Assessment of paprika geographical origin fraud by high-performance liquid chromatography with fluorescence detection (HPLC-FLD) fingerprinting

Guillem Campmajó[a,b,*], Luis R. Rodríguez-Javier[a], Javier Saurina[a,b], 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), E08921 Santa Coloma de Gramenet, Spain.

[c] Serra Húnter Fellow, Generalitat de Catalunya, Rambla de Catalunya 19-21, E08007 Barcelona, Spain.

* Corresponding author: Guillem Campmajó

(Tel.: 34-93-402-1286, e-mail: gcampmajo@ub.edu)

Abstract

Paprika production under the protected designation of origin (PDO) standardized procedures leads to more quality products. However, it is also related to higher retail prices, making them susceptible to adulteration with low-quality paprika or its agricultural origin's mislabeling. Therefore, in this study, high-performance liquid chromatography with fluorescence detection (HPLC-FLD) fingerprints, strongly related to phenolic acid and polyphenolic compounds, were proposed as chemical markers to assess the classification of paprika from five European regions (three Spanish PDO, Hungary, and the Czech Republic), through a classification decision tree constructed by partial least squares-regression discriminant analysis (PLS-DA) models. After external validation, an excellent classification accuracy of 97.9% was achieved. Moreover, the chromatographic fingerprints were also proposed to detect and quantitate two different paprika geographical origin blend scenarios by partial least squares (PLS) regression. Low external validation and prediction errors —with values below 1.6 and 10.7%, respectively— were obtained.

Keywords: Paprika; HPLC-FLD; Fingerprinting; Chemometrics; Food Authentication; Protected Designation of Origin

1. INTRODUCTION

In the last decades, society has been increasingly interested not only in food safety but also in its quality, encompassing attributes such as the presence of specific ingredients, the production system (e.g., organic products), or the region of origin. In the case of geographical indication, the European Union (EU) has established three labels — protected designation of origin (PDO), protected geographical indication (PGI), and geographical indication (GI)— that protect the intellectual property rights, as well as the inherent characteristics and reputation, of a food or beverage product directly linked to its production area (The European Parliament and the Council of the European Union, 2012). Among them, PDO distinction demands the strictest requirements since all the steps involved in the agricultural foodstuff production have to be carried out in a specific area through well-described methodologies.

Several spices, which are widely employed as a food seasoning in the main European cuisines because of their organoleptic properties, are currently registered with the PDO status: one cumin, five saffron, and seven paprika products (European Comission, 2020). Focusing on the latter, a valued red powdered spice is obtained from the drying and grinding of red pepper fruits of the genus *Capsicum* (Solanaceae family), with three PDO products coming from Spain (*Pimentón de La Vera*, *Pimentón de Murcia*, and *Pebre bord de Mallorca*), two from Hungary (*Kalocsai fűszerpaprika-őrlemény* and *Szegedi fűszerpaprika-őrlemény*), and one from Slovakia (*Žitavská paprika*) and France (*Piment d'Espelette - Ezpeletako Biperra*). Moreover, besides its particular intense red color, taste, and flavor, paprika is also well-known to be an essential source of antioxidant compounds such as capsaicinoids, carotenoids, tocopherols, ascorbic acid, and phenolic and polyphenolic compounds, which provide important health benefits and have a crucial impact on the fruit quality (Hassan, Yusof, Yahaya, Rozali, & Othman, 2019;

Škrovánková, Mlček, Orsavová, Juríková, & Dřímalová, 2017; Topuz, Dincer, Özdemir, Feng, & Kushad, 2011). Nevertheless, herbs and spices are among the goods most vulnerable to fraudulent practices in the EU (European Comission, 2019). In this line, paprika production under the PDO standardized procedures leads to more quality products with higher retail prices (Danezis, Tsagkaris, Camin, Brusic, & Georgiou, 2016), making them susceptible to adulteration and mislabeling practices. Therefore, the development of analytical methodologies to authenticate paprika origin is necessary.

Because of the lack of specific markers directly related to food origin (primary markers), classic targeted analysis can not solve geographical origin authentication issues. For that reason, according to its agricultural origin, the paprika classification has been assessed through profiling or fingerprinting strategies (Ballin & Laursen, 2019) in combination with chemometrics. On the one hand, both multi-elemental or bioactive substance (i.e., carbohydrates, capsaicinoids, carotenoids, and phenolic and polyphenolic compounds) profiles have been commonly proposed as chemical descriptors to achieve paprika authenticity. While the former has been determined by energy-dispersive X-ray fluorescence (ED-XRF) (Fiamegos, Dumitrascu, Papoci, & de la Calle, 2021), or inductively coupled plasma with optical emission spectroscopy (ICP-OES) (Palacios-Morillo, Jurado, Alcázar, & De Pablos, 2014) or mass spectrometry (ICP-MS) (Ördög et al., 2018), liquid chromatography coupled to low- (LC-MS) (Barbosa, Campmajó, Saurina, Puignou, & Núñez, 2020) or high-resolution mass spectrometry (LC-HRMS) (Arrizabalaga-Larrañaga et al., 2020; Barbosa, Saurina, Puignou, & Núñez, 2020b; Mudrić et al., 2017) have been used to record the latter. On the other hand, several fingerprinting approaches (Cuadros-Rodríguez, Ruiz-Samblás, Valverde-Som, Pérez-Castaño, & González-Casado, 2016) with an analytical strategy favoring phenolic and polyphenolic detection have also been applied. For that purpose, liquid chromatography

with different detection systems, such as ultraviolet (LC-UV) (Cetó, Sánchez, Serrano, Díaz-Cruz, & Núñez, 2020; Cetó et al., 2018), electrochemical detection (LC-ECD) (Serrano et al., 2018), or LC-HRMS (Barbosa, Saurina, Puignou, & Núñez, 2020a), has been evaluated.

This study aimed to prove the applicability of high-performance liquid chromatography with fluorescence detection (HPLC-FLD) fingerprints, strongly related to phenolic acid and polyphenolic compounds, as a chemical marker to authenticate the geographical origin of paprika by chemometrics. Therefore, the classification of samples belonging to the three Spanish PDO and two eastern European countries (the Czech Republic and Hungary) was evaluated by partial least squares regression-discriminant analysis (PLS-DA). Instead, partial least squares (PLS) regression was used to detect and quantitate paprika adulterations.

2. MATERIAL AND METHODS

2.1 Reagents and solutions

An Elix® 3 coupled to a Milli-Q® system (Millipore Corporation, Bedford, MA, USA) was used to obtain purified water, correspondingly filtered with a 0.22-µm nylon membrane. Moreover, UHPLC-supergradient methanol and acetonitrile were purchased from Panreac (Castellar del Vallès, Spain), while formic acid (96%) from Merck (Darmstadt, Germany).

2.2 Instrumentation

The HPLC-FLD fingerprints were obtained using a chromatographic system consisting of an Agilent 1100 Series HPLC instrument from Agilent Technologies (Waldbronn, Germany), equipped with a binary pump (G1312A), a degasser (G1379A),

an autosampler (G1329B), and a fluorescence detector (G1321A). Besides, the ChemStation software (Agilent Technologies) allowed the HPLC-FLD system control and data acquisition and processing.

For the LC separation, core-shell technology Kinetex C_{18} column (100 mm × 4.6 mm id., 2.6 µm particle size) and guard column (2 mm × 4.6 mm id, 2.6 µm particle size), both from Phenomenex (Torrance, CA, USA), as well as 0.1% (*v/v*) formic acid aqueous solution (solvent A) and acetonitrile (solvent B) as the mobile phase components, were used. This study's gradient elution program started with a 2 min-isocratic step at 40% solvent B, followed by a linear gradient elution up to 80% in 1 min, and an isocratic step at this last condition for 5 min. Subsequently, after a 2 min-lineal increase up to 100% solvent B, the mobile phase's composition was isocratically kept for 2 min. Afterward, 1 min-lineal decrease back to the initial conditions, and 5 min of isocratic elution for column re-equilibration, were set. The mobile phase flow rate was 500 $\mu L \cdot min^{-1}$, and the injection volume 5 µL. Moreover, for FLD acquisition, 310 and 380 nm were chosen as the excitation and emission wavelengths, respectively.

2.3 Samples

2.3.1 Paprika samples for the qualitative classification study

A total of 122 paprika samples from different countries —Spain, Hungary, and the Czech Republic— and types —hot, bittersweet, and sweet— were analyzed in this study for classificatory purposes. Among the Spanish paprika samples, 45 were distinguished with *La Vera* PDO, 18 with *Murcia* PDO, and 16 with *Mallorca* PDO. Instead, the 28 Hungarian samples came from the *Kalocsa* region, while the 15 Czech paprika samples' region was not labeled. *La Vera* samples were directly purchased from paprika production companies, whereas the others were bought in Czech, Hungarian, and Spanish

commercial supermarkets and markets. Moreover, a quality control (QC) sample, constituted by pooling 50 µL of each analyzed paprika sample extract, was also prepared.

2.3.2 Adulterated paprika samples for the quantitative study

Two different geographical origin adulteration cases —*La Vera vs. Murcia* and the Czech Republic *vs. Murcia*— were under study. In each case, calibration and external validation blends were prepared (five and three replicates of each set, respectively), mixing sweet samples (sweet smoked in the case of the Czech Republic) from the corresponding origins in proportions from 0 to 100% as shown in Table S1, and giving a total of 45 samples. Besides, for testing purposes, three replicates at 25, 50, and 75%, using different sweet samples to those previously employed, were also prepared in both cases. For these sets of samples, QC samples consisted of additional 50% adulterated samples.

2.4 Sample treatment and analysis

A straightforward sample treatment, previously developed for polyphenolic compound extraction from Spanish paprika (Cetó et al., 2018), based on solid-liquid extraction (SLE) with water:acetonitrile (20:80 *v/v*) was carried out. Briefly, 0.3 g of the sample were extracted with 3 mL of the extracting mix, stirred in a Vortex (Stuart, Stone, United Kingdom) for 1 min, sonicated (5510 Branson ultrasonic bath, Hampton, NH, USA) for 15 min, and centrifuged (ROTANTA 460 RS Centrifuge, Hettich, Germany) for 30 min at 4500 rpm. Finally, the resulting polyphenolic extract was filtered with a 0.22-µm nylon filter and preserved at 4°C in a glass injection vial until its analysis by HPLC-FLD.

In each sample sequence, samples were randomly injected to minimize instrumental drifts' influence in the chemometric models. Moreover, at the beginning and after every ten sample injections, an extracting solvent blank and a QC sample were also injected to control cross-contamination and metabolite behavior in the analytical system, respectively.

2.5 Data analysis

The obtained HPLC-FLD raw data were exported to Microsoft Excel (Microsoft, Inc., Redmond, WA, USA) spreadsheet for preprocessing, and then, the constructed matrices were subjected to principal component analysis (PCA), PLS-DA, or PLS regression, using the Solo 8.6 chemometrics software from Eigenvector Research (Manson, WA, USA). Details of the theoretical background of these statistical methodologies are addressed elsewhere (Massart et al., 1997).

While PCA was used to evaluate QC sample behavior, PLS-DA was employed for sample classificatory purposes and PLS regression to detect and quantitate paprika geographical origin fraud. Indistinctly of the chemometric method used, the construction of different data matrices was required: the X-data matrix, consisting of the HPLC-FLD fingerprints obtained, and the Y-data matrix for PLS-DA and PLS regression, defining each sample class or the corresponding percentage of adulteration, respectively. Moreover, before chemometric analysis, chromatographic fingerprints were pretreated by smoothing, baseline-correcting, aligning, and autoscaling to improve data quality. The most appropriate number of latent variables (LVs) in PLS-DA and PLS regression models was established at the first significant minimum point of the venetian blinds cross-validation (CV) error.

Both PLS-DA models' classification performance and PLS regression models' prediction ability were evaluated through external validation. On the one hand, because of the complexity of the classification issue, where many sample particularities (e.g., geographical origin and type) were involved, a classification decision tree constituted by consecutive PLS-DA models was built using the hierarchical model builder (HMB). While 60% of paprika samples (stratified random chosen) constructed the calibration PLS-DA models, the remaining 40% were used as the validation set. On the other hand, Table S1 shows the adulteration percentages used in the calibration and validation sets employed in the PLS regression analysis.

3. RESULTS AND DISCUSSION

3.1 HPLC-FLD chromatographic separation

In the last decades, FLD has emerged as an alternative and also as a complement to UV and MS detection of phenolic and polyphenolic molecules (Monasterio, Olmo-García, Bajoub, Fernández-Gutiérrez, & Carrasco-Pancorbo, 2016), since many of them —hydroxybenzoic, hydroxyphenylacetic, and hydroxycinnamic acids, tyrosols (Godoy-Caballero, Acedo-Valenzuela, & Galeano-Díaz, 2012), lignans (Selvaggini et al., 2006), and flavanols (Bakhytkyzy, Nuñez, & Saurina, 2018)— are susceptible to be detected by this detection system. In this context, and because of the high potential of polyphenols to address food authentication issues (Barbosa, Pardo-Mates, Puignou, & Núñez, 2017), HPLC-FLD fingerprinting has recently gained interest in this field. Its application has already proven excellent descriptive performance when analyzing phenolic/polyphenolic food extracts. For instance, HPLC-FLD fingerprints were successfully proposed as chemical descriptors to address the origin, variety, and roasting degree of coffee (Núñez, Martínez, Saurina, & Núñez, 2021), as well as to assess the varietal origin of extra-virgin

olive oil (Bajoub et al., 2017). Besides, in some applications, such as the nuts classification, they provided better discrimination ability than HPLC-UV (Campmajó, Saez-Vigo, Saurina, & Núñez, 2020).

This study aimed to develop an HPLC-FLD fingerprinting approach, strongly related to phenolic acid and polyphenolic composition, for paprika classification according to its geographical origin, as well as its fraud quantitation. Thus, characteristic and representative sample chromatograms were required for each given class. Because of the non-targeted nature of the proposed method, the optimization of the chromatographic gradient elution relied on obtaining chromatographic fingerprints with enough discriminant information in a suitable time (below 20 min) rather than looking for baseline resolved peaks. For that purpose, different binary gradient elution modes, using a C_{18} column and 0.1% (*v/v*) formic acid aqueous solution and an organic solvent (methanol or acetonitrile), which are the most common chromatographic separation conditions for the analysis of phenolic compounds in food samples (Lucci, Saurina, & Núñez, 2017), were tested in a sweet *La Vera* paprika sample. Methanol as the organic solvent of the mobile phase was discarded since many compounds were not eluted until reaching 95%. Therefore, different initial acetonitrile percentages and the combination of gradient and isocratic steps were applied. As a compromise between the number of detected peaks and the analysis time, Fig. 1 shows the chosen HPLC-FLD fingerprints (excitation and emission wavelengths of 310 and 380 nm, respectively), which follow the gradient elution program detailed in Section 2.2, for the different varieties of studied samples. For the subsequent chemometric analysis, only the range from 0 to 12 min was considered, avoiding the column re-equilibration step.

3.2 Geographical origin classification

The visual inspection of the obtained HPLC-FLD fingerprints, depicted in Fig. 1, allowed the detection of considerable qualitative variations in chromatographic peak distribution and intensity among the different geographical origin paprika samples under study. For instance, *La Vera* and *Mallorca* samples were characterized by distinctive chromatographic fingerprints comparing to the remaining regions, which at first glance showed more similarities. In addition, these features were reproducible among samples belonging to the same geographical origin since, in general, the paprika type slightly modified the HPLC-FLD fingerprint shape, allowing their use as a chemical marker to address paprika geographical origin authentication through chemometrics.

First, an exploratory chemometric analysis through PCA was performed to assess the results' validity and ensure the lack of systematic errors during the sample sequence by studying QC sample behavior. Thus, a 135 × 1667 (samples × variables) dimension data matrix, constructed with the chromatographic fingerprints registered for each paprika and QC samples, was subjected to PCA. As a result, Fig. S1 presents the scatter plot for scores on the PC2-PC1 (explaining 85.78% of the variance), showing a clear group of QC samples in the middle of the plot, and therefore, the absence of a trend associated with the analytical system and the suitability of data pretreatment. Moreover, according to geographical origin, several sample groups and trends can be observed in the PCA scores plot, indicating the suitability of HPLC-FLD fingerprints when used as sample chemical descriptors.

Then, after excluding QC samples, the resulting X-data matrix (122 × 1667) and the Y-data matrix (122 × 1), giving the geographical origin, were exploited in a preliminary supervised approach by PLS-DA. Three LVs were selected to build the PLS-DA model that remarkably allowed the discrimination of some paprika regions. In this line, Fig. 2A shows the scores plot of LV1 *vs.* LV2, where *La Vera* —in the right side of the plot

displaying positive LV1 values— and *Mallorca* samples —on the bottom of the diagram displaying negative LV2 values— are distinguished. Besides, the remaining samples seem to follow a trend along the LV2 according to their geographical origin. Since in this PLS-DA model, LV construction was mainly influenced by *La Vera* and *Mallorca* classes, complete visual discrimination between the remaining three regions was not achieved. Therefore, a new supervised model for *Murcia*, Hungarian, and Czech paprika samples was built with six LVs, explaining a Y-variance of 94.48%. The corresponding scatter plot for scores on the LV2-LV1 is depicted in Fig. 2B, providing a great classification for the studied classes, although sweet smoked Czech paprika appears separated from the other Czech samples. This may indicate that the smoking procedure affects the phenolic profile, which agrees with Barbosa et al. that reported variations in the content of several phenolic compounds —such as syringaldehyde, ferulic acid, nepetin 7-glucoside, and hesperidin— found in sweet smoked and non-smoked Czech paprika (Barbosa, Campmajó, et al., 2020).

Because of the arduousness of the classification under study, which involved a wide number of classes, a single PLS-DA model was inadequate to solve the authentication issue and, therefore, a classification decision tree, constituted by smaller two-input class PLS-DA calibration models —acting as the rule nodes—, was proposed. In this line, Fig. S2 depicts the flow-chart of the designed classification decision tree and details data matrices dimensions and LVs used in the four rule nodes: 1) *La Vera vs.* others, 2) *Mallorca vs.* others (without including *La Vera* samples), 3) the Czech Republic *vs.* others (without including *La Vera* and *Mallorca* samples), and 4) *Murcia vs.* Hungary. As previously mentioned in Section 2.5, PLS-DA calibration models were built using 60% of the analyzed paprika samples, while the remaining 40% were used to carry out external validation. Table 1 summarises some statistical parameters such as the root-mean-square

error of calibration (RMSEC), cross-validation (RMSECV), or external validation (RMSEEV), and the corresponding values of R^2. The low values of RMSECV and their similarity to RMSEC ones ensured good internal consistency and prevented overfitting. Besides, the high values of R^2 for the prediction and the low RMSEEV values, suggested a satisfactory predictive capability of the developed PLS-DA models. In this line, after performing external validation, an excellent classification accuracy of 97.9% was achieved. Moreover, all sample classes showed a sensitivity (capability to detect true positives) and specificity (capability to detect true negatives) of 100%, except for Hungary that presented a sensitivity of 91.7%, and *Murcia* that provided a specificity of 97.6%.

3.3 Detection and quantitation of geographical origin fraud

Because of the excellent classification results obtained with the proposed methodology, HPLC-FLD fingerprints were also used to detect and quantitate paprika geographical origin fraud. Thus, as previously mentioned in Section 2.3.2, two different paprika adulteration scenarios were evaluated (*La Vera vs. Murcia* and the Czech Republic *vs. Murcia*) by analyzing a set of mixed sweet samples (sweet smoked in the case of the Czech Republic) as detailed in Table S1.

Since the obtained chromatographic fingerprints varied according to blend percentage, they were subjected to PLS regression to predict the blending degree. However, before PLS regression analysis and aiming to check the correct behavior of the QC samples, which corresponded to a 50% adulterated sample, PCA was performed for both data matrices (61 × 1667). When observing the corresponding scores plots, QCs are located in the center in *La Vera vs. Murcia* set (Fig. S3A), while they are grouped displaying negative values of PC2 in the Czech Republic *vs. Murcia* set (Fig. S3B),

proving the reliability of the subsequent chemometric results. Besides, in both PCA scores plots, PC1 could be related to the blending percentage as it seems to increase in samples from the left to the right.

As previously indicated in the Material and Methods Section, PLS regression models were established from the calibration data set of standard samples. In this line, for both of the adulteration scenarios under study, an X-data matrix (30 × 1667) —containing the HPLC-FLD fingerprints of calibration samples— and a Y-data matrix (30 × 1) — specifying the percentage of adulteration— were exploited by this multivariate regression technique. Afterward, external validation was performed to evaluate the prediction ability of the build PLS regression models. Table 2 sums up the number of LVs used in each calibration PLS regression model, as well as some statistical parameters related to calibration, cross-validation, and external validation performance. Good calibration models were constructed, as indicated by the low RMSEC values, bias values tending towards 0 and determination coefficients R^2 (C) ≥ 0.999. Besides, excellent results were obtained for the external validation (see scatter plots of measured *vs.* predicted percentages of adulteration in Fig S4), with overall RMSEEV values below 1.6%.

Given these results, the built models' applicability was tested by analyzing new independent mixtures at 25, 50, and 75% (expected adulteration percentages in real fraud) not used for building the calibration models. As reported in Table 2, satisfactory root-mean-square error of prediction (RMSEP) values of 10.7 and 3.7% were obtained for *La Vera vs. Murcia* and the Czech Republic *vs. Murcia* cases, respectively. Moreover, as shown in Fig 3, where the obtained scatter plots of measured *vs.* predicted percentages of adulteration are depicted, prediction ability slightly decreased when increasing the percentage of *La Vera* sample in the test mix of *La Vera vs. Murcia* case (Fig 3A). Instead,

a similar predictive performance was observed between the studied percentages in the Czech Republic *vs. Murcia* one (Fig 3B).

It should be pointed out that each PLS model is exclusively valid for the specific problem for which it has been designed. For instance, a model to predict *La Vera* paprika's adulteration with *Murcia* samples has provided suitable results for that given problem. On the contrary, the model prediction performance may be poor when the adulteration comes from another region. In a more general context, if *La Vera* samples were adulterated with paprika from any (known or unknown) region, the calibration matrix should reflect this variability, including samples from different geographical areas at various blending percentages. This situation represents a higher experimental and chemometric challenge since it requires the preparation of a wide range of standards to cover all the experimental variance.

4. CONCLUSIONS

This study suggests the suitability of phenolic and polyphenolic extract HPLC-FLD fingerprints (using an excitation wavelength of 310 nm and an emission wavelength of 380 nm), when combined with chemometrics, as chemical markers to classify European paprika samples according to their geographical origin and detect and quantitate their blend percentage. In this line, an excellent classification accuracy of 97.9%, as well as prediction errors below 10.7% have been reached, respectively. Therefore, the proposed HPLC-FLD fingerprinting method can be used as a reliable and straightforward complementary method to prevent geographical origin fraud of paprika of European origin. Moreover, although the phenolic HPLC-FLD fingerprints could be slightly modified by the harvesting year, mainly because of climate conditions, differences related to geographical origin may prevail. In this context, the maintenance of the analysis's

representativeness will require the inclusion of further paprika samples in the calibration chemometric models.

Declaration of competing interest

There are no conflicts of interest to declare.

Acknowledgements

Guillem Campmajó thanks the University of Barcelona for the PhD APIF fellowship.

Funding

The authors are grateful for the support received from the Agency for Administration of University and Research Grants (Generalitat de Catalunya), through the grants 2017SGR-171 and 2017SGR-310.

References

Arrizabalaga-Larrañaga, A., Campmajó, G., Saurina, J., Núñez, O., Santos, F. J., & Moyano, E. (2020). Determination of capsaicinoids and carotenoids for the characterization and geographical origin authentication of paprika by UHPLC–APCI–HRMS. *Lwt*, (November). https://doi.org/10.1016/j.lwt.2020.110533

Bajoub, A., Medina-Rodríguez, S., Gómez-Romero, M., Ajal, E. A., Bagur-González, M. G., Fernández-Gutiérrez, A., & Carrasco-Pancorbo, A. (2017). Assessing the varietal origin of extra-virgin olive oil using liquid chromatography fingerprints of phenolic compound, data fusion and chemometrics. *Food Chemistry*, *215*, 245–255. https://doi.org/10.1016/j.foodchem.2016.07.140

Bakhytkyzy, I., Nuñez, O., & Saurina, J. (2018). Determination of flavanols by liquid chromatography with fluorescence detection. Application to the characterization of cranberry-based pharmaceuticals through profiling and fingerprinting approaches. *Journal of Pharmaceutical and Biomedical Analysis*, *156*, 206–213. https://doi.org/10.1016/j.jpba.2018.04.031

Ballin, N. Z., & Laursen, K. H. (2019). To target or not to target? Definitions and nomenclature for targeted versus non-targeted analytical food authentication. *Trends in Food Science and Technology*, *86*, 537–543. https://doi.org/10.1016/j.tifs.2018.09.025

Barbosa, S., Campmajó, G., Saurina, J., Puignou, L., & Núñez, O. (2020). Determination of Phenolic Compounds in Paprika by Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectrometry: Application to Product Designation of Origin Authentication by Chemometrics. *Journal of Agricultural and Food Chemistry*, *68*, 591–602. https://doi.org/10.1021/acs.jafc.9b06054

Barbosa, S., Pardo-Mates, N., Puignou, L., & Núñez, O. (2017). The Role of

Polyphenols and Polyphenolic Fingerprinting Profiles in the Characterization and Authentication of Natural Food Products. In J. C. Taylor (Ed.), *Advances in Chemistry Research. Volume 40*. Nova Science Publishers, Inc.

Barbosa, S., Saurina, J., Puignou, L., & Núñez, O. (2020a). Classification and authentication of paprika by UHPLC-HRMS fingerprinting and multivariate calibration methods (PCA and PLS-DA). *Foods*, *9*, 486. https://doi.org/10.3390/foods9040486

Barbosa, S., Saurina, J., Puignou, L., & Núñez, O. (2020b). 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*, 1623. https://doi.org/10.3390/molecules25071623

Campmajó, G., Saez-Vigo, R., Saurina, J., & Núñez, O. (2020). High-performance liquid chromatography with fluorescence detection fingerprinting combined with chemometrics for nut classification and the detection and quantitation of almond-based product adulterations. *Food Control*, *114*, 107265. https://doi.org/10.1016/j.foodcont.2020.107265

Cetó, X., Sánchez, C., Serrano, N., Díaz-Cruz, J. M., & Núñez, O. (2020). Authentication of paprika using HPLC-UV fingerprints. *Lwt*, *124*, 109153. 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*, 4479. https://doi.org/10.3390/s18124479

Cuadros-Rodríguez, L., Ruiz-Samblás, C., Valverde-Som, L., Pérez-Castaño, E., & González-Casado, A. (2016). Chromatographic fingerprinting: An innovative

approach for food “identitation” and food authentication - A tutorial. *Analytica Chimica Acta*, *909*, 9–23. https://doi.org/10.1016/j.aca.2015.12.042

Danezis, G. P., Tsagkaris, A. S., Camin, F., Brusic, V., & Georgiou, C. A. (2016). Food authentication: Techniques, trends & emerging approaches. *TrAC - Trends in Analytical Chemistry*, *85*, 123–132. https://doi.org/10.1016/j.trac.2016.02.026

European Comission. (2019). *2019 Annual Report. The EU Food Fraud Network and the Administrative Assistance and Cooperation System*. 1–19. https://doi.org/10.2875/326318

European Comission. (2020). eAmbrosia - the EU geographical indications register. Retrieved December 13, 2020, from https://ec.europa.eu/info/food-farming-fisheries/food-safety-and-quality/certification/quality-labels/geographical-indications-register/

Fiamegos, Y., Dumitrascu, C., Papoci, S., & de la Calle, M. B. (2021). Authentication of PDO paprika powder (Pimentón de la Vera) by multivariate analysis of the elemental fingerprint determined by ED-XRF. A feasibility study. *Food Control*, *120*, 107496. https://doi.org/10.1016/j.foodcont.2020.107496

Godoy-Caballero, M. P., Acedo-Valenzuela, M. I., & Galeano-Díaz, T. (2012). Simple quantification of phenolic compounds present in the minor fraction of virgin olive oil by LC-DAD-FLD. *Talanta*, *101*, 479–487. https://doi.org/10.1016/j.talanta.2012.09.063

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*, 1–25. https://doi.org/10.3390/antiox8100469

Lucci, P., Saurina, J., & Núñez, O. (2017). Trends in LC-MS and LC-HRMS analysis and characterization of polyphenols in food. *TrAC - Trends in Analytical*

Chemistry, *88*, 1–24. https://doi.org/10.1016/j.trac.2016.12.006

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

Monasterio, R. P., Olmo-García, L., Bajoub, A., Fernández-Gutiérrez, A., & Carrasco-Pancorbo, A. (2016). Potential of LC coupled to fluorescence detection in food metabolomics: Determination of phenolic compounds in virgin olive oil. *International Journal of Molecular Sciences*, *17*, 1627. https://doi.org/10.3390/ijms17101627

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

Núñez, N., Martínez, C., Saurina, J., & Núñez, O. (2021). High-performance liquid chromatography with fluorescence detection fingerprints as chemical descriptors to authenticate the origin, variety and roasting degree of coffee by multivariate chemometric methods. *Journal of the Science of Food and Agriculture*, *101*, 65–73. https://doi.org/10.1002/jsfa.10615

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*, 521–530. https://doi.org/10.5601/jelem.2017.22.4.1497

Palacios-Morillo, A., 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

Selvaggini, R., Servili, M., Urbani, S., Esposto, S., Taticchi, A., & Montedoro, G. F. (2006). Evaluation of phenolic compounds in virgin olive oil by direct injection in high-performance liquid chromatography with fluorometric detection. *Journal of Agricultural and Food Chemistry*, *54*, 2832–2838. https://doi.org/10.1021/jf0527596

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*, 296–301. https://doi.org/10.1016/j.talanta.2018.06.085

Škrovánková, S., Mlček, J., Orsavová, J., Juríková, T., & Dřímalová, P. (2017). Polyphenols content and antioxidant activity of paprika and pepper spices. *Potravinarstvo*, *11*, 52–57. https://doi.org/10.5219/695

The European Parliament and the Council of the European Union. (2012). Regulation (EU) No 1151/2012 of the European Parliament and of the Council of 21 November 2012 on quality schemes for agricultural products and foodstuffs. *Official Journal of the European Union*, *L343*, 1–29.

Topuz, A., Dincer, C., Özdemir, K. S., Feng, H., & Kushad, M. (2011). Influence of different drying methods on carotenoids and capsaicinoids of paprika (Cv., Jalapeno). *Food Chemistry*, *129*, 860–865. https://doi.org/10.1016/j.foodchem.2011.05.035

Figure captions

Figure 1. HPLC-FLD fingerprints, acquired at an excitation wavelength of 310 nm and an emission wavelength of 380 nm, for a selected sample within each paprika region and type. Black and light grey indicate the hot and sweet type, respectively, while medium gray corresponds to bittersweet in *La Vera* samples or smoked sweet in the Czech Republic ones.

Figure 2. (A) PLS-DA scores plot of LV1 *vs.* LV2, using the HPLC-FLD fingerprints acquired for all the paprika samples analyzed. (B) PLS-DA scores plot of LV1 *vs.* LV2, using only *Murcia*, Hungary, and the Czech Republic HPLC-FLD fingerprints.

Figure 3. Prediction test PLS results: *La Vera vs Murcia* (on the left side) and the Czech Republic *vs Murcia* (on the right side) scatter plot of measured *vs* predicted percentages of paprika blend level. Red and black symbols indicate prediction and calibration samples, respectively. Black dashed line corresponds to the theoretical diagonal line, while the red line to the experimental adjusted one.

Table 1. Calibration, cross-validation, and external validation statistical parameters obtained for each of the PLS-DA models used to build the classification decision tree.

	1) *La Vera vs.* others	**2) *Mallorca vs.* others**	**3) Czech R. *vs.* others**	**4) *Murcia vs.* Hungary**
RMSEC	0.166	0.126	0.121	0.152
RMSECV	0.179	0.135	0.154	0.211
RMSEEV	0.135	0.133	0.141	0.260
R^2 (C)	0.882	0.905	0.920	0.906
R^2 (CV)	0.863	0.892	0.871	0.818
R^2 (EV)	0.925	0.889	0.898	0.730

Table 2. Number of LVs and calibration, cross-validation, external validation, and prediction statistical parameters obtained for each of the PLS regression models used to determine the paprika blend percentage.

	1) *La Vera vs. Murcia*	**2) Czech R. *vs. Murcia***
LVs	4	4
RMSEC	0.732	0.933
RMSECV	1.440	1.343
RMSEEV	1.543	0.974
RMSEP	10.701	3.730
R^2 (C)	1.000	0.999
R^2 (CV)	0.998	0.998
R^2 (EV)	0.997	0.999
R^2 (P)	0.996	0.995

FIGURE 1

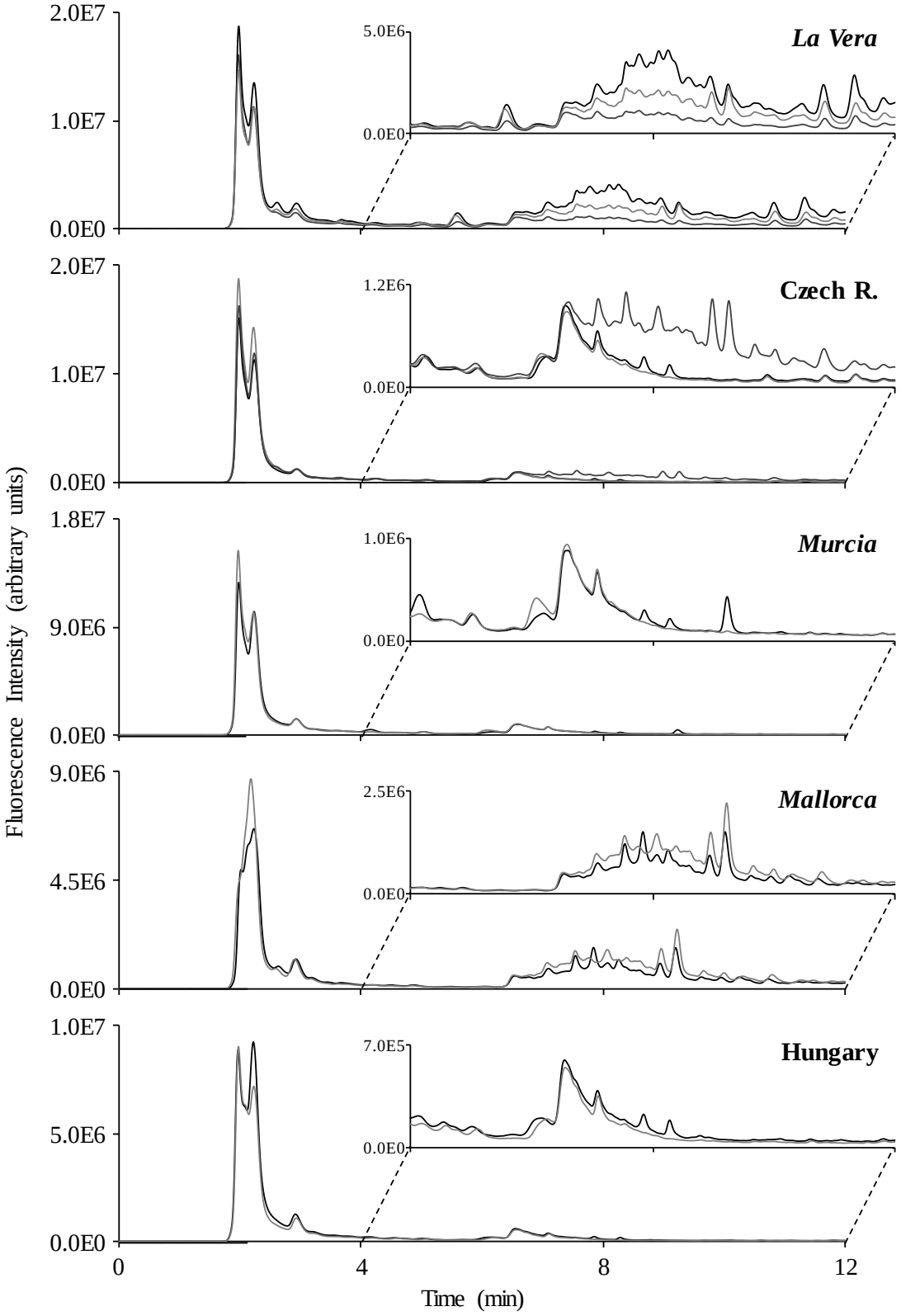

La Vera
Czech R.
Murcia
Mallorca
Hungary
Fluorescence Intensity (arbitrary units)
Time (min)
2.0E7
1.0E7
0.0E0
5.0E6
1.2E6
1.8E7
9.0E6
2.5E6
4.5E6
7.0E5
0
4
8
12

FIGURE 2

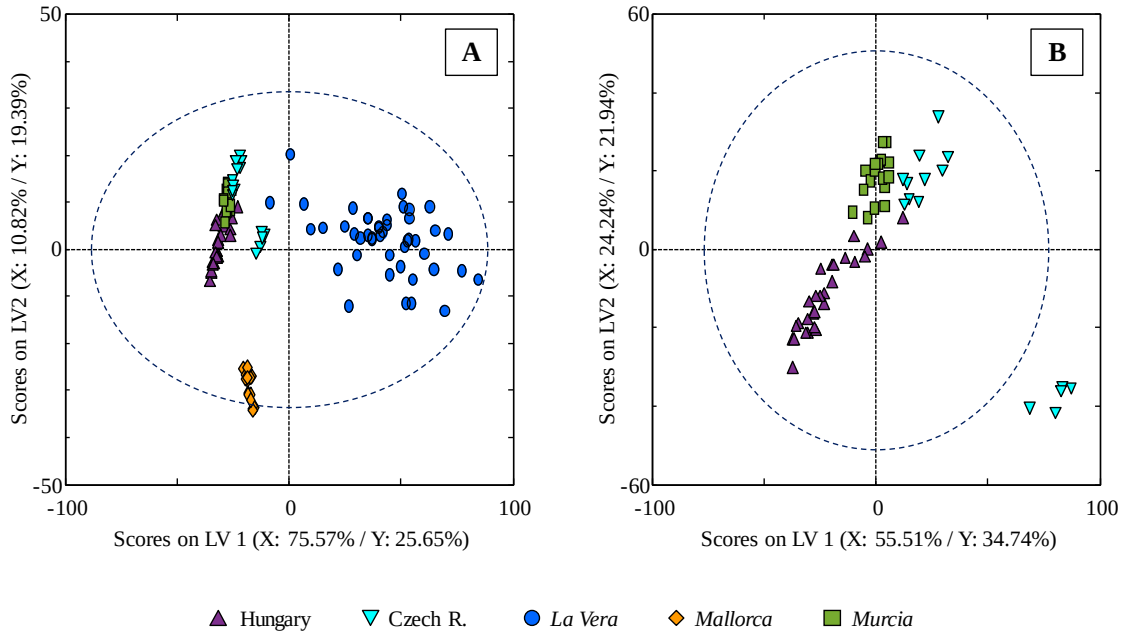
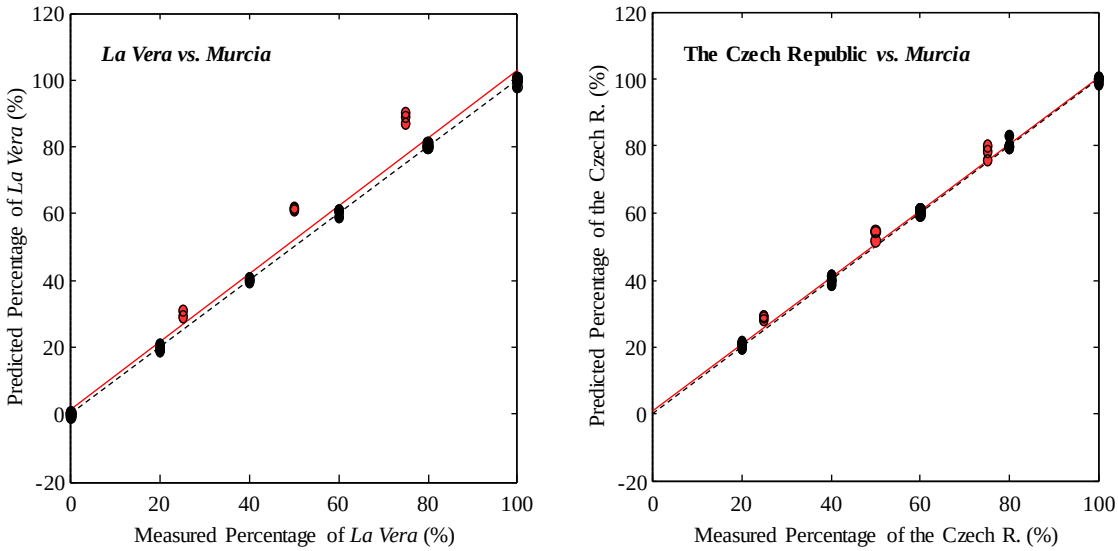

A
B
Scores on LV2 (X: 10.82% / Y: 19.39%)
Scores on LV 1 (X: 75.57% / Y: 25.65%)
Scores on LV2 (X: 24.24% / Y: 21.94%)
Scores on LV 1 (X: 55.51% / Y: 34.74%)
50
0
-50
-100
0
100
60
0
-60
-100
0
100
Hungary
Czech R.
La Vera
Mallorca
Murcia

FIGURE 3


La Vera vs. Murcia
Predicted Percentage of La Vera (%)
Measured Percentage of La Vera (%)
120
100
80
60
40
20
0
-20
0
20
40
60
80
100
The Czech Republic vs. Murcia
Predicted Percentage of the Czech R. (%)
Measured Percentage of the Czech R. (%)

Supplementary Material

Assessment of paprika geographical origin fraud by high-performance liquid chromatography with fluorescence detection (HPLC-FLD) fingerprinting

Guillem Campmajó[a,b,*], Luis R. Rodríguez-Javier[a], Javier Saurina[a,b], 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), E08921 Santa Coloma de Gramenet, Spain.

[c] Serra Húnter Fellow, Generalitat de Catalunya, Rambla de Catalunya 19-21, E08007 Barcelona, Spain.

* Corresponding author: Guillem Campmajó

(Tel.: 34-93-402-1286, e-mail: gcampmajo@ub.edu)

Table S1. Description of the samples analyzed in the PLS regression adulteration studies as calibration or validation set. *La Vera vs. Murcia* and the Czech Republic *vs. Murcia* cases were under study.

	La Vera / the Czech R., %	*Murcia*, %
Calibration set	0	100
	20	80
	40	60
	60	40
	80	20
	100	0
Validation set	15	85
	25	75
	50	50
	75	25
	85	15

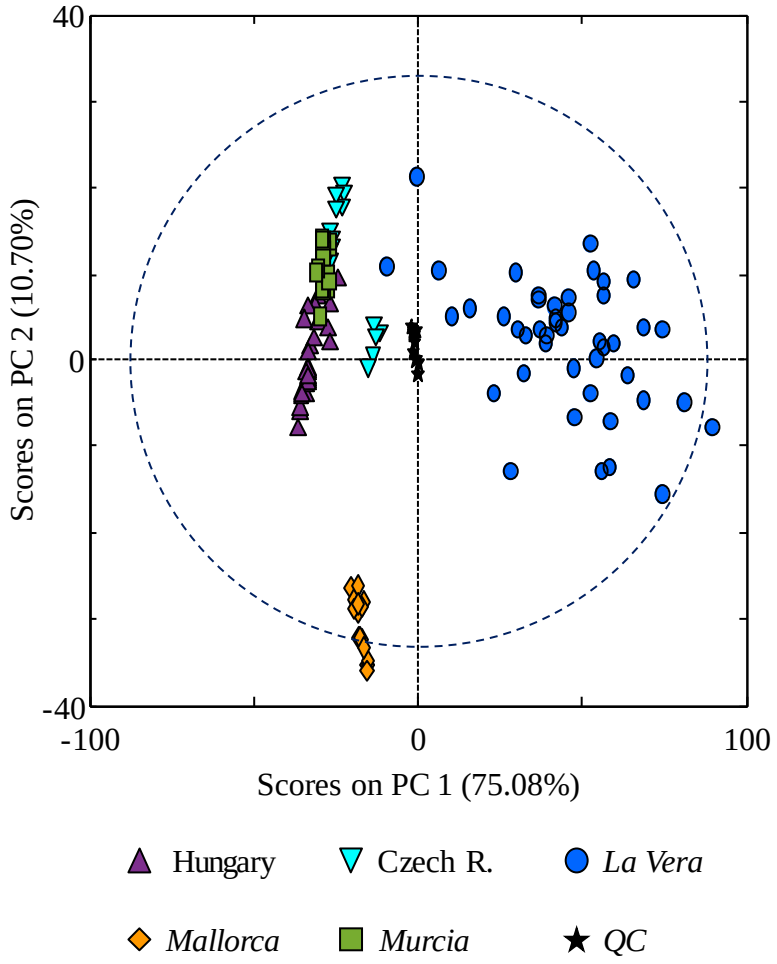


Figure S1. PCA scores plot showing the QC samples' correct behavior in the classification study and some trends associated to sample geographical origin.

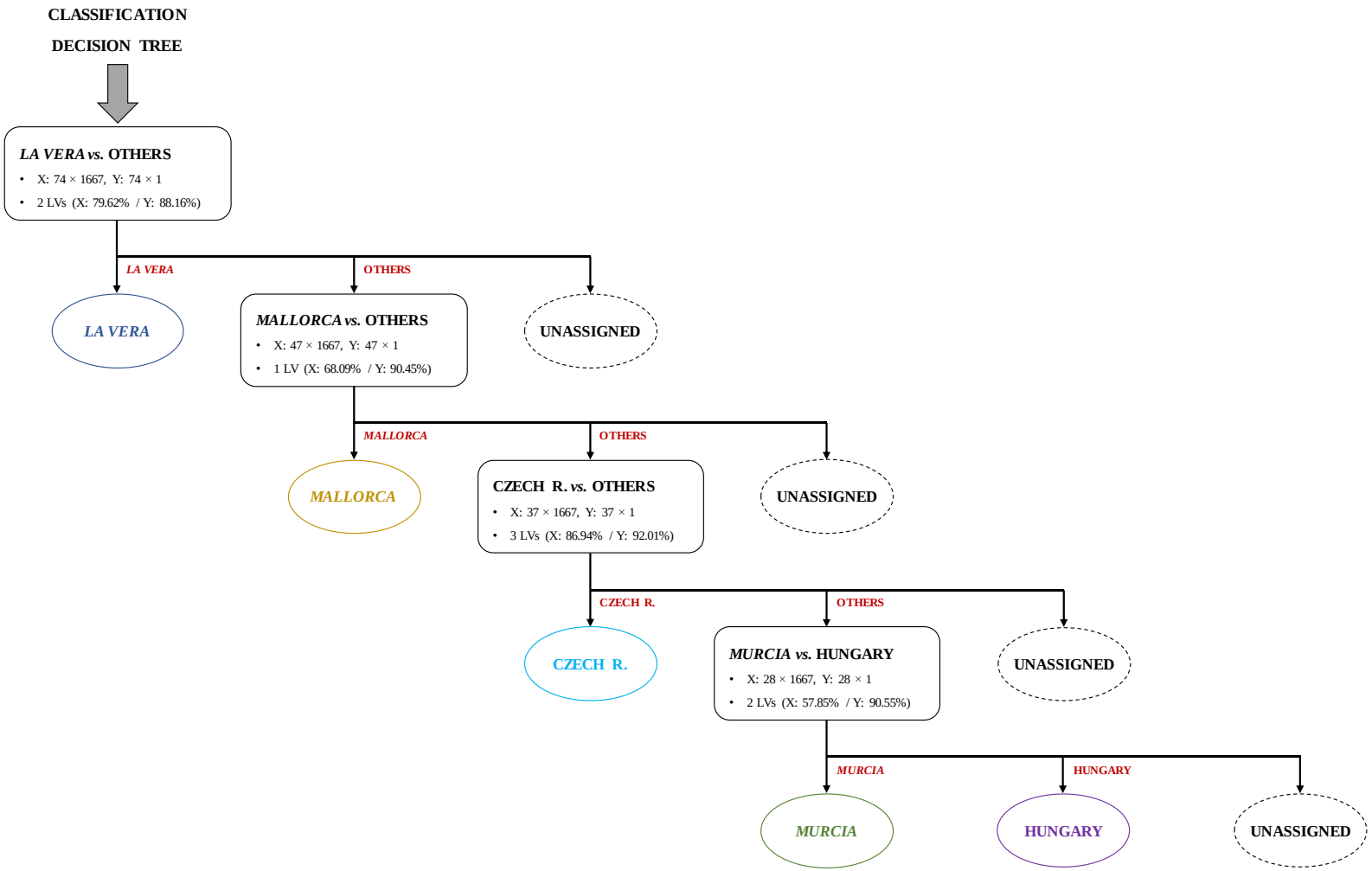


Figure S2. Flow-chart of the designed classification decision tree build using PLS-DA models as the rule nodes. Data matrices dimensions and LVs used to construct the calibration models are detailed.

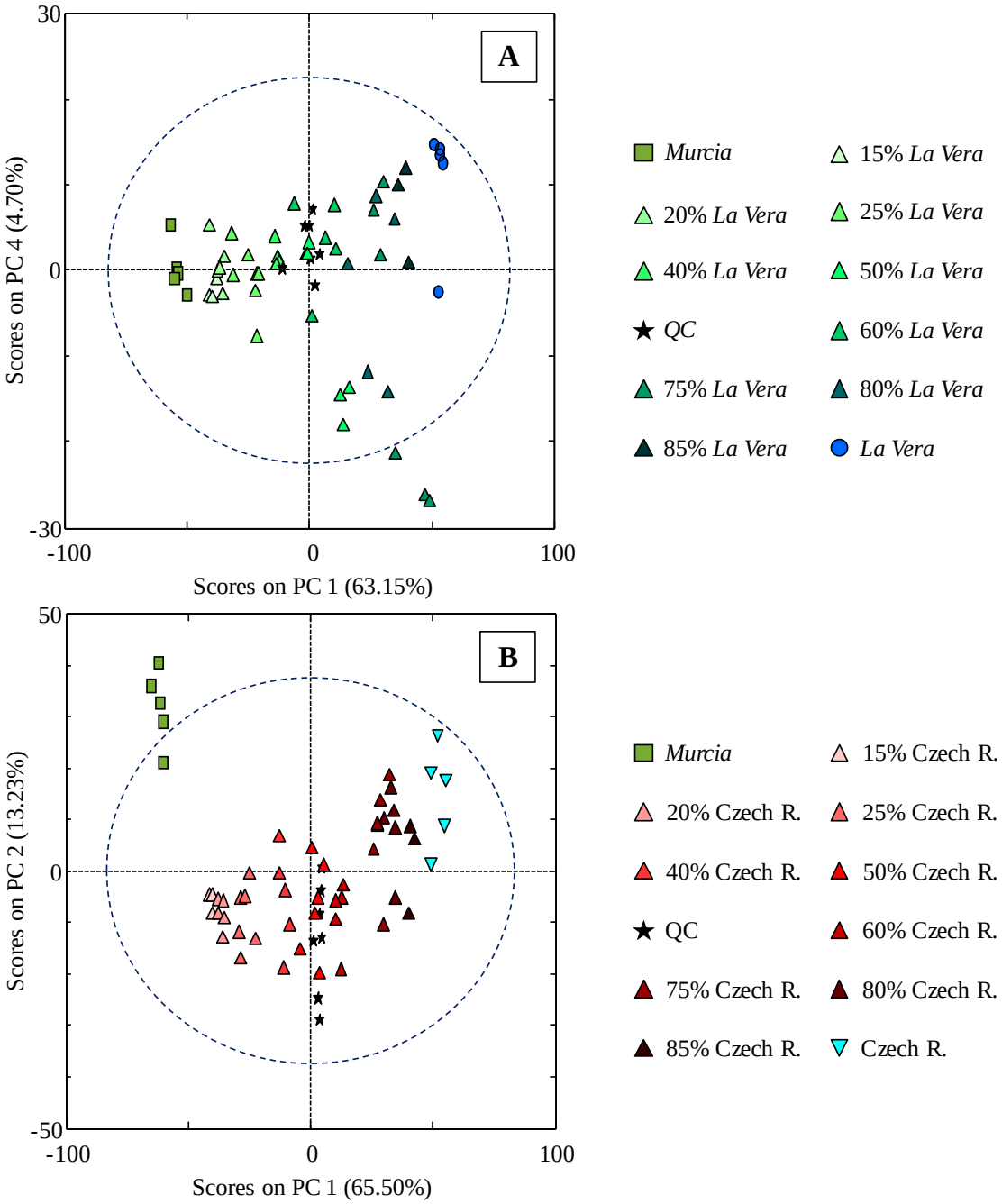


Figure S3. PCA scores plot showing the QC samples' correct behavior in the study to detect and quantitate paprika geographical origin fraud: (A) La Vera *vs*. Murcia PLS case and (B) the Czech Republic *vs*. Murcia case.

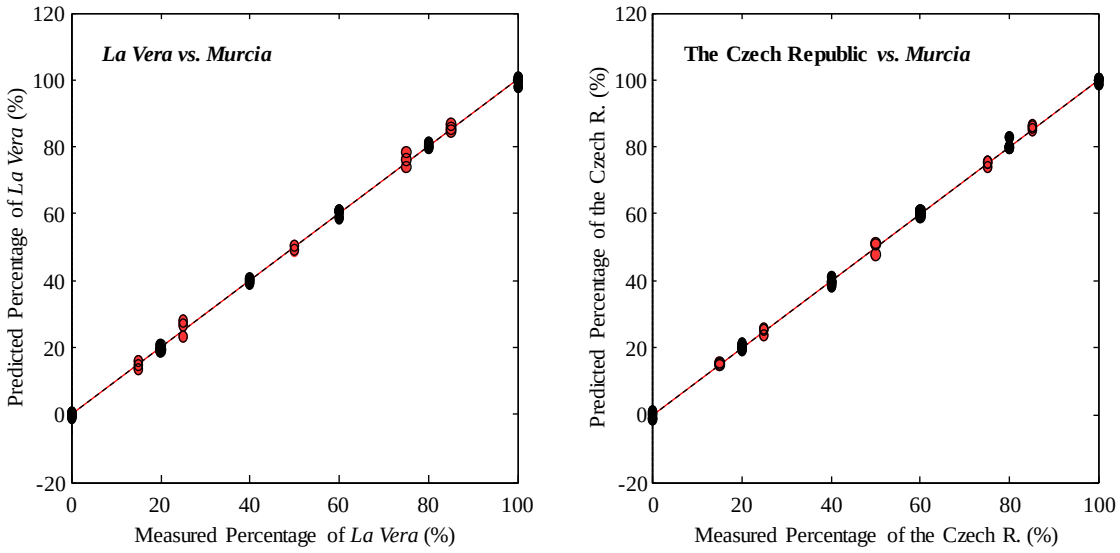


Figure S4. External validation PLS results: *La Vera vs. Murcia* (on the left side) and the Czech Republic *vs. Murcia* (on the right side) scatter plot of measured *vs* predicted percentages of paprika blend level. Red and black symbols indicate external validation and calibration samples, respectively. Black dashed line corresponds to the theoretical diagonal line, while the red line to the experimental adjusted one.