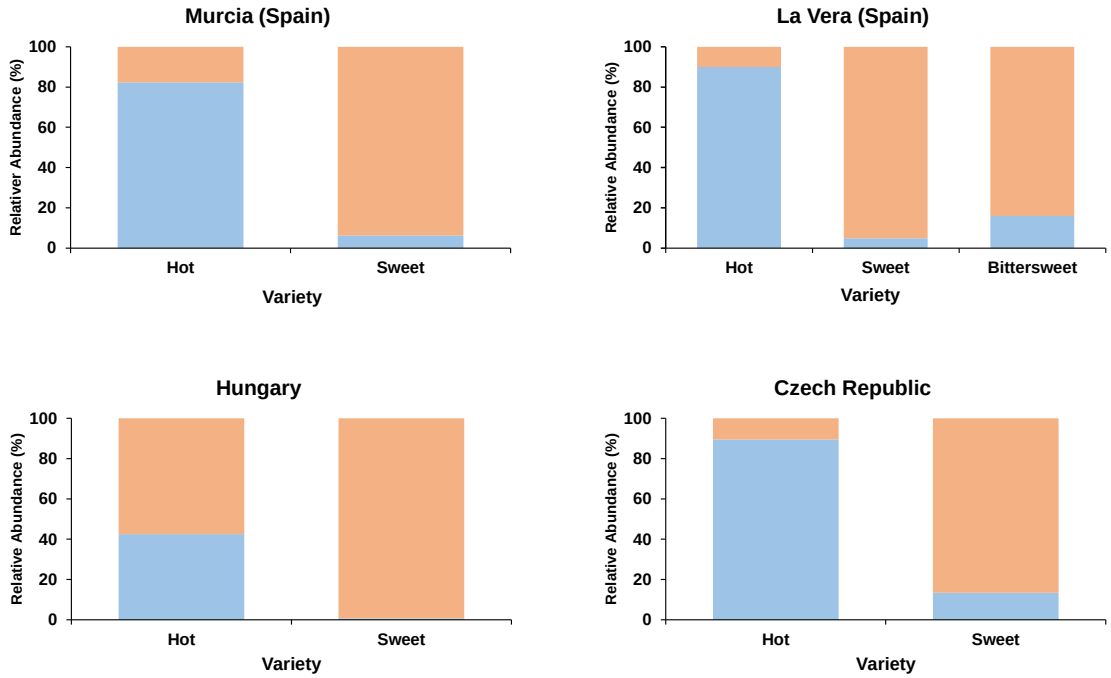


Figure S1: Capsaicinoid (blue) and carotenoid (orange) distribution of Paprika from different origins and varieties.

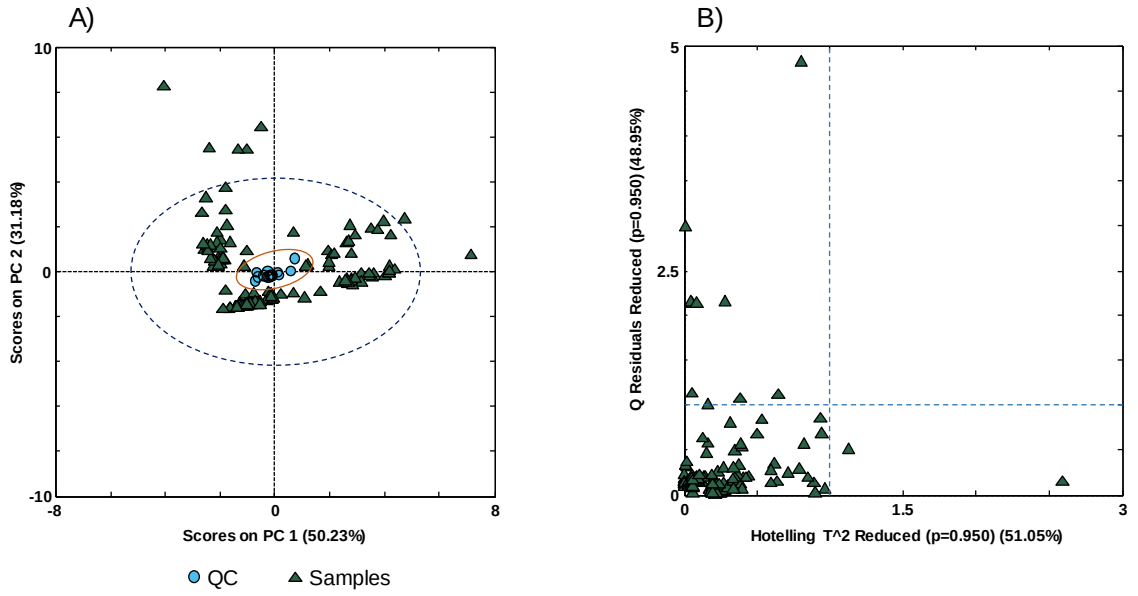


Figure S2: (A) PCA Scores plot of PC1 *vs.* PC2, showing a correct behavior of QC samples. (B) Hotelling T2 *vs.* Q residual values plot for the detection of outlier samples.

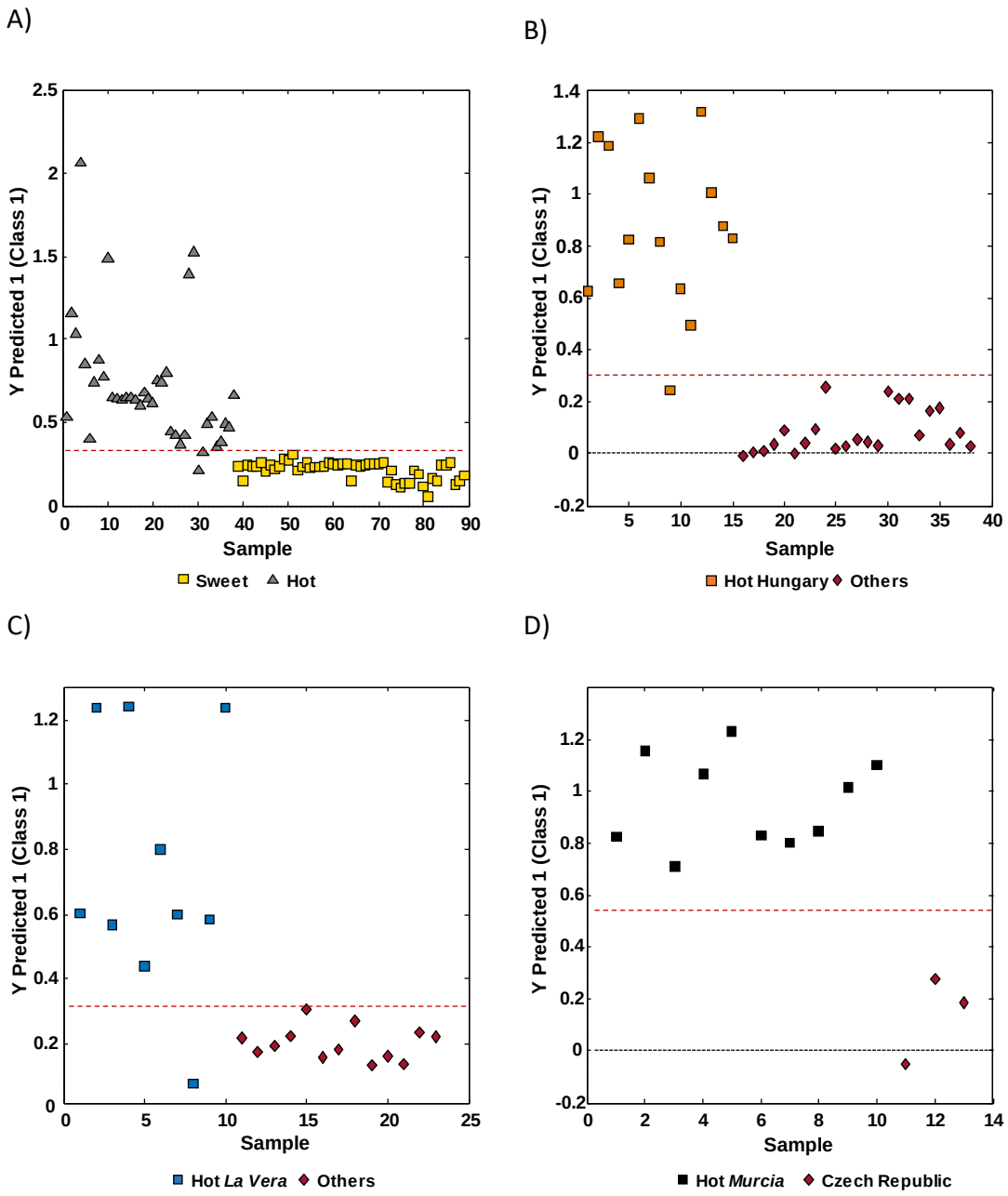


Figure S3: Classification plot depicting Samples *vs.* Y Predicted 1 Scores plot for the PLS-DA calibration models of (A) hot *vs.* sweet, (B) hot Hungary *vs.* others, (C) hot *La Vera vs.* others, and (D) hot *Murcia vs.* Czech Republic.