Authentication of paprika using HPLC-UV fingerprints

Xavier Cetó[a], Cristina Sánchez[a], Núria Serrano[a], José Manuel Díaz Cruz[a,*], Oscar Núñez [a,b,c]

[a] *Department of Chemical Engineering and Analytical Chemistry, Faculty of Chemistry, Universitat de Barcelona, 08028 Barcelona, Spain*

[b] *Research Institute in Food Nutrition and Food Safety, University of Barcelona, 08901, Santa Coloma de Gramanet, Spain*

[c] *Serra Hunter Fellow, Generalitat de Catalunya, Barcelona, Spain*

*to whom correspondence should be addressed, e-mail: josemanuel.diaz@ub.edu

Abstract

In this work we combine simple extraction and HPLC-UV methodologies with chemometric pattern-recognition strategies in order to obtain characteristic fingerprints of phenolic compounds that allow the authentication of paprika samples. To illustrate the potential of the proposed approach, two different adulteration scenarios were considered, namely adulteration of paprika based on its type (sweet, bittersweet and spicy) as well as on its region (Murcia, la Vera and Czech Republic). Upon preparation of a proper set of samples, they were analysed using a C_{18} reversed-phase column and registered chromatograms were then compressed employing fast Fourier transform (FFT) to reduce the large dimensionality of the data set, while preserving all relevant features. Next, data were analysed using linear discriminant analysis (LDA) for the

qualitative discrimination of adulterated samples, followed by partial least-squares regression (PLS) modelling to quantitatively assess the adulteration degree.

Keywords: paprika; food authentication; adulteration; liquid chromatography; partial least-squares regression

1. Introduction

In an effort to promote and protect the quality of regional foods, geographical indications (e.g. protected designation of origin, PDO), regulatory boards assess that producers comply with the specific technical production conditions and performs regular controls (including sensory and analytical examinations, as well as stock statements and verification of movements) (Dias & Mendes, 2018). However, despite these controls, there is a demand of new analytical low-cost methods needed to assess both authenticity and fulfilment of quality standards (Danezis, Tsagkaris, Camin, Brusic, & Georgiou, 2016; Galvin-King, Haughey, & Elliott, 2018). Particularly, this is critical when trying to assess the authenticity of local natural foods. Unfortunately, there is a lack of methods able to classify food samples, since usually there is not any specific compound directly related to food origin or quality that could be determined using conventional analytical techniques.

A system capable to perform such task should simultaneously detect a large spectrum of compounds and provide comprehensive information of the sample. In this regards, current approaches for quality control are shifting from compound-oriented to pattern-oriented strategies (Cavanna, Righetti, Elliott, & Suman, 2018; Cuadros-Rodríguez, Ruiz-Samblás, Valverde-Som, Pérez-Castaño, & González-Casado, 2016; Esteki, Shahsavari, & Simal-Gandara, 2019; Zeng et al., 2008). This means developing methodologies for the simultaneous detection of many compounds and the further pattern recognition analysis of the data, instead of focusing on the quantification of a few specific substances. The main advantage of pattern-oriented approaches is that they do not require any prior knowledge of the sample composition in order to succeed, but even more, they can be used to assess those key (bio)markers.

In recent years, there has been an increased interest and knowledge on the presence of bioactive compounds in food, as well as on the role of such substances on the quality and health benefits of food products, which has to be guaranteed (Johanningsmeier, Harris, & Klevorn, 2016; Kris-Etherton et al., 2002). Bioactive compounds distribution in most natural foods can be related to the specific products varieties, processing technologies, production regions and climate conditions (Baenas, Belović, Ilic, Moreno, & García-Viguera, 2019; Mudrić et al., 2017). Most of these compounds are powerful antioxidants needed for the functioning of plant cells, with huge health benefits upon its ingestion as they can act as free radical scavengers and inhibitors of lipoprotein oxidation, providing a protective effect against aging pathologies like cardiovascular diseases or cancers mutation (Kim et al., 2016; Quideau, Deffieux, Douat-Casassus, & Pouységu, 2011).

Paprika, sometimes also referred to as chilli pepper, is a characteristic red seasoning powder obtained from the drying and grinding of certain varieties of red peppers (*Capsicum annuum L.*) (Pérez-Gálvez, et al., 2005). There are three important varieties of paprika: sweet, bittersweet, and spicy. The two most known varieties of paprika in Spain, and the only ones with a PDO, come from the region of "la Vera" in Cáceres (Extremadura) and from "Murcia" (Commission Regulation (EEC), 11 February 2000, 24 November 2006). Among the different bioactive substances found in paprika, phenolic compounds are especially important, and their distribution may be related to the different red pepper varieties (Baenas et al., 2019; Mudrić et al., 2017; Quideau et al., 2011).

In this context, herein we investigate on the capabilities of combining liquid chromatography, to obtain a profile of the phenolic content of paprika's, with chemometric methods such as linear discriminant analysis (LDA) or partial least-

squares regression (PLS) for the extraction of a characteristic fingerprint that allow the authentication of paprika samples.

2. Experimental

2.1 Reagents and materials

Methanol (UHPLC-gradient grade), formic acid 98%, acetonitrile, absolute ethanol and acetone were purchased from Panreac (Barcelona, Spain). Standards of phenolic compounds were obtained from Sigma-Aldrich (St. Louis, MO, USA), from which stock solutions of 1000 mg/L were prepared in methanol and stored in amber glass vials. Deionized water was obtained from a Milli-Q system (Millipore, Billerica, MA, USA).

2.2 Paprika samples

Authentication of paprika was studied from two different points of view. On the one side, adulteration with paprika from other regions was considered, whereas on the other side, adulteration with paprika from other varieties was also evaluated.

To this aim, paprika samples from three different regions (La Vera, Murcia and Czech Republic) were considered. Among the samples of every region, there were different types of paprika (sweet, bittersweet and spicy in the samples from La Vera, and sweet and spicy in the samples from Murcia and Czech Republic). The samples were purchased directly from producers or from different local shops. Adulteration of paprika was made in two ways. In the study about paprika types, 24 mixtures were prepared with different proportions of the varieties sweet, bittersweet and spicy of paprika from La Vera. For each variety, 12 different proportions were considered (0, 0.01, 0.02, 0.1, 0.2, 0.4, 0.6, 0.8, 0.9, 0.98, 0.99 and 1), according to the design of 24

experiments summarised in Table 1. Every mixture was prepared twice, which means 48 samples to be analysed. In the study about paprika regions, 24 mixtures were prepared with samples of the type spicy from the regions of La Vera, Murcia and Czech Republic. As in the previous study, the experimental design of Table 1 was used and the samples were prepared twice, also producing a total of 48 samples.

Prior to its analysis, paprika samples were subjected to a extraction stage by sonication and centrifugation in water:acetonitrile (20:80 v/v) (Cetó et al., 2018). Briefly, 0.3 g of paprika were weighted, dispersed in 3 mL of solution and vortexed for 1 min. Next, samples were sonicated for 15 min and centrifuged at 4500 rpm for 30 min. Finally, the supernatant was filtered through 0.45 μm nylon filters, and samples stored at 4 ºC until their analysis.

2.3 Chromatographic analysis

HPLC analysis was carried out in an Agilent 1200 Series instrument (Palo Alto, CA, USA) equipped with a G1311A quaternary pump, a G1322A vacuum degasser, a G1329A autosampler and a G1314B ultraviolet-visible detector; all of them controlled with the Agilent ChemStation software package.

Chromatographic fingerprints were obtained with a reverse phase Kinetex C_{18} column (2.6 μm C18 100 Å, 100 x 4.6 mm) from Phenomenex (Torrance, CA, USA) at room temperature. For the elution, a mixture of Milli-Q water containing 0.1% formic acid (solvent A) and methanol (solvent B) were used as the mobile phase components at a flow rate of 1 mL/min, and with the following gradient: 0-2 min, isocratic step at 5% B; 2-4 min linear gradient from 5 to 25% B; 4-12 min, at 25% B; 12-14 min, from 25 to 45% B; 14-16 min, at 45% B; 16-18 min, from 45 to 95% B; 18-20 min, at 95% B; 20-21 min, back to initial conditions at 5% B; and from 21-30 min, at 5% B for column re-

equilibration (Cetó et al., 2018). Injection volume was 20 μL, and UV absorption registered at 280 nm every 291 ms.

Every paprika sample was injected by triplicate, which generated 144 chromatograms in the study about paprika types and 144 chromatograms more in the study about paprika regions.

2.4 Chemometric analysis

The resulting chromatograms were first baseline corrected by polynomial fitting and subtraction of the background and compressed using fast Fourier transform (FFT), and then submitted to linear discriminant analysis (LDA) and partial least squares regression (PLS) by means of home-made programs implemented in Matlab 7.1 (MathWorks, Natick, MA, USA) (Cetó, Céspedes, & del Valle, 2013).

Briefly, FFT was used to reduce the large dimensionality of the recorded data, while LDA was used to actually attempt its categorization based on the adulteration degree. Finally, in order to numerically quantify the degree of adulteration, PLS was employed. In both cases, the set of samples was randomly split between two subsets: training and testing, in the ratio 2:1 to ensure unbiased results were obtained from the models.

3. Results and Discussion

As discussed earlier, it is very complicated to achieve the authentication of food samples from the concentration profiles of specific compounds obtained from their targeted analysis. Oppositely, completely non-targeted analysis has also the drawback that much more features or descriptors will be required (including many that will turn to be non-relevant), thus hindering the data processing stage as well as possibly demoting model performance.

In this direction, we have developed and applied a chromatographic method for the profiling of phenolic compounds present in paprika (Figure 1), and we hypothesize on the potential of this method in combination with chemometric analysis for the authentication of paprika samples. To illustrate its potential, two different scenarios were explored taking into account paprika classification, namely the adulteration with different varieties and the adulteration with paprika from different regions (PDO's). Moreover, not only the qualitative authentication was considered, but also the quantification of the adulteration degree was attempted. The results obtained are presented over the next sections.

3.1 Authentication of paprika based on its type

The first study case was to attempt the authentication of adulterated sweet, bittersweet and spicy paprika samples from La Vera according to the levels reported in Table 1. Upon preparation of the set of adulterated samples, they were subjected to the extraction procedure and the chromatographic analysis described above, which produced a set of 144 chromatograms with characteristic fingerprints as these shown in Figure 1.

Upon measurement of all the set of samples, the next step was to attempt its discrimination with the aid of chemometric methods. However, given the large dimensionality of the recorded data, chromatograms were first compressed down to 512 coefficients with the aid of FFT algorithm. This allowed a reduction of over 95.8% on the pattern matrix without any loss of significant information and also a notorious decrease in the instrumental noise (Cetó et al., 2013).

The chosen pattern recognition method to attempt the discrimination of the adulterated samples was LDA, taking the different adulteration levels (*i.e*, the 24

mixture proportions) as the classes into which the samples were divided and the calculated Fourier coefficients as the pattern matrix. To further remove non-relevant variables with lower or none relevance to the classification task, a stepwise inclusion method was used so as to select the minimum set of coefficients to perform the prediction task with the optimum performance (Johnson & Wichein, 2007).

The two dimensional score plot obtained after LDA is shown in Figure 2. Despite its complexity given the large number of classes considered (24), there are some interesting trends that can be observed. Firstly, it is important to note how the three classes corresponding to the pure (non-adulterated) paprika samples are the ones taking the extreme values for both discriminating functions (DFs), or in other words, appear at the extremes of the plot. That is, C1 corresponding to sweet samples in the right bottom of the plot, opposite to it there is C17 corresponding to spicy samples and on the top in between those two there is C9 corresponding to bittersweet samples. More interestingly, it can also be observed how two big clusters appear distinguishing spicy adulterated samples from sweet and bittersweet adulterated ones. That is, if we imagine a line going from the top left to the bottom right, we can see how those clusters would be separated by it. Even more, we can notice how intra-clusters distance is bigger for this subset of samples compared to the other, thus indicating that adulteration of spicy paprika samples with other types of paprika is much more noticeable. This fact might be due to the much higher concentrations of capsaicinoids in spicy paprika in comparison to the other two types, which leads to a significant decrease of its concentration in the mixtures. Lastly, despite the apparent overlapping that might be seen with some classes in this 2D representation (e.g. C18/C20), this is not an issue as when also DF3 is plotted (which represents 9.48% of the total model variance), it can be seen how the clusters are clearly discriminated (with respective centroids coordinates of ca. 15 and -5,

respectively). In this regard, it has to be kept on mind that the actual model has a total of 23 DFs, which are the ones used to numerically assign the samples to each of the classes.

Next, in order to numerically assess the performance of the model, confusion matrixes were built (data not shown). Classification rate for the training and testing subsets was 100% and 81.3%, respectively; the latter being slightly lower due to some miss-classification between sweet and bittersweet adulterated samples at the lower considered levels. Moreover, performance of the model was also evaluated in terms of sensitivity, specificity and precision values (averaged for the classes) (Cetó, Voelcker, & Prieto-Simón, 2016), achieving a 81.3%, 99.2% and 85.4%, respectively.

Upon confirmation of the capability of the method to discriminate adulterated samples, and even more, to actually discriminate between different levels of adulteration, the next step was to attempt to numerically predict the actual degree of adulteration. To this aim, PLS was used instead of LDA as the modelling tool, using as before the calculated Fourier coefficients as the pattern matrix, but taking the actual percentage of adulteration rather than the classes as the target matrix.

As an example, the comparison graph of the predicted vs. expected percentage of adulteration for the mixtures of sweet and bittersweet samples is shown in Figure 3. As can be seen, a good trend is obtained, with fitted regression lines for both training and testing subsets almost indistinguishable from the ideal comparison line (y=x). That is, with slope and correlation coefficients close to 1, and intercept value close to 0; being the theoretical values included in the 95% confidence interval. In this way, confirming the potential of the approach not to only qualitatively discriminate between pure and paprika adulterated with other varieties, but also to numerically quantify the degree of adulteration.

3.2 Authentication of paprika based on its region

To further assess the suitability of the proposed method for the authentication of paprika, not only adulteration within different varieties was considered, but also the potential fraud of not respecting the PDOs. That is, mixing paprika produced in different regions. As done with the previous scenario, the same experimental design was employed, considering spicy samples from three different regions: Murcia, la Vera and Czech Republic.

As before, upon measurement of all samples, the set of 144 chromatograms was compressed employing FFT and a qualitative LDA model was built to attempt its discrimination. The resulting scores plot is shown in Figure 4. In this case a very clear trend was observed ~~in the score plot~~, with the different classes taking almost a perfect triangular shape where each vertex corresponds to the unadulterated paprika samples and the mixed samples are distributed along the faces of the triangle, being sorted according to the degree of adulteration. That is, *la Vera* samples (C1) is in the left top of the plot and opposite to *Murcia* samples (C9), which appear on the right sharing similar scores for DF2; meanwhile *Czech Republic* samples appear on the bottom, in between those two, with clear different values for DF2 evidencing that that coordinate basically discriminates Spanish and Czech samples. Very significant is also the increase in the percentage of accumulated variance only with the first two DFs, as in this case, the value goes up to ca. 91.6%. A huge value that helps to explain why such a clear trend has been obtained in the scores plot.

In order to numerically assess such a promising output, a confusion matrix was built (data not shown), from which the classification rate for the training and testing subsets was estimated as 100% and 95.8%, respectively. Performance of the model was also

evaluated in terms of sensitivity, specificity and precision values (averaged for the classes) (Cetó et al., 2016), achieving a 95.8%, 99.8% and 97.2%, respectively. All these metrics confirm what could be somehow expected from the LDA score plot, *i.e.* the fact that a clear identification is obtained for the authentication of paprika's region. In this direction, we suspect that the superior performance observed in this case as compared to the previous one might be due to the higher impact on the phenolic profile derived from the different geographical climate conditions, but also to the fact that different varieties might be cultivated over different areas, which exalts further this different profiling (Mudrić et al., 2017).

Finally, a PLS model was also built to confirm what seems to be very clear from the LDA scores plot in this case, and is the fact that the proposed chromatographic approach has huge potential to be used to numerically predict the degree of adulteration. As an example, the comparison graph of the predicted vs. expected percentage of adulteration for the mixtures of *la Vera* and *Murcia* samples is shown in Figure 5. A very good trend is obtained, with fitted regression lines for both training and testing subsets almost indistinguishable from the ideal comparison line (y=x), containing the theoretical values of slope (1) and intercept (0) in the 95% confidence interval. Thus, the proposed methodology allows both the qualitative identification and the quantitative determination of the degree of adulteration of paprika with paprika from other regions.

4. Conclusions

Based on these results, we can confirm the huge potential of chromatographic methods in combination with chemometric analysis for the authentication of paprika samples. More specifically, we can confirm the hypothesis that the broad phenolic profile of paprika is significant enough to allow the discrimination of paprika samples

given that phenolic distribution and content in natural food products seems to be related to food features such as the plant/fruit/seed variety, the geographical climate conditions of their production area, and the cultivation and manufacturing practices, among others. Consequently, they could be a rich source of analytical information to carry out the characterization, classification and authentication of food products as well as to detect possible adulterations.

Overall, this work aims to demonstrate the advantages derived from the use of chemometric methods as an alternative to specific-compound targeted classical analysis. In this way, a biomimetic approach generates an overall fingerprint of the food products analysed which allows to overcome the lack of knowledge of the compounds responsible for certain characteristics and/or perceptions.

Acknowledgments

The authors acknowledge the support of *Generalitat de Catalunya* through the projects 2017SGR-311 and 2017SGR-310. X. Cetó thanks the support of the *Beatriu de Pinós* postdoctoral programme of the Government of Catalonia's Secretariat for Universities and Research of the Ministry of Economy and Knowledge.

References

Baenas, N., Belović, M., Ilic, N., Moreno, D. A., & García-Viguera, C. (2019). Industrial use of pepper (Capsicum annum L.) derived products: Technological benefits and biological advantages. *Food Chemistry, 274*, 872-885. doi: https://doi.org/10.1016/j.foodchem.2018.09.047

Cavanna, D., Righetti, L., Elliott, C., & Suman, M. (2018). The scientific challenges in moving from targeted to non-targeted mass spectrometric methods for food fraud analysis: A proposed validation workflow to bring about a harmonized approach. *Trends in Food Science & Technology, 80*, 223-241. doi: https://doi.org/10.1016/j.tifs.2018.08.007

Cetó, X., Céspedes, F., & del Valle, M. (2013). Comparison of methods for the processing of voltammetric electronic tongues data. *Microchimica Acta, 180*(5-6), 319-330. doi: 10.1007/s00604-012-0938-7

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, 18*(12), 4479.

Cetó, X., Voelcker, N. H., & Prieto-Simón, B. (2016). Bioelectronic tongues: New trends and applications in water and food analysis. *Biosensors and Bioelectronics, 79*, 608-626. doi: http://dx.doi.org/10.1016/j.bios.2015.12.075

Publication of an application for registration pursuant to Article 6(2) of Regulation (EEC) No 2081/92 on the protection of geographical indications and designations of origin (11 February 2000).

Publication of an application pursuant to Article 6(2) of Council Regulation (EC) No 510/2006 on the protection of geographical indications and designations of origin for agricultural products and foodstuffs (24 November 2006).

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. doi: 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. doi: https://doi.org/10.1016/j.trac.2016.02.026

Dias, C., & Mendes, L. (2018). Protected designation of origin (PDO), Protected geographical indication (PGI) and traditional speciality guaranteed (TSG): A bibiliometric analysis. *Food Research International, 103*, 492-508. doi: https://doi.org/10.1016/j.foodres.2017.09.059

Esteki, M., Shahsavari, Z., & Simal-Gandara, J. (2019). Food identification by high performance liquid chromatography fingerprinting and mathematical processing. *Food Research International, 122*, 303-317. doi: https://doi.org/10.1016/j.foodres.2019.04.025

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

Johanningsmeier, S. D., Harris, G. K., & Klevorn, C. M. (2016). Metabolomic technologies for improving the quality of food: Practice and promise. *Annual Review of Food Science and Technology, 7*(1), 413-438. doi: 10.1146/annurev-food-022814-015721

Johnson, R. A., & Wichein, D. W. (2007). *Applied multivariate statistical analysis* (Vol. 6th). Harlow, GB: Pearson Education.

Kim, H.-G., Bae, J.-H., Jastrzebski, Z., Cherkas, A., Heo, B.-G., Gorinstein, S., & Ku, Y.-G. (2016). Binding, Antioxidant and anti-proliferative properties of bioactive compounds of sweet paprika (Capsicum annuum L.). *Plant Foods for Human Nutrition, 71*(2), 129-136. doi: 10.1007/s11130-016-0550-9

Kris-Etherton, P. M., Hecker, K. D., Bonanome, A., Coval, S. M., Binkoski, A. E., Hilpert, K. F., . . . Etherton, T. D. (2002). Bioactive compounds in foods: their role in the prevention of cardiovascular disease and cancer. *The American Journal of Medicine, 113*(9), 71-88. doi: 10.1016/S0002-9343(01)00995-0

Mudrić, S. Ž., Gašić, U. M., Dramićanin, A. M., Ćirić, I. Ž., 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. doi: https://doi.org/10.1016/j.foodchem.2016.09.038

Pérez-Gálvez, A., Jarén-Galán, M., Mínguez-Mosquera, M.I. (2005), Processing of red pepper fruits (*Capsicum Annuum L.*) for production of paprika and paprika oleoresin. *Handbook of Fruits and Fruit Processing*, Y. H. Hui Ed. (pp. 565-579).

Quideau, S., Deffieux, D., Douat-Casassus, C., & Pouységu, L. (2011). Plant polyphenols: Chemical properties, biological activities, and synthesis. *Angewandte Chemie International Edition, 50*(3), 586-621. doi: 10.1002/anie.201000044

Zeng, Z., Chau, F.-t., Chan, H.-y., Cheung, C.-y., Lau, T.-y., Wei, S., . . . Liang, Y. (2008). Recent advances in the compound-oriented and pattern-oriented approaches to the quality control of herbal medicines. *Chinese Medicine, 3*(1), 9. doi: 10.1186/1749-8546-3-9

Table 1. Composition of the set of samples prepared to evaluate paprika adulteration based on its type [(A) sweet, (B) bittersweet and (C) spicy] as well as on its region [(A) La Vera, (B) Murcia and (C) Czech Republic].

Class	A	B	C	Class	A	B	C	Class	A	B	C
1	1	0	0	9	0	1	0	17	0	0	1
2	0.99	0.01	0	10	0	0.99	0.01	18	0.01	0	0.99
3	0.98	0.02	0	11	0	0.98	0.02	19	0.02	0	0.98
4	0.9	0.1	0	12	0	0.9	0.1	20	0.1	0	0.9
5	0.8	0.2	0	13	0	0.8	0.2	21	0.2	0	0.8
6	0.6	0.4	0	14	0	0.6	0.4	22	0.4	0	0.6
7	0.4	0.60	0	15	0	0.4	0.60	23	0.60	0	0.4
8	0.2	0.8	0	16	0	0.2	0.8	24	0.8	0	0.2

FIGURE CAPTIONS

Figure 1. Representative raw chromatograms obtained for spicy paprika samples extracts of (top to bottom) *La Vera*, *Murcia* and *Czech Republic* under the conditions described in section 2.3.

Figure 2. Score plot obtained after LDA analysis for the authentication of paprika's type. In this study 144 chromatograms were analysed, corresponding to 24 different proportions of the sweet, bittersweet and spicy types of paprika from La Vera (two samples for every proportion and three injections per sample). Numbers indicate the class of every sample (i.e., the proportion of paprika types) according to Table 1. Coloured filled symbols correspond to the training subset and black empty ones to the testing subset, whereas the centroid for each of the classes is also plotted (★).

Figure 3. Performance of the optimized FFT-PLS model for the authentication of paprika's type. For every sample, the predicted *versus* expected percentage of adulteration of sweet paprika from La Vera with the bittersweet variety of the same PDO is shown, including training (●, solid line) and testing (○, dotted line) subsets. The dashed line corresponds to the theoretical diagonal line.

Figure 4. Score plot obtained after LDA analysis for the authentication of paprika's region. In this study 144 chromatograms were analysed, corresponding to 24 different proportions of the spicy type of paprika from La Vera, Murcia and Czech Republic (two samples for every proportion and three injections per sample). Numbers indicate the class of every sample (i.e., the proportion of paprika types) according to Table 1.

Coloured filled symbols correspond to the training subset and black empty ones to the testing subset, whereas the centroid for each of the classes is also plotted (★).

Figure 5. Performance of the optimized FFT-PLS model for the authentication of paprika's region. For every sample, the predicted *versus* expected percentage of adulteration of spicy paprika from La Vera with the same variety from Murcia is shown, including training (●, solid line) and testing (○, dotted line) subsets. The dashed line corresponds to the theoretical diagonal line.

Figure 1

Fig1

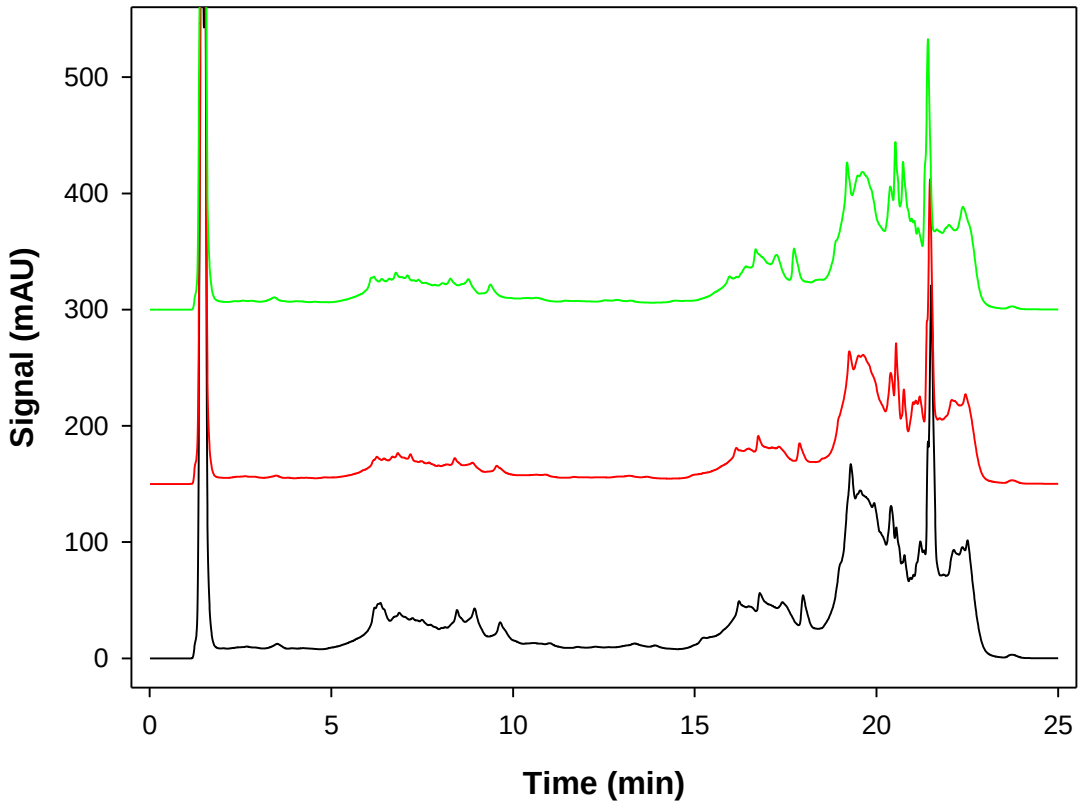


Figure 2

Fig2

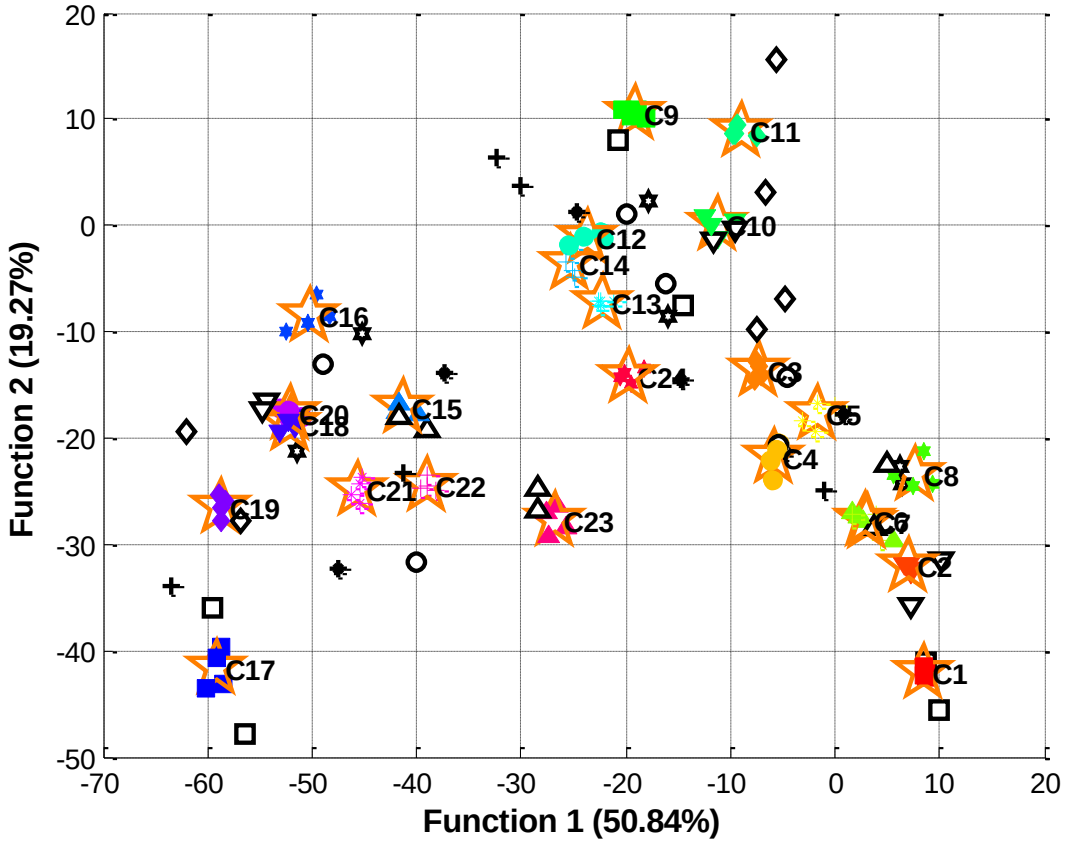


Figure 3

Fig3

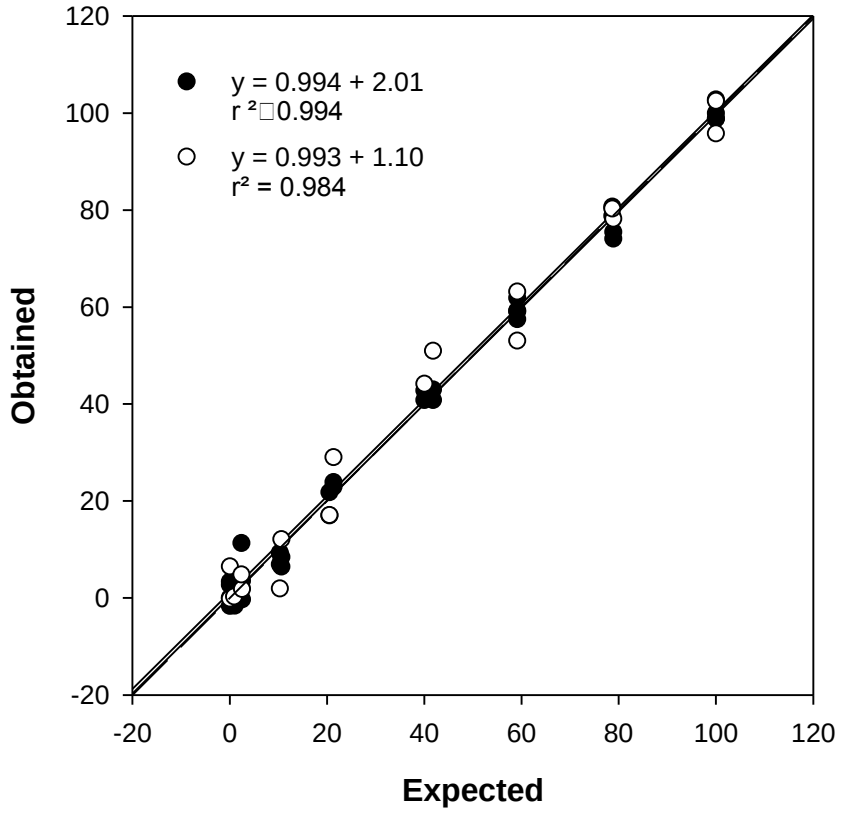


Figure 4

Fig4

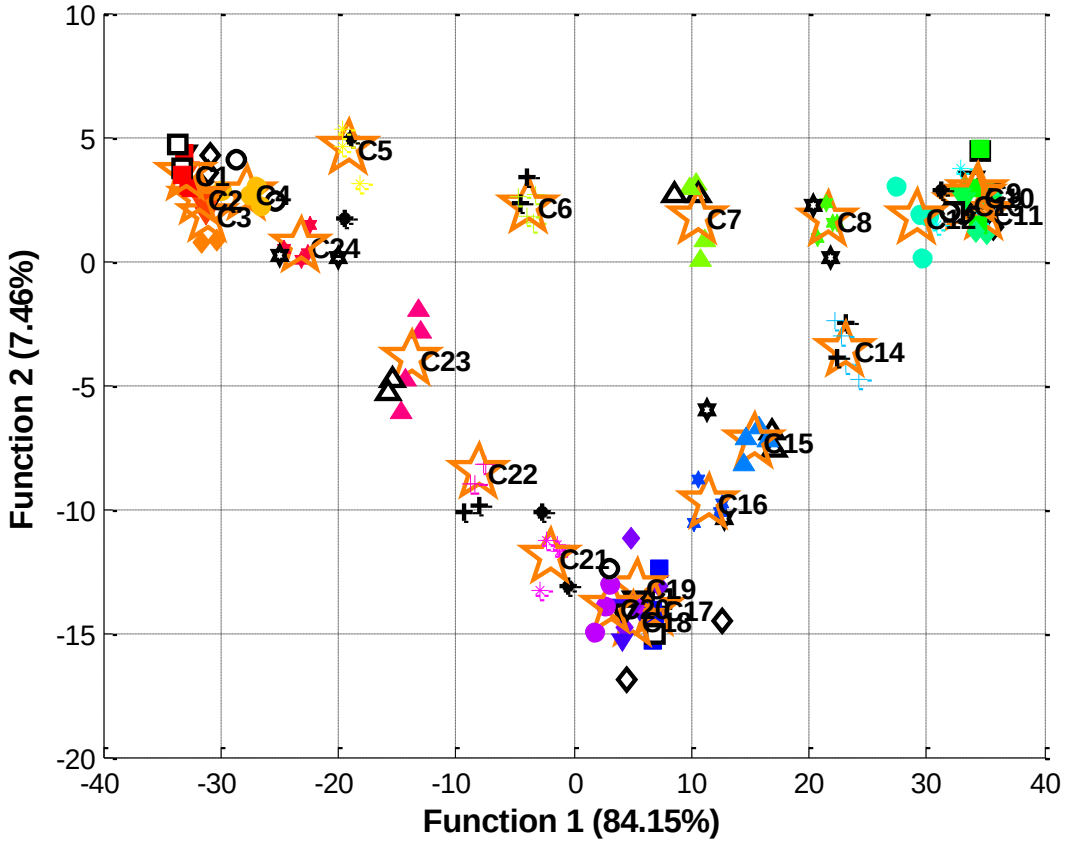
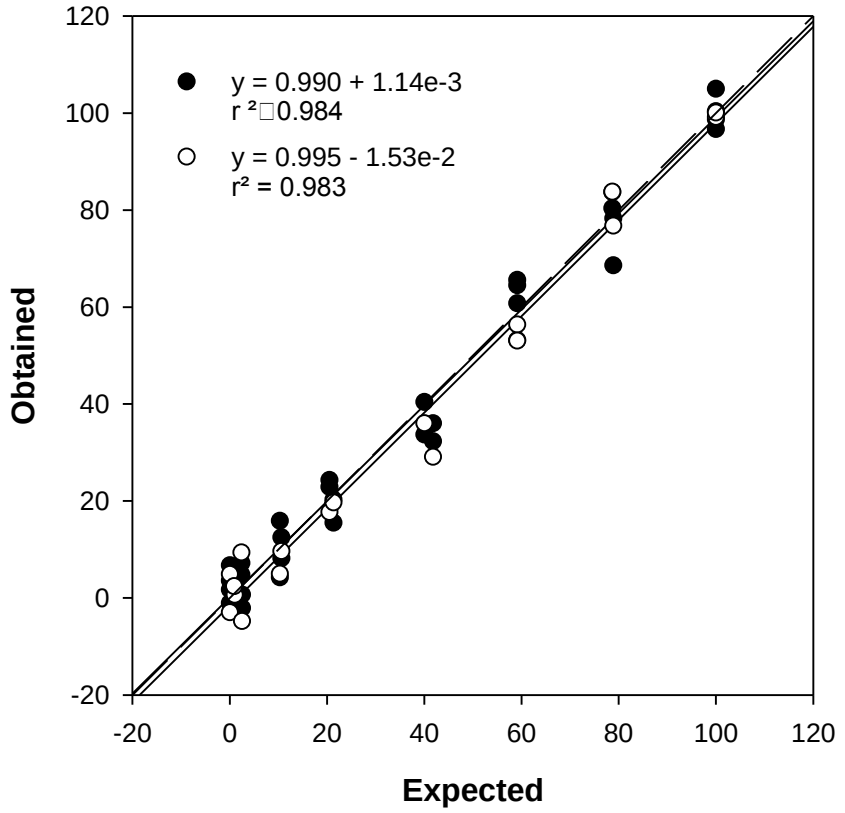


Figure 5

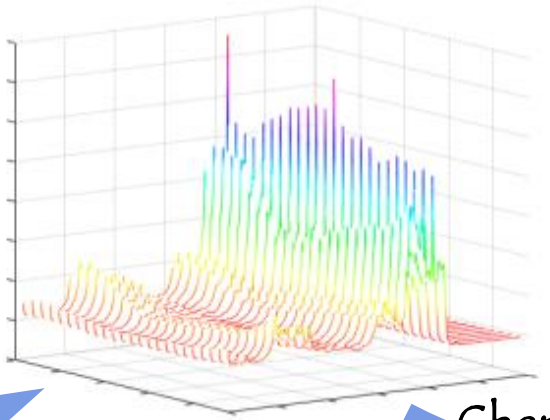
Fig5



Graphical Abstract

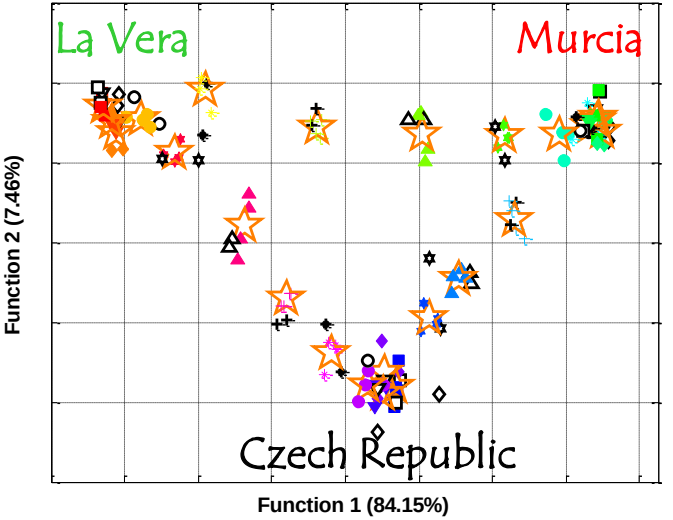
Sample extraction

HPLC analysis



Phenolics fingerprint

Chemometric analysis



Conflicts of interest

Declarations of interest: none

Authors contribution section

This work was originally designed by Núria Serrano, Oscar Núñez and José Manuel Díaz-Cruz. Cristina Sánchez carried out the experimental work by liquid chromatography. Xavier Cetó contributed to the design of experiments and carried out the data treatment. He also wrote the initial version of the manuscript that was later improved with contributions by Núria Serrano, Oscar Núñez and José Manuel Díaz-Cruz.