UNIVERSITAT DE BARCELONA

Application of metabolomic profiling and fingerprinting approaches to food fraud cases

Guillem Campmajó Galván

Guillem Campmajó Galván

UNIVERSITAT DE BARCELONA

PhD THESIS

Application of metabolomic profiling and fingerprinting approaches to food fraud cases

2023

Guillem Campmajó Galván

UNIVERSITAT DE BARCELONA

FACULTAT DE QUÍMICA

DEPARTAMENT D'ENGINYERIA QUÍMICA I QUÍMICA ANALÍTICA

Programa de doctorat:

Química Analítica i Medi Ambient

Application of metabolomic profiling and fingerprinting approaches to food fraud cases

---

Guillem Campmajó Galván

Memòria presentada per tal d'optar al títol de Doctor per la Universitat de Barcelona

Sota la direcció de

Dr. Oscar Núñez Burcio

Departament d'Enginyeria Química i Química Analítica

Universitat de Barcelona

Barcelona, maig del 2023

El Dr. Oscar Núñez Burcio, professor catedràtic del Departament d'Enginyeria Química i Química Analítica de la Universitat de Barcelona,

FA CONSTAR,

que la present memòria titulada "Application of metabolomic profiling and fingerprinting approaches to food fraud cases" ha estat realitzada sota la meva direcció pel Sr. Guillem Campmajó Galván, en el Departament d'Enginyeria Química i Química Analítica, i que tots els resultats presentats són fruit del treball experimental realitzat per l'esmentat doctorand.

I perquè així consti, expedeixo i signo el present certificat.

Barcelona, maig del 2023

Dr. Oscar Núñez Burcio

Dedicat als meus avis Fulgenci i Pepe,

i a les meves àvies Carmen i Mª Sol

En primer lloc, vull agrair al meu director, el Dr. Oscar Núñez, la seva supervisió, seguiment i tutoria en el desenvolupament d'aquesta tesi doctoral. En especial, vull agrair-li la confiança que ha dipositat en mi i les oportunitats que m'ha brindat.

El meu agraïment al Dr. Javier Saurina per la seva ajuda, orientació i disponibilitat quan ha estat necessària. També, als professors del grup de recerca Fluorescència —Dra. Sònia Sentellas, Dra. Mercè Granados, Dr. José Francisco García, i Dr. José Luis Beltran— per la bona acollida i la seva proximitat, així com a la resta de treballadors i professorat de la Secció de Química Analítica.

I would also like to thank Dr. Alina Astefanei for giving me the opportunity of working in her laboratory and for her support during my stay in Amsterdam.

També, vull fer especial menció dels estudiants de TFG i TFM amb els que he tingut el plaer de treballar —Gemma, Laura, Ruben, Judith, Luis i Xavi—, i dels investigadors estrangers amb qui he pogut col·laborar i aprendre —Alessandra, Braulio i Carolina—.

La meva gratitud a tots els companys i amics de laboratori i departament —Aina, Alejandro, Ane, Clara, Claudia, Hiba, Irene, Juanfra, Laura, Lídia, María, Montse, Nerea, Nil, Oumaima, Paulina, Raquel, Rocío, Roger, Sara, i Sergio— per les experiències dins i fora de la universitat. També a tots els companys del CSIC que he tingut el plaer de conèixer en aquest període.

Vull fer una menció especial a la Noemí, l'Albert C., l'Albert V., l'Òscar, la Laura i l'Adrià, per haver compartit aquesta etapa junts i per la seva amistat. També a l'Andrea, el Miguel i la Francesca per fer de l'estada a Amsterdam una experiència inoblidable. De la mateixa manera, vull agrair als meus amics dels Jesuïtes de Casp per la seva estima, així com el seu fidel suport.

Finalment, vull agrair a la meva família, i en concret als meus pares i a la meva germana, per la confiança i el recolzament incondicional al llarg de tots aquests anys. Sense ells, realitzar aquesta tesi no hauria estat possible.

CONTENTS

ABSTRACT

Food fraud is an intentional and misleading act in food that generally does not comply with food law and is motivated by economic gain. It encompasses several fraudulent practices such as deception during manufacture, diversion into illicit supply chains, interventions with the food product, or misrepresentation. In this context, the coming to light of the horse meat scandal at the beginning of 2013 highlighted the shortcomings of the European system against food fraud, increasing concern and interest among European citizens and administrative bodies.

Under these circumstances, in recent years, omics tools —comprising genomics, transcriptomics, proteomics, metabolomics, and elementomics/isotopollomics— have been applied to solve food fraud issues, along with biostatistics and chemometrics. In most cases, their application has relied on profiling (focusing on determining targeted secondary chemical markers) or fingerprinting approaches (based on the unspecific detection of instrumental responses without assuming any previous knowledge about the sample composition), overcoming the traditional targeted analysis. In particular, since a food product's metabolome varies according to its biological nature and several external conditions (*i.e.*, either from a natural or anthropogenic origin), metabolomics has shown excellent potential to assess several issues related to its authenticity and quality.

Therefore, in this thesis, several metabolomic profiling and fingerprinting approaches were developed to address different food fraud cases. In this line, liquid chromatography coupled to low- or high-resolution mass spectrometry (LC–LRMS, LC–HRMS) was proposed for the targeted approaches. In contrast, non-targeted methods were based on liquid chromatography with ultraviolet detection (LC-UV) or fluorescence detection (LC-FLD), LC–HRMS, or direct mass spectrometry (MS)-based techniques. Furthermore, non-supervised and supervised chemometric techniques allowed sample assignation and classification. As a result, the proposed analytical methodologies were successfully applied to several food products

—including paprika, nuts and seeds, hen eggs, vegetable oils, and red wine— guaranteeing their classification and authentication regarding the geographical origin, botanical origin, production system, or quality category.

RESUM

El frau alimentari és un acte intencionat i enganyós produït en els aliments que, generalment, no compleix amb la legislació alimentària i que està motivat per un benefici econòmic. Comprèn diverses pràctiques fraudulentes, com ara l'engany durant la producció, la desviació a cadenes de subministrament il·lícites, les manipulacions del producte alimentari o la tergiversació. En aquest context, la sortida a la llum de l'escàndol de la carn de cavall a principis del 2013 va posar de manifest les mancances del sistema europeu contra el frau alimentari, augmentant la preocupació i l'interès entre els ciutadans i els organismes administratius europeus.

En aquestes circumstàncies, en els darrers anys, s'han aplicat eines òmiques —que inclouen la genòmica, la transcriptòmica, la proteòmica, la metabolòmica i l'elementòmica/isotopol·lòmica— per resoldre qüestions relacionades amb el frau alimentari, juntament amb bioestadística i quimiometria. En la majoria dels casos, la seva aplicació s'ha efectuat mitjançant estratègies basades en perfils (centrant-se en la determinació dirigida de marcadors químics secundaris) o empremtes dactilars (basades en la detecció inespecífica de respostes instrumental sense assumir cap coneixement previ sobre la composició de la mostra), superant l'anàlisi dirigida tradicional. En concret, com que el metaboloma d'un producte alimentari varia segons la seva naturalesa biològica i un seguit de condicions externes (siguin d'origen natural o antropogènic), la metabolòmica ha demostrat un excel·lent potencial per avaluar diverses qüestions relacionades amb la seva autenticitat i qualitat.

Per tant, en aquesta tesi, es van desenvolupar diverses estratègies de perfils i empremtes dactilars metabolòmiques per abordar alguns casos de frau alimentari. En aquesta línia, es va proposar la cromatografia líquida acoblada a l'espectrometria de masses de baixa o alta resolució (LC–LRMS, LC–HRMS) per als enfocaments dirigits. En canvi, els mètodes no dirigits es van basar en la cromatografia líquida amb detecció ultraviolada (LC-UV) o fluorescent (LC-FLD), LC–HRMS o tècniques basades en l'espectrometria de masses (MS) directa. A més, tècniques

quimiomètriques no supervisades i supervisades van permetre l'assignació i classificació de les mostres. Com a resultat, les metodologies analítiques proposades es van aplicar amb èxit a diferents productes alimentaris —incloent el pebre vermell, fruits secs i llavors, ous de gallina, olis vegetals i vi negre— garantint-ne la classificació i autenticació pel que fa a l'origen geogràfic, l'origen botànic, el sistema de producció o la categoria de qualitat.

ABBREVIATIONS AND ACRONYMS

AAC	Administrative Assistance and Cooperation
AAS	Atomic absorption spectroscopy
ACN	Alert and Cooperation Network
ADIS	Animal Diseases Information System
AES	Atomic emission spectroscopy
AESAN	Spanish Agency for Food Safety and Nutrition
AIF	All-ion fragmentation
AIMS	Ambient ionisation mass spectrometry
ANN	Artificial neural network
ANOVA-PCA	Analysis of variance-principal component analysis
APCI	Atmospheric pressure chemical ionisation
ATR	Attenuated total reflectance
CE	Capillary electrophoresis
CEN	European Committee for Standardisation
CID	Collision-induced dissociation
COI	Cytochrome c oxidase 1
CV	Cross-validation
CZE	Capillary zone electrophoresis
d	Reduced distance
DA	Discriminant analysis
DAD	Diode array detector
DDA	Data-dependent acquisition
ddPCR	Digital droplet polymerase chain reaction
DG	Directorate General
DG AGRI	Directorate General for Agriculture and Rural Development
DG MARE	Directorate General for Maritime Affairs and Fisheries
DG SANTE	Directorate General for Health and Food Safety
DIA	Data-independent acquisition
DIHRMS	Direct infusion high-resolution mass spectrometry
DIMS	Direct infusion mass spectrometry
DIR	Directive
DMS−MS	Differential mobility spectrometry coupled to mass spectrometry
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
EC	European Commission
ECN	Equivalent carbon number
ED-XRF	Energy dispersive-X ray fluorescence
EEA	European Economic Area
EFSA	European Food Safety Authority
EI	Electron ionisation

ELISA	Enzyme-linked immunosorbent assay
ESI	Electrospray ionisation
ETD	Electron transfer dissociation
EU	European Union
EUROJUST	European Union Agency for Criminal Justice Cooperation
EUROPHYT	European Union Notification System for Plant Health Interceptions
EUROPOL	European Union Agency for Law Enforcement Cooperation
EVOO	Extra virgin olive oil
FA	Fatty acid
FAES	Flame atomic emission spectroscopy
FAME	Fatty acid methyl ester
FAO	Food and Agriculture Organization
FBO	Food business operator
FFN	Food Fraud Network
FIA−HRMS	Flow injection analysis coupled to high-resolution mass spectrometry
FIA−MS	Flow injection analysis coupled to mass spectrometry
FIR	Far-infrared
FT	Fourier transformation
GC	Gas chromatography
GC-FID	Gas chromatography with flame ionisation detector
GC−HRMS	Gas chromatography coupled to high-resolution mass spectrometry
GC−MS	Gas chromatography coupled to mass spectrometry
GI	Geographical indication
GMO	Genetically modified organism
HACCP	Hazard Analysis and Critical Control Point
HCA	Hierarchical cluster analysis
HCD	High-energy collision dissociation
H-ESI	Heated electrospray ionisation
HILIC	Hydrophilic interaction chromatography
^{1}H-NMR	Proton nuclear magnetic resonance
HPLC	High-performance liquid chromatography
HRM	High-resolution melting
HRMS	High-resolution mass spectrometry
kNN	k-nearest neighbour
ICP−MS	Inductively coupled plasma mass spectrometry
ICP-OES	Inductively coupled plasma atomic emission spectroscopy
iLOD	Instrumental limit of detection
iLOQ	Instrumental limit of quantification
IMS	Ion mobility spectrometry
IMSOC	Information management system for official controls
IOC	International Olive Council
IR	Infrared

IRMS	Isotope ratio mass spectrometry
IT	Ion trap
JRC	Joint Research Centre
KC-FFQ	Knowledge Centre for Food Fraud and Quality
LAMP	Loop-mediated isothermal amplification
LFD	Lateral flow devices
LC	Liquid chromatography
LC-CAD	Liquid chromatography with charged aerosol detection
LC-ECD	Liquid chromatography with electrochemical detection
LC-FLD	Liquid chromatography with fluorescent detection
LC−HRMS	Liquid chromatography coupled to high-resolution mass spectrometry
LC−LRMS	Liquid chromatography coupled to low-resolution mass spectrometry
LC−MS	Liquid chromatography coupled to mass spectrometry
LC−MS/MS	Liquid chromatography coupled to tandem mass spectrometry
LC-RID	Liquid chromatography with a refractive index detector
LC-UV	Liquid chromatography with ultraviolet detection
LDA	Linear discriminant analysis
LIT	Linear ion trap
LLE	Liquid-liquid extraction
LOD	Limit of detection
LOQ	Limit of quantification
LRMS	Low-resolution mass spectrometry
LTQ-Orbitrap	Linear trap quadrupole-Orbitrap
LV	Latent variable
MALDI−MS	Matrix-assisted laser desorption ionisation−mass spectrometry
MANOVA-PCA	Multivariate analysis of variance-principal component analysis
matK	Maturase K
MC-ICP−MS	Multi-collector-inductively coupled plasma mass spectrometry
MIR	Mid-infrared
MLR	Multiple linear regression
MRM	Multiple-reaction monitoring
MRMS	Magnetic resonance mass spectrometer
MS	Mass spectrometry
MUFA	Monounsaturated fatty acids
NGS	Next generation sequencing
NIR	Near-infrared
NIST	National Institute of Standards and Technology
NMR	Nuclear magnetic resonance
OD	Orthogonal distance
OIV	International Organisation of Vine and Wine
OLAF	European Anti-Fraud Office

OO	Olive oil
OPLS-DA	Orthogonal partial least squares regression-discriminant analysis
PC	Principal component
PCA	Principal component analysis
PCR	Polymerase chain reaction
PCR-LFD	Polymerase chain reaction-lateral flow device
PDO	Protected designation of origin
PFF	Peptide fragmentation fingerprinting
PFP	Pentafluorophenyl
PGI	Protected geographical indication
PLS	Partial least squares
PLS-DA	Partial least squares regression-discriminant analysis
PMF	Peptide mass fingerprinting
PTM	Post-translational modification
PUFA	Polyunsaturated fatty acids
Q-Orbitrap	Quadrupole-Orbitrap
qPCR	Quantitative polymerase chain reaction
QqQ	Triple quadrupole
Q-TOF	Quadrupole-time-of-flight
RASFF	Rapid Alert System Feed and Food
rbcL	Carboxylase
REG	Regulation
RF	Radio frequency
RMSEC	Root-mean-square error of calibration
RMSECV	Root-mean-square error of cross-validation
RMSEP	Root-mean-square error of prediction
ROC	Receiver operating characteristic
R^2	Determination coefficient
SD	Score distance
SHU	Scoville heat units
SIMCA	Soft independent modelling of class analogies
SLE	Solid-liquid extraction
SPE	Solid-phase extraction
SSAFE	Safe Supply of Affordable Food Everywhere
SVM	Support vector machine
TAG	Triacylglycerol
TLC	Thin-layer chromatography
TOF	Time-of-flight
TRACES	Trade Control and Expert System
TSG	Traditional specialities guaranteed
TWIMS	Travelling-wave ion mobility spectrometry
UK	United Kingdom

USA	United States of America
UV	Ultraviolet
UV-Vis	Ultraviolet-visible
VACCP	Vulnerability Assessment and Critical Control Point
VOC	Volatile organic compound
VOO	Virgin olive oil

OBJECTIVES AND STRUCTURE

This thesis' main goal is the development of metabolomic profiling and fingerprinting approaches, consisting of chromatographic- and mass spectrometric-based methods combined with chemometrics, to solve different food fraud cases. Thus, the subgoals set in order to attain the primary objective are detailed below:

- Development of targeted liquid chromatography coupled to mass spectrometry (LC–MS) for the determination of phenolic compounds and capsaicinoids and carotenoids in paprika. Then, evaluation of these compounds as chemical markers for classifying paprika samples according to their geographical origin.
- Application of liquid chromatography with spectroscopic detection fingerprinting —liquid chromatography with ultraviolet detection (LC-UV) or fluorescence detection (LC-FLD)— to address different food fraud issues regarding nuts, hen eggs, and paprika samples.
- Use of non-targeted liquid chromatography coupled to high-resolution mass spectrometry (LC–HRMS) to address the classification of nut samples according to their type and to tentatively identify type-related chemical markers. Then, evaluation of the found chemical markers in a case study of adulterating almond-based products with hazelnut or peanut.
- Assess several food authentication issues through direct mass spectrometry (MS)-based fingerprinting methods, based on flow injection analysis coupled to high-resolution mass spectrometry (FIA–HRMS) and differential mobility spectrometry coupled to mass spectrometry (DMS–MS). Application to red wine, paprika, and vegetable oil samples.
- Evaluation of different chemometric tools —including the principal component analysis (PCA), partial least squares regression-discriminant analysis (PLS-DA), soft independent modelling of class analogies (SIMCA), and partial least squares (PLS) regression— depending on the aim of each food fraud study.

Thereby, this thesis is presented as a compendium of scientific publications and is structured in four chapters.

Chapter 1 is divided into two main concepts: food fraud and foodomics. On the one hand, it introduces food fraud by providing historical background, defining its most relevant related terms, detailing the current European Union (EU) legislation and structure to combat it, and describing some current tools for its vulnerability assessment. Moreover, common threats and fraudulent practices carried out on the food products analysed in this thesis (*i.e.*, paprika, hen eggs, olive oil, wine, and nuts and seeds) are also reported. On the other hand, and from a foodomics perspective, Chapter 1 introduces targeted and non-targeted approaches, describes the application of omics tools —genomics and transcriptomics, proteomics, metabolomics, and elementomics and isotopollomics— in food authentication, and summarises the most used chemometric techniques.

Chapter 2 focuses on applying metabolomic profiling approaches, combined with chemometrics, in food authentication. Hence, this chapter includes an introduction, which distinguishes between targeted and suspect profiling approaches and describes the main metabolites used for food authenticity purposes, the results section, and a discussion of the obtained results. In this context, the results shown in this chapter correspond to the next scientific articles:

> **Publication I** – Barbosa, S.; **Campmajó, G.**; Saurina, J.; Puignou, L.; Núñez, O. 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. 2020, 68, 591.
>
> https://dx.doi.org/10.1021/acs.jafc.9b06054
>
> **Publication II** – Arrizabalaga-Larrañaga, A.; **Campmajó, G.**; Saurina, J.; Núñez, O.; Santos, F. J.; Moyano, E. Determination of capsaicinoids and carotenoids for the

characterization and geographical origin authentication of paprika by UHPLC–APCI–HRMS. LWT. 2021, 139, 110533.

https://doi.org/10.1016/j.lwt.2020.110533

In contrast, Chapter 3 evaluates metabolomic fingerprinting approaches based on chromatographic and mass spectrometric methods to authenticate food products using multivariate analysis. The chapter is structured in three sections: the introduction, the results, and the discussion. In this case, the introduction contains the next scientific book chapter and is complemented with further information:

Publication III – **Campmajó, G.**; Núñez, O. Chromatographic fingerprinting approaches in food authentication. In G. Campmajó & O. Núñez (Eds.), Chromatographic and related separation techniques in food integrity and authenticity. Volume A: Advances in chromatographic techniques. World Scientific Publishing. 2021, 137 – 165.

https://doi.org/10.1142/9781786349958_0006

Furthermore, the results contained in Chapter 3 correspond to the following scientific articles:

Publication IV – **Campmajó, G.**; Navarro, G.J.; Núñez, N.; Puignou, L.; Saurina, J.; Núñez, O. Non-targeted HPLC-UV fingerprinting as chemical descriptors for the classification and authentication of nuts by multivariate chemometric methods. Sensors. 2019, 19, 1388.

https://doi.org/10.3390/s19061388

Publication V – **Campmajó, G.**; Cayero, L.; Saurina, J.; Núñez, O. Authentication of hen eggs by HPLC-UV fingerprinting and chemometric methods. Foods. 2019, 8, 310.

https://doi.org/10.3390/foods8080310

Publication VI – **Campmajó, G.**; Saez-Vigo, R.; Saurina, J.; Núñez, O. 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. 2020, 114, 107265.

https://doi.org/10.1016/j.foodcont.2020.107265

Publication VII – **Campmajó, G.**; Rodríguez-Javier, L. R.; Saurina, J.; Núñez, O. Assessment of paprika geographical origin fraud by high-performance liquid chromatography with fluorescence detection (HPLC-FLD) fingerprinting. Food Chemistry. 2021, 352, 129397.

https://doi.org/10.1016/j.foodchem.2021.129397

Publication VIII – **Campmajó, G.**; Saurina, J.; Núñez, O. Liquid chromatography coupled to high-resolution mass spectrometry for nut classification and marker identification. Food Control. 2023, 152, 109834.

https://doi.org/10.1016/j.foodcont.2023.109834

Publication IX – **Campmajó, G.**; Saurina, J.; Núñez, O. FIA−HRMS fingerprinting subjected to chemometrics as a valuable tool to address food classification and authentication: application to red wine, paprika, and vegetable oil samples. Food Chemistry. 2022, 373, 131491.

https://doi.org/10.1016/j.foodchem.2021.131491

Publication X – **Campmajó, G.**; Saurina, J.; Núñez, O.; Sentellas, S. Differential mobility spectrometry coupled to mass spectrometry (DMS−MS) for the classification of Spanish PDO paprika. Food Chemistry. 2022, 390, 133141.

https://doi.org/10.1016/j.foodchem.2022.133141

Finally, this PhD thesis ends with its general conclusions that are summarised in Chapter 4.

CHAPTER 1

INTRODUCTION

1.1. FOOD FRAUD

1.1.1. FROM ANCIENT TIMES TO NOWADAYS

Food fraud goes back to ancient times, arguably near the origin of trading, which was initially based on barter. At that time, most of the products offered in the markets were raw, and mainly processed foods (*e.g.*, ales, honey, wine, and oils) were vulnerable to deceptive practices. Nevertheless, raw products could also be altered by fraudulent actions like overweighting [1]. Furthermore, as the complexity of goods and the food chain increased over time, fraudsters gained more knowledge, which led to subtler and more sophisticated frauds. For instance, as the spices trade grew in the Medieval Period, being highly valued and expensive products, their adulteration with local dried herbs did too. Subsequently, the Industrial Revolution caused a massive migration from rural communities into urban areas in the most developed countries. Thus, urbanisation distanced consumers from primary food production, raising the opportunities for food fraud [2].

In 1820, Friedrich Accum published the *Treatise on Adulterations of Food and Culinary Poisons* [3], containing the first comprehensive study of food fraud. It comprised several adulteration cases that could impact consumers' health and provided some methodologies for their detection (*i.e.*, microscopy, physical property measurements, or wet chemistry). Moreover, the combination of different simultaneous factors —growing concern about food quality, progress in health and scientific knowledge, and food fraud scandals causing thousands of deaths— led to several food regulatory laws set up during the 19th and early 20th centuries. In this regard, the United Kingdom (UK) Public Health Act in 1875 [4] (consolidated later through the Food and Drugs Adulteration Act in 1928) [5], the French Food Law in 1905 [6], and the United States of America (USA) Federal Meat Inspection Act [7] and Pure Food and Drug Act [8] in 1906 established a benchmark for public administrations. In addition, throughout the 20th century, the development of both

food fraud legislation and analytical techniques used to detect it (gaining sensitivity and specificity) continued [9,10].

However, the progress and application of strict food regulations, supported by more sophisticated analytical techniques, have still not impeded fraudsters from committing fraud without being detected. Several food fraud scandals during the last decades have highlighted the shortcomings of the worldwide detection system. Besides, some indirectly became a food safety issue because they threatened human health. For instance, in 1981, olive oil was adulterated with rapeseed oil denatured with aniline, which had an industrial use, causing more than 1,000 deaths in Spain. More recently, in 2008, milk powder was adulterated with melamine in China. Melamine is a toxic organic compound, and its addition aimed to increase milk powder nitrogen content, shaming a more extensive protein abundance. As a consequence, six babies died, and almost 300,000 were severely intoxicated. Finally, beef products were found to contain horse meat in 2012 illegally. Although this fraud was originally detected in the UK, further research indicated its spread to other European countries. This latter case proved the danger of large-scale fraud due to a globalised supply system with cut-price agri-food productivism and an incomplete labelling system [11,12]. Besides, despite not entailing a public health threat, it questioned the effectiveness of the existing controls at national and European levels.

Furthermore, besides food safety, consumers' interest in food quality —encompassing attributes such as the presence of a specific ingredient, the production system (*e.g.*, organic products), or the region of origin— has increased. These attributes are often related to higher prices, making these products susceptible to fraud. In this context, although this kind of fraudulent practice does not generally pose a threat to human health, it misleads both consumers and official authorities.

In summary, nowadays, it is easy to conduct food fraud without being detected because of factors such as the complexity of the food chain (many players are involved between production and consumption), the massive diversity of products,

and the lack of analytical solutions. Indeed, the actual extent of fraudulent practices is unknown, and only some are discovered. Combating food fraud is, therefore, a complex challenge that requires an international effort from a multidisciplinary perspective.

1.1.2. DEFINITIONS AND RELATED TERMS

Up to now, the term '*food fraud*' still misses an international harmonised legal definition. As a result, researchers and regulatory bodies have diversely interpreted it alongside its associated terminology. However, several common aspects can be pointed out from these food fraud definitions. In this context, it is an intentional and misleading act in food (by modification of the food itself or related documentation), which generally does not comply with food law and is motivated by economic gain [13]. Besides, it is considered that food fraud becomes food crime when it evolves from random individual acts to organised activities perpetrated on a large scale [14].

Similarly, it has also been inconsistently categorised into different types. The categorisation of the food fraud incidents proposed by Manning *et al*. [15] is taken as a reference in this thesis with some modifications. In this line, Table 1.1 summarises the types and sub-types of food fraud. Briefly, '*deception during manufacture*' includes two sub-types: overtreating and overrun. On the one hand, the undeclared or incorrectly labelled addition of water in frozen seafood products —either through phosphate or non-phosphate treatment or glazing— is an example of an overtreating practice. In this case, fraudsters aim to manipulate the product price by increasing its weight [16]. On the other hand, overrun includes fraudulent acts, such as intentionally underreporting the total amount of production, which usually result in subsequent '*diversion into illicit supply chains*' activities. This latter food fraud type does not necessarily imply a food law violation (the foodstuff might comply with the safety requirements of the food legislation) but may infringe upon other regulations. Foodstuff smuggling to evade specific tax payments is an example of these practices.

Instead, the food fraud type named '*interventions with the food product*' encompasses practices such as adding substances to accentuate specific organoleptic properties (*e.g.*, adding Sudan dyes to enhance chilli powder colour) or adulterating by adding lower quality and cheaper products to the authentic one (*e.g.*, mixing extra-virgin olive oil with other vegetable oils). Finally, this classification's last food fraud type is '*misrepresentation*', which mainly includes deceptions by presence or omission in the product label. For instance, it comprises acts such as non-declaration of allergens or banned substances and mislabelling of the product's geographical origin or production system.

Table 1.1. Types and sub-types of food fraud. Adapted from Manning *et al*. classification [15].

Type	**Sub-type**
Deception during manufacture	Overtreating
	Overrun
Diversion into illicit supply chains	Diversion
	Smuggling
	Theft
Interventions with the food product	Addition
	Adulteration
	Substitution
	Product tampering
	Unapproved processes
Misrepresentation	Misdescription
	Record tampering
	Misrepresentation of food characteristics, country of origin, food ingredients, or food packaging
	Claim violation
	False or misleading statements

One of the difficulties of classifying food fraud practices is that, in most cases, a given fraudulent practice can be categorised into multiple types. In this context, most lead to misrepresentation or mislabelling in the final product. For instance, while the fraudulent adulteration of honey with cheaper sweeteners (*i.e.*, corn sugar or rice

syrups) is performed by omission in the product label, a false statement is carried out when substituting a seafood species for a cheaper one. Furthermore, '*duplication*' frauds —simulation and counterfeiting—, described in the original classification of Manning *et al*. [15], have not been considered in this thesis since they can also be classified as '*interventions with the food product*' or '*misrepresentation*'.

From a broader point of view, the concept of food integrity goes beyond food fraud. According to the European Union (EU) Food Integrity Project [17], it is '*the state of being whole, entire, or undiminished or in perfect condition*'. Different complementary categorisations of this term have been proposed. On the one hand, Robson *et al*. [13] defined it as an overarching term in the so-called food protection risk matrix [18]. Hence, as shown in Figure 1.1, food integrity is composed of four key elements —food quality, food safety, food fraud, and food defence— whose risks differ in both their intentionality of the cause and final motivation.

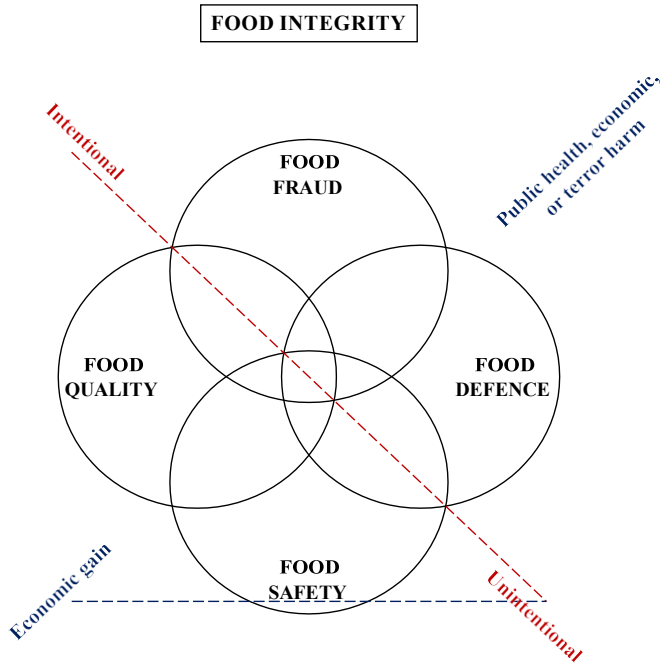


Figure 1.1. Food protection risk matrix.

In this context, food quality includes negative and positive attributes influencing foodstuff value and consumer acceptance. Thus, unlike food fraud, its risk emerges from an unintentional event, such as a mishandling in the food supply chain, that alters the product reducing the expected characteristics and properties. Therefore, even with the reduced brand equity, if not detected, a food quality incident does not impede food business operators (FBOs) from obtaining an economic gain unwittingly. Furthermore, it is noteworthy to mention that although the motivation neither of food quality nor food fraud is to threaten human health, on some occasions, it can be an effect. Instead, food safety and defence risks give rise to potential harm, which is generally related to public health. In this line, while food safety risk derivates unintentionally (*e.g.*, due to contamination along the production step), a food defence case is born from the will to produce damage.

On the other hand, Manning [19] broke food integrity down into the items that can be affected by fraud along the supply chain:

- Product integrity refers to the inherent and intrinsic characteristics of the foodstuff.
- Process integrity encompasses extrinsic characteristics (such as the production method, organic or vegan label, or country of origin), often highly valued and related to the food production step.
- People integrity relies on the honesty of each actor involved in the whole food chain not committing fraud.
- Data integrity mainly applies to the logistics and qualitative information that concerns a specific foodstuff.

Another relevant term commonly employed in food fraud is '*food authenticity*'. Authenticity is the quality of being genuine, ergo authentic. In this line, according to the European Committee for Standardisation (CEN), food product authenticity is correlated to the '*match between the actual characteristic of the food product and the claim made about it*' [20]. Therefore, it is directly linked to product and process

integrity since it involves the foodstuff's intrinsic and extrinsic properties. Furthermore, Popping *et al*. [21] recently proposed using the term '*inauthentic*' rather than '*food fraud*' since fraud is a legal term and, therefore, requires a successful prosecution in a court of law. Nevertheless, '*food fraud*' is preferred in this thesis due to its widespread use in the field.

Moreover, data integrity is strongly associated with traceability, which was defined as '*the ability to trace and follow a food, feed, food-producing animal or substance intended to be, or expected to be incorporated into a food or feed, through all stages of production, processing and distribution*' in the European General Food Law [22].

1.1.3. EUROPEAN UNION FOOD LEGISLATION

As established by the European Union (EU) Treaties, the European legislature —composed of the European Parliament, the Council of the EU, and the European Commission (EC)— is responsible for legislation on certain issues such as food. Thus, while the European Parliament acts on behalf of the citizens and the Council of the EU does it on the Member States, the EC is the day-to-day European administration. In this context, legislation can be mainly enacted as a Regulation (REG) or a Directive (DIR). The first must be directly applied across the EU (affecting both people and businesses), whereas the last fixes a specific objective that the Member States must achieve by harmonising their laws [23].

Regarding food law, although food fraud practices are illegal, most do not threaten public health, and the EU authorities did not prioritise them. Thus, regulatory agencies initially focused on food safety, ergo the unintentional presence of residues or contaminants in food or feed implying a potential danger to human health. In this context, the White Paper on Food Safety [24], presented by the EC in 2000, reflected the EU's will to assure the highest food safety standards through proper legislation and organisation. As a result, in 2002, REG (EC) Nº 178/2002 [22], known as the General Food Law Regulation, was subsequently implemented and approved by the

European Council and the European Parliament. Establishing the general requirements and procedures to assess food and feed safety in the whole food chain and the creation of the European Food Safety Authority (EFSA) and the Rapid Alert System Feed and Food (RASFF) were among its scope. Moreover, several guidelines and obligations mentioned in it were further developed through complementary independent regulations.

For instance, on the one hand, the General Food Law's Article 17 makes responsible FBOs to ensure that food complies with food law at all stages of the food chain and the Member States to check it. In this line, in 2004, REG (EC) Nº 882/2004 [25] and REG (EC) Nº 854/2004 [26] were established. The first dealt with the official controls performed to verify compliance with feed and food law, animal health, and animal welfare rules. Instead, REG (EC) Nº 854/2004 detailed the specific rules for the organisation of official controls on products of animal origin intended for human consumption. Therefore, they did not only provide a legal framework for inspection agencies to assess FBOs' law compliance but also the corresponding measures to be taken in the opposite case.

Additionally, rules regarding the use of food additives in food products were laid down in REG (EC) Nº 1333/2008 [27]. Food additives encompass any substance not consumed as food nor used as a characteristic ingredient. Only those substances that do not threaten consumers' health, meet a technological need (*i.e.*, food preservation, product supplementation for consumers with special dietary needs, or organoleptic properties enhancement), and do not mislead consumers, are included in the so-called positive lists. Therefore, while additives in the lists are authorised under the specified conditions (*i.e.*, maximum level), the not included ones are prohibited.

On the other hand, REG (EC) Nº 1169/2011 [28] provided the bases to ensure a high level of consumer protection regarding food information (food labelling), as requested in Article 16 of the General Food Law, which specifies that the labelling of food shall not mislead consumers. This regulation, which repealed the DIR

2000/13/EC [29], contains the mandatory, prohibited, and voluntary information that a food product can be labelled with, among other instructions. Hence, for instance, the indication of up to 12 particulars (*e.g.*, the product's name, the list of ingredients, the presence of allergens, or the date of durability) is mandatory in all food products, except for tiny labels. Instead, messages misleading the consumers or making health claims are forbidden. In this line, only nutrition or health claims following REG (EC) Nº 1924/2006 [30] can be voluntarily labelled in food.

Nevertheless, despite the development and implementation of the above-mentioned food laws (mainly focusing on the food safety field), food fraud was not explicitly addressed in any of them. Indeed, only the general stipulation that consumers may not be misled, contained in REG (EC) Nº 178/2002 [22], indirectly assessed this issue. Under these circumstances, the coming to light of the horse meat scandal at the beginning of 2013 was an inflexion point. In response to the food crisis, the European Parliament Committee on the Environment, Public Health and Food Safety drew up a report [12] emphasising the needs of the European food fraud system and making recommendations to improve it. In brief, the report's main request to the CE was to pay full attention to combat food fraud becoming of crucial priority. As a result, the EU policies concerning food fraud changed, and some of the suggestions were established. For instance, because of the call to implement a system to collect and exchange data related to fraud cases, the Agri-Food Fraud Network (FFN) and the Administrative Assistance and Cooperation (AAC) system were set up in 2013 and 2015, respectively.

Furthermore, this new approach towards food fraud also ended up in law modifications. Hence, some of the requests made in the report were established in 2019 through the REG (EU) Nº 2017/625 [31], repealing the previous legislation in the field: REG (EC) Nº 882/2004 [25] and REG (EC) Nº 854/2004 [26]. Thus, the in-force regulation aimed to harmonise the existing regulatory framework (official controls and other activities) of the entire agri-food chain, focusing not only on food safety issues but also on preventing fraud. This purpose was reflected in Article 9,

which makes the competent authorities responsible for regularly performing official controls to identify intentional fraudulent or deceptive practices in the areas included in the regulation. In this regard, Article 1 of the regulation widened its field of application by including the following sectors: plant health, trade and use of protected plant products, organic production, and use and labelling of protected designation of origin (PDO), protected geographical indication (PGI) or traditional specialities guaranteed (TSG) products. Furthermore, Article 9 contained other report recommendations, such as performing the official controls based on a risk-based approach and without prior notice (unannounced inspections).

Moreover, Article 131 of the regulation on official controls described the CE's duty to establish the information management system for official controls (IMSOC) in collaboration with the Member States. In this context, Commission implementing REG (EU) 2019/1715 [32] laid down the rules for its set-up and management. Further details about its operation are given in Section 1.1.4.

Besides the previously described general food laws, specific legislation concerning food labelling has also been established. In the last decades and due to greater access to the food supply, European consumers have become increasingly interested in products with attributes socially associated with high-quality standards (*i.e.*, geographical origin, production or processing system, or product variety), aside from their nutritional value. In this context, the EU has developed different regulations that establish the legal requisites to ensure a specific standard and properly label it. In this line, REG (EC) Nº 1151/2012 [33], repealing REG (EC) Nº 509/2006 [34] and REG (EC) Nº 510/2006 [35], covers the quality schemes to identify agricultural products and foodstuffs with value-added characteristics or attributes. Both PDO and PGI distinctions (see Figure 1.2 to observe the corresponding logos) link the products with a specific geographical area, recognising their unique and distinctive characteristics. Thus, the PDO label identifies a product originated in a specific region, whose environment (natural and human factors) influences its quality and where the whole production chain is carried out. Instead, although the PGI label demands similar

guidelines, it allows more flexibility since it only requires one production step in the defined geographical area. Currently, there are 1861 PDO and 1372 PGI registered products [36]. It is noteworthy to mention that this regulation does not apply to the geographical indication (GI) of spirit drinks or grapevine products, which follows REG (EU) 2019/787 [37]. However, according to Article 16 of the regulation, they can use the PGI symbol. Besides, REG (EC) Nº 1151/2012 did neither apply to aromatised wines, but this was recently amended to simplify the legal framework due to the limited number of registrations [38].

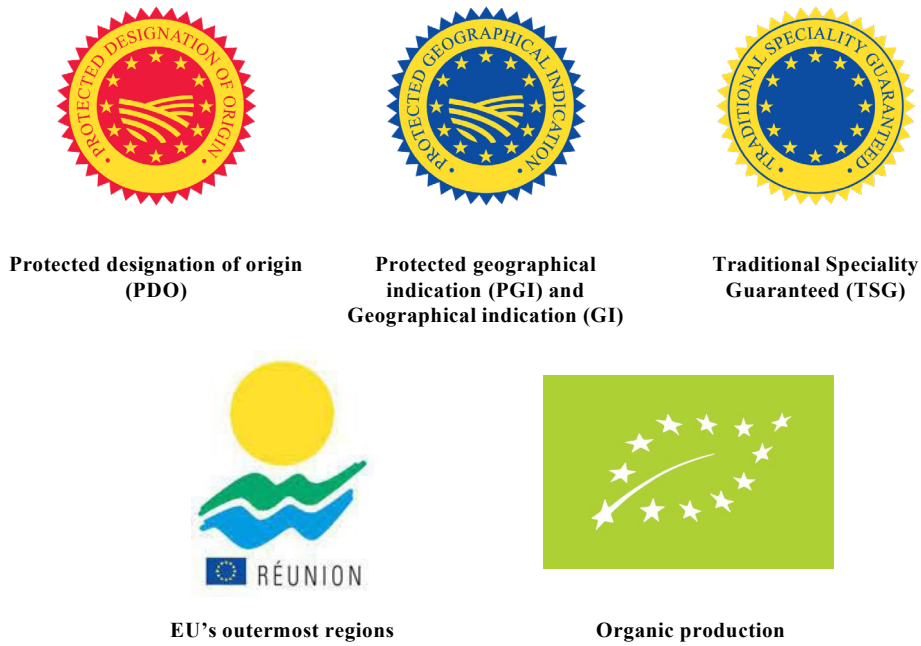


Figure 1.2. EU geographical origin and production method quality certifications.

Furthermore, REG (EC) Nº 1151/2012 describes the scheme for TSG, whose logo can be observed in Figure 1.2. In this case, a TSG foodstuff shall result from a traditional practice (either in the production, processing, or composition) with specific ingredients. Up to now, 69 products are registered as TSG, such as the Spanish *Panellets* and *Jamón Serrano*, or the Italian *Amatriciana Tradizionale* and *Pizza Napoletana* [39]. In addition, this regulation also includes two other optional

quality terms: the mountain product and the product of island farming labels. In both cases, raw materials and feedstuff must come from a mountain area or island as appropriate, and in processed products, processing must also occur in such locations.

An additional EU law related to the product's geographical origin is REG (EU) Nº 228/2013 [40], which specifies agricultural measures for the EU's outermost regions: French overseas departments (France), the Azores and Madeira (Portugal) and the Canary Islands (Spain). Hence, among other objectives, this regulation aims to strengthen the competitiveness of crops and agricultural products from these areas. In this line, Figure 1.2 shows the logo used to enhance consumer awareness and consumption.

Instead, regarding the food production and processing system, consumers' interest in organic farming products has exponentially risen in the last few years. This fact has been mainly caused due to increasing environmental consciousness within society and the belief that organic foodstuffs are healthier than conventional ones. In this regard, REG (EC) Nº 834/2007 [41] ruled in the EU in matters of organic production and the labelling of its products until 2022, when REG (EU) 2018/848 [42] repealed it. Briefly, the new regulation comprised a scope extension to cover other products linked to agriculture (*e.g.*, essential oils, cotton, or sea salt), a trade agreement with recognised third countries having equivalent organic standards, and the simplification for small farmers to join the organic scheme.

Moreover, according to the regulation, some of the main principles applied to agricultural and aquaculture activities of organic production and the processing of organic food are summed up next:

- Organic production shall base on practices respecting nature's system and cycles, preserving natural landscapes, and sustainably exploiting energy and natural resources (*i.e.*, water, soil, organic matter, and air) and, therefore, protecting the environment. For instance, using non-renewable resources is limited, while recycling waste and by-products is desired for further use.

- It shall lead to producing a wide variety of high-quality goods, preferably locally produced and distributed through short channels.
- High-level of biodiversity is pursued through methods employing living organisms and mechanical production techniques, excluding genetically modified organisms (GMOs), and based on risk assessment approaches.
- Choosing appropriate plant varieties and animal breeds is required for a high degree of genetic diversity, with disease resistance, longevity, and adaptation to local conditions. Besides, plant agronomic performance and animal breeding value shall also be considered.
- A high level of animal health and welfare is essential. Livestock must be fed with organic feed, and animal husbandry practices enhancing their immune system must be applied.
- Finally, processed organic food shall originate from organic agricultural ingredients without adding food additives or micronutrients. Indeed, the integrity of organic production shall be ensured at any stage of the food chain.

Regarding their presentation and labelling, organic products shall contain the corresponding logo (see Figure 1.2) and be labelled as detailed in Article 30. For instance, although only the term '*organic*' may be used to label them in English, other languages accept different terms too (*e.g.*, Spanish includes terms such as '*ecológico*' and '*biológico*').

In summary, despite the above-mentioned initial prioritisation of food safety issues and the lack of specific and consistent nomenclature for food fraud, the EU has made recent efforts to include food fraud in the scope of its legislation. Thus, no single law can holistically address its broad scope, but several laws indirectly do it. The following Section presents the current EU organisation to fight against it.

1.1.4. CURRENT EUROPEAN UNION STRUCTURE REGARDING FOOD FRAUD

As mentioned in the previous Section, the General Food Law founded the EFSA in 2002 [22]. The EFSA is an independent agency of the EU whose mission focuses on providing scientific advice on issues related to the food chain (mainly regarding food safety), communicating to the common public its outputs, and collecting data to identify and monitor current and emerging risks. Moreover, it cooperates closely with Member States and institutional partners, such as the Directorate General for Health and Food Safety (DG SANTE) or the Joint Research Centre (JRC) [43].

At the European level, a total of 31 departments, known as Directorate Generals (DGs), constitute the EC [44]. Among them, the DG SANTE has a crucial role in health —protecting human, animal, and plant health and animal welfare— and food areas —promoting a high level of food safety and sustainability along the supply chain—. Indeed, regarding the latter case, the DG SANTE is currently contributing to the '*European Green Deal*', one of the general objectives set by the CE's President, Ursula von der Leyen. Hence, within this general objective, three specific objectives are detailed: 1) the food and feed safety, 2) the sustainable food systems – the '*Farm to Fork*' strategy, and 3) the international promotion of the EU food safety standards. Among them, some refer to food fraud [45].

For instance, within the first specific objective, DG SANTE works on properly implementing REG (EU) Nº 2017/625 on official controls [31] and the management and maintenance of the IMSOC. In this regard, the IMSOC was established in 2021 in the application of the Commission implementing REG (EU) 2019/1715. Briefly, it consists of four EU information systems: the Animal Diseases Information System (ADIS), the European Union notification system for plant health interceptions (EUROPHYT), the Trade Control and Expert System (TRACES), and the iRASFF platform [32]. Specifically, the iRASFF platform allows information exchange between the Alert and Cooperation Network (ACN) members regarding non-

compliances with the EU legislation of cross-border nature through the food chain. This information is obtained from the networks that compose the ACN [46]:

- The RASFF allows exchanging information (RASFF notifications) concerning food safety. Hence, only direct or indirect risks to human health due to food or feed products are reported.
- Instead, detected cross-border EU food legislation' violations, without posing *a priori* a health risk, are released by non-compliance notifications in the AAC.
- Finally, the FFN deals with potential food fraud cases. Therefore, food non-compliances suspected to be intentional for an economic purpose are confidentially shared (fraud notifications) within it.

ACN members comprise the EU Member States, the European Economic Area (EEA) countries, the EFSA, Switzerland, and the EC. Besides, in the case of FFN notifications, only contact points designated previously by each member can have access. For instance, the Spanish contact point is the Spanish Agency for Food Safety and Nutrition (AESAN), through the sub-directorate general of alerts and official control programming [47]. In addition, as depicted in Figure 1.3, cooperation between the ACN (especially by the FFN), the police and customs agencies —the European Union Agency for Law Enforcement Cooperation (EUROPOL) and the European Anti-Fraud Office (OLAF), respectively—, and the European Union Agency for Criminal Justice Cooperation (EUROJUST) is carried out when needed.

Instead, one of the main goals of the '*Farm to Fork*' strategy is fighting against food fraud. Thereby, DG SANTE conducts different studies (*e.g.*, the launch of a coordinated control plan on the authenticity of herbs and spices) [48] to check the effective implementation of Member States' national measures. Finally, the specific objective of the international promotion of EU food safety standards is based on initiatives to improve bilateral trade relations with non-European countries.

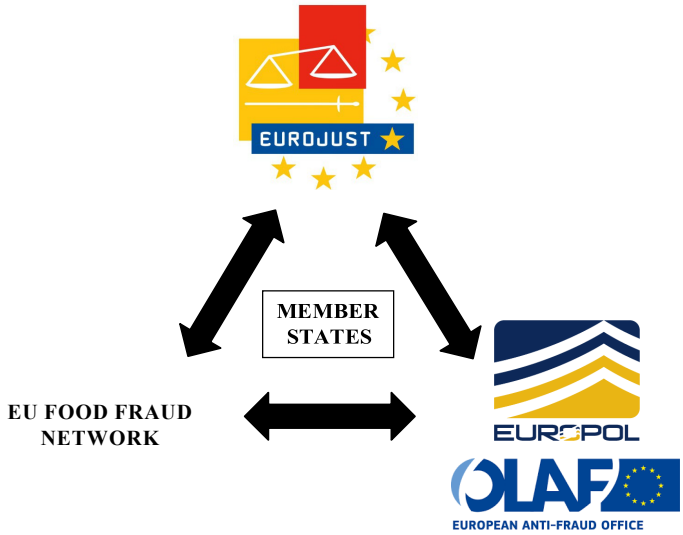


Figure 1.3. EU cooperative approach to combat food fraud.

Nevertheless, the DG SANTE is not the only DG actively preventing and detecting food fraud. On the one hand, the Knowledge Centre for Food Fraud and Quality (KC-FFQ) unit, created within the JRC, shares updated scientific publications about food fraud and quality, creates food fraud databases (*e.g.*, European Wine DataBank or the Oleum DataBank), or develops harmonised methodologies for assessing food quality [49]. On the other hand, the Directorate General for Agriculture and Rural Development (DG AGRI) or the Directorate General for Maritime Affairs and Fisheries (DG MARE) are also involved in the marketing regulations and the development of analytical methodologies to check compliance of agricultural and seafood products, respectively [21].

1.1.5. FRAUD VULNERABILITY ASSESSMENT

As the General Food Law describes, FBOs are responsible for ensuring food safety and protecting consumers from fraudulent practices [22]. Thus, they must assess and monitor both by developing and applying effective internal systems.

As previously discussed, food safety threatens public health due to unintentional risks derived from the presence of known contaminants or pathogens (*e.g.*, pesticides, veterinary drug residues, and mycotoxins). Food safety events are inevitable and frequently occur because of their involuntary nature. Thus, these issues have traditionally been handled following internal process controls based on mitigation strategies (risk assessment), trying to reduce the negative consequences of particular emergencies. In this context, predictive risk-based assessment tools —being risk the probability of loss or injury from a hazard [50]— share common steps such as risk identification, risk analysis, and risk evaluation [15]. For instance, the Hazard Analysis and Critical Control Point (HACCP) plan, which is one of the most extended systems, is structured in the following seven principles: risk identification is performed by 1) conducting hazard analysis, risk analysis by 2) determining the critical control points, and finally, risk evaluation by establishing 3) the critical limits, 4) monitoring procedures, 5) corrective actions, 6) verification procedures, and 7) record-keeping and documentation procedures [51].

Otherwise, food fraud originates from fraudsters' intentionality to obtain economic gain. In this case, the primary strategy followed by FBOs to address fraud is based on preventing the event from occurring rather than detecting it once it has happened. Therefore, food fraud assessment focuses on reducing its vulnerabilities, which are gaps, weaknesses, or flaws of the food system that create opportunities for fraudsters [52]. These vulnerabilities can be intrinsic —at the individual or organisational levels— or extrinsic —related to the suppliers and customers, the wider food chain network, or the international environment— to the food business. Furthermore, as reported by van Ruth *et al*. [53], they can be defined by three key elements that enhance and favour fraud likelihood: opportunities, motivations, and control measures.

First, opportunities can be divided into technical opportunities and opportunities in time and place. The former encompasses factors such as the ease of adulterating specific food products or the absence of detection methods. Instead, the latter is

strongly related to the length and complexity of the food supply chain since the more stages (*i.e.*, production, processing, storage, distribution, retailing, administration requirements, and consumption) and actors involved in it, the greater fraud vulnerability. Second, both economic and cultural or behavioural motivations can increase fraud vulnerability. For instance, economic motivations comprise low availability of expensive goods, products with value-added attributes (*e.g.*, PDO, PGI, or TSG), or business economic health, among others. In contrast, factors such as business strategy, ethical culture, or corruption are related to cultural motivations. Finally, control measures are capable of counteracting and decreasing the vulnerability arising from opportunities and motivations when properly executed, or on the contrary, intensifying it due to lack or deficiencies. In this context, fraud detection or prevention can be carried out by technical (*i.e.*, development of effective analytical methods and application of traceability tools) or managerial measures (*i.e.*, ethical codes of conduct, integrity screening, or whistleblowing systems), respectively [53].

Unlike food safety, where HACCP measures have been internationally recognised through their incorporation into Codex Alimentarius standards, no food fraud prevention tool has been validated [54]. Nevertheless, the Vulnerability Assessment and Critical Control Point (VACCP), which follows the principles of HACCP, the Safe Supply of Affordable Food Everywhere (SSAFE), or other in-house food fraud vulnerability assessments, are currently the preferred models to address fraud in food businesses [55].

1.1.6. FOOD MATRICES UNDER STUDY

Food production and distribution offer possibilities for evasion of taxes or other duties through fraudulent practices mentioned in Section 1.1.2, such as diversion or smuggling. However, this thesis evaluates fraud directly aimed at the buyer's level. In this context, food fraud threatens various products, either from plant or animal

origin, and can usually affect their composition, cultivar or variety, species, geographical origin, production system, or processing [56].

As a case of study, frauds related to different food matrices —including paprika, hen eggs, olive oil, wine, and nuts and seeds— are herein addressed. In this line, the current context of each analysed good is detailed below.

1.1.6.1. Paprika

Spices —encompassing products such as paprika, saffron, pepper, cinnamon, and turmeric— are well-known for their organoleptic properties, which make them good seasonings for culinary purposes. They are also used as the main ingredient of different food supplements or over-the-counter pharmaceuticals due to their high content of bioactive compounds with beneficial effects on human health (*e.g.*, anti-inflammatory or antimicrobial activities) [57]. Besides, other uses include cosmetics, promoting these products as '*natural*'.

The plant species used for producing spices cannot be properly cultivated in the European area due to climate conditions, except for dried paprika and chilli. In this line, paprika is a valued red powder spice obtained from drying and grinding red pepper fruits of the *Capsicum annuum* species (*Capsicum* L. genus and Solanaceae family) [58]. In Europe, up to seven paprika products are distinguished with the PDO label: *Pimentón de la Vera* (La Vera, Extremadura, Spain), *Pimentón de Murcia* (Murcia, Spain), *Pebre bord de Mallorca* (Mallorca, Spain), *Kalocsai fűszerpaprika-őrlemény* (Kalocsa, Hungary), *Szegedi fűszerpaprika-őrlemény* (Szeged, Hungary), *Žitavská paprika* (Žitavská, Slovak Republic), and *Piment d'Espelette – Ezpeletako Biperra* (Espelette, France) [36]. Nevertheless, Asia predominantly leads the spice market, while the EU needs to import them to supply its population, running a sizeable trade deficit for these products [59]. For instance, 94,000 tons of paprika and allspice were imported to the EU in 2019, 69% from China [48].

In this context, paprika (like other spices) is at high risk of adulteration because of different factors or vulnerabilities that increase fraud opportunities. For example, its supply chain is long and complex, requiring several intermediates' participation and passing through different countries [60]. Additional difficulties in keeping paprika free from fraud comprise its presentation form, which makes easier adulteration practices from being detected, and its high economic value [61].

Typical forms of adulteration that have been detected in spices are the substitution for foreign matter with physical similarities (*i.e.*, granulometry or colour) or inferior production-own materials, the addition of synthetic dyes to reach a fresh appearance, and the geographical origin mislabelling [62]. In this line, due to a coordinated control plan on the authenticity of herbs and spices launched by the DG SANTE, almost 6% of the paprika samples tested were suspicious of adulteration [48]. Among these samples, 63% contained non-declared constituents such as ash (with a content above 10%, which is the maximum level established), maize, carrot, tomato, sunflower seed, onion, or garlic. Instead, the remaining 37% contained dyes (*e.g.*, Sudan dyes) whose use is not authorised in food [27]. However, other deception practices related to the misdescription of the geographical origin or production system were not evaluated. Therefore, analytical methods to detect paprika geographical origin fraud are developed in this thesis.

1.1.6.2. Hen eggs

Hen egg is consumed worldwide due to its high nutritional value, high culinary potential (*i.e.*, it is a coagulant, foaming, and emulsifying), and affordable price. Indeed, it is considered one of the cheapest foods of animal origin whose intake can provide the essential amino acids, lipids, vitamins, and minerals, while offering a moderate calorie source [63]. From 2010 to 2020, Asia was the leading hen egg producer yielding 61.4%, followed by America and Europe, with 19.7% and 13.5%, respectively [64].

Although hen eggs can be subject to different kinds of fraud —such as adding specific dyes or adulterating with melamine—, their freshness and correct identification category have become complex and critical issues to be solved. On the one hand, egg freshness is strongly related to its quality since it affects the properties mentioned above while increasing the risk of developing undesired microorganisms or degradation products [65]. In this regard, the European legislation sets the following deadlines related to egg ageing: the terms '*extra*' or '*extra fresh*' can only be used until the ninth day after laying, eggs must reach the consumer within 21 days after being laid, and finally, their minimum durability shall be fixed at most up to the twenty-eighth day after laying [66,67]. Therefore, for instance, falsification of the expiry date, aiming to length the product's commercial life, can be carried out by fraudsters. In order to detect age-related fraudulent practices in hen eggs, REG (EC) Nº 853/2004 establishes 3-hydroxybutyric and lactic acids as chemical markers, since their high content indicates severe product deterioration [67].

On the other hand, different European directives and regulations limit the legal framework for producing, trading, and selling hen eggs. In this line, the DIR 1999/74/EC [68] established the minimum standards that the Member States shall ensure to protect the laying hens, distinguishing three types of rearing systems: the unenriched cage, the enriched cage, and the alternative systems. Besides, the same directive had foreseen the prohibition of the unenriched cage system (often called battery cages), where laying hens have at least 550 cm^2 of cage area per hen, for animal welfare reasons by 2012, as it was done [69]. Furthermore, the DIR 2002/4/EC [70] demanded the Member States to properly register the establishments keeping laying hens, while REG (EC) Nº 589/2008 [66] detailed the rules for egg marketing standards. Thereby, hen eggs must contain the so-called producer code, which consists of a distinguishing number composed of a digit indicating the farming method, followed by the Member State's code (*e.g.*, ES for Spain, FR for France, or IT for Italy), and an identification number defined by the Member State to trace the producer.

According to the regulation, class A eggs are destined for human consumption and must comply with several quality features. Among them, four types of eggs can be found in the European market concerning the farming method. Their corresponding digit and some of their specific characteristics are summarised below:

- 3 – Eggs from cage hens. In this case, hens are bred in cages of at least 750 cm^2 per hen. Each cage has a nest, litter that allows pecking and scratching, and perches.
- 2 – Eggs from barn hens. In this farming system, hens can move freely around the farm building, which consists of multiple aviaries where drinking and feeding facilities are equally distributed. Besides, a maximum allowed stocking density of 9 hens per m^2 (at least 1110 cm^2 per hen) is established.
- 1 – Eggs from free range hens. Barn system requirements must be fulfilled and complemented with continuous daytime access to open-air runs, mainly covered by vegetation. In this line, hen density in open-air runs must not exceed 1 hen per 4 m^2.
- 0 – Organic eggs. Hens producing these eggs are free range but follow organic production principles in terms of the origin of the animals, nutrition, animal welfare, and veterinary treatment, among others [42]. Moreover, the maximum indoor stocking density is 6 hens per m^2, while the same conditions as the free range farming system are permitted outdoors.

In 2019, 49.5% of the European hen eggs were from cages, 32.5% from barns, 11.8% from free range hens, and 6.2% from organic production [71,72]. However, in response to the European citizens' initiative '*End the cage age*', which requests a transition to more ethical and sustainable farming systems, the EC is planning to prohibit the use of the cage system (possibly by 2027) [73].

Furthermore, there is an increase in the hen egg's price from type 3 to 0 due to the differences in the stocking density, feed costs, and productivity [74]. This fact makes eggs of higher category susceptible to mislabelling with ones of lower category.

Currently, this fraudulent practice can only be detected by competent authorities through the administrative traceability of the product [75]. For this reason, in this thesis, the authentication of the hen egg category is addressed with a particular focus on identifying eggs obtained through organic production.

1.1.6.3. Olive oil

Virgin olive oil is mechanically and physically produced from the fruits of the olive tree (*Olea europaea* L.). Its appreciated quality, sensorial attributes, nutritional properties, and health benefits make it the most characteristic and representative product of the Mediterranean area. In this line, according to the Food and Agriculture Organization (FAO) of the United Nations, Spain, Italy, and Greece constituted the largest olive oil producers from 2010 to 2020 [64]. Furthermore, these countries possess 29, 42, and 20 olive oil products registered as PDO, respectively [36].

It is known that parameters such as the olive cultivar, the environmental conditions, the agronomic practice, the harvest season, the olive maturation, the storage, and the technological processes strongly affect olive oil quality [76]. In this context, REG (EU) Nº 1308/2013, in agreement with the International Olive Council (IOC), classifies olive oils in accordance to their quality category [77]:

- As previously mentioned, '*Virgin olive oil*' comprises oils obtained from *Olea europaea* L. fruits only through mechanical or physical means. They are classified as extra virgin olive oil (EVOO), virgin olive oil (VOO), or lampante olive oil depending on their free acidity value in terms of oleic acid ($\leq 0.8\%$, $\leq 2.0\%$, and $> 2.0\%$, respectively).
- '*Refined olive oil*' consists of oils obtained by refining virgin olive oil.
- '*Olive oil*' (OO) encompasses olive oils resulting from blending refined and virgin olive oils (except for lampante oil).

Besides, the regulation's classification includes oils obtained from olive-pomace: crude olive-pomace oil, refined olive-pomace oil, and olive-pomace oil.

In this context, olive oil is one of the most regulated food commodities in the world. For instance, in the EU, Commission REG (EU) 2022/2104 and Commission implementing REG (EU) 2022/2105 [78,79], which recently repealed the REG (EEC) Nº 2568/91 [80], detail what parameters have to be checked to evaluate olive oil quality and purity, which are their established limits, and which analytical method have to be applied to determine them. Briefly, testing laboratories verify olive oil's category by following a decision tree (each rule node consists of a parameter) until one final decision is reached. First, rule node parameters are based on quality criteria —*i.e.*, acidity, peroxide value, ultraviolet (UV) spectroscopy, or organoleptic assessment— and then on purity criteria —*i.e.*, stigmastadienes, *trans*-isomers of fatty acids, fatty acid composition, equivalent carbon number (ECN) 42, sterol composition and total content, erythrodiol and uvaol, waxes, and 2-glyceryl-monopalmitate—.

However, despite being highly regulated, the olive oil sector is still a target for fraudulent practices due to the high-profit margin. Thus, according to the 2021 ACN annual report [46], the '*fats and oil*' product category was the third with the most fraud notifications in 2021. Remarkably, most remain related to OO or VOO sold as EVOO. In this context, the misdescription of the olive oil category and the addition of lower quality oils from the same species (refined or olive-pomace oils) or from different species (other vegetable oils) are practices commonly carried out [81]. Furthermore, undeclared pigments, such as copper complexes of chlorophylls or carotenoids, whose use is banned in this product, may also be added to mask tampering [82]. In addition, as reported by Conte *et al.* [83], current regulatory methods do not cover the misdescription of the cultivar or the geographical origin nor the detection of selected blends of virgin olive oil with lower-quality vegetable oils (including soft deodorised olive oils).

In addition, it is noteworthy that different global and local projects have been recently launched to combat olive oil fraud. On the one hand, at the European level, the Oleum project (from 2016 to 2020) aimed to assure olive oil quality and authenticity by 1) developing new analytical methodologies, 2) establishing the Oleum Databank, and 3) promoting a worldwide community of the sector stakeholders [84]. On the other hand, at the Catalan level, the *Federació de Cooperatives Agràries de Catalunya* led the Autenfood project between 2018 and 2020. Similarly, it aimed to find new analytical methods capable of detecting olive oil fraud [85].

Therefore, under these circumstances, the present thesis proposes a high-throughput analytical method to address olive oil authenticity in terms of botanical origin and category.

1.1.6.4. Wine

Wine is an alcoholic beverage obtained from the fermentation of grapes, usually from one or more varieties of the *Vitis vinifera* species. It is internationally consumed due to its enjoyable flavour, its historic cultural relevance (*i.e.*, it promotes conviviality and social engagement), and its link to the Mediterranean diet, which includes its moderate consumption (especially red wine) and is associated with heart disease prevention [86]. In this line, the USA, France, and Italy led the consumption ranking in 2020, according to the International Organisation of Vine and Wine (OIV) [87]. Instead, regarding its production, almost 27 million tonnes of wine were produced worldwide that year, with Europe being the leading producer. Indeed, Italy, France, and Spain have been the top three wine producers in the last decade [64].

Wine quality is strongly influenced by diverse parameters such as the type of grape varieties, viticultural practices, winemaking techniques, ageing conditions, and *terroir*. In this last case, the French term t*erroir* refers to the combination of environmental factors (*i.e.*, geographical, climatic, and pedological conditions) affecting the grape's phenotype [88]. Therefore, due to its significant connection with

the origin area, wine is one of the agricultural products most regulated by the EU through REG (EC) Nº 1151/2012 [33]. Hence, for instance, 1184 wines are distinguished with the PDO label (362 are French, 408 Italian, and 101 Spanish) and 445 with the PGI one [36].

Furthermore, wine is not exempt from fraud as a product with high commercial value and desirability in the market. On the contrary, it is one of the beverages most manipulated or mislabelled for economic purposes. In this line, while practices such as the dilution with water, the addition of alcohol, the addition of sugar (chaptalisation), or the addition of colouring or flavouring substances alter its intrinsic properties; others, including misrepresentation of the grape variety, the age of the wine, or the geographical origin, affect its extrinsic properties [89]. Particularly, wine authentication according to its geographical designation is of great concern, considering that most consumers choose wine depending on it.

Currently, Commission implementing REG (EU) 2019/34 [90] lays down the rules for the application of REG (EU) Nº 1308/2013 [77] as regards PDO, PGI, and TSG labels in the wine sector. Thereby, in agreement with the regulation, PDO and PGI correct compliance shall be verified through the revision of mandatory documentation and organoleptic (*i.e.*, visual appearance, odour, and taste) and physicochemical analytical tests (*i.e.*, control of alcoholic strength, sugars, acidity, volatile acidity, sulphur dioxide, and carbon dioxide). In addition, the analytical databank of wine isotopic data (called European Wine DataBank), which the JRC operates, was recently established and implemented through Commission delegated REG (EU) 2018/273 [91] and Commission implementing REG (EU) 2018/274 [92]. It contains data on the isotopic profile of authentic wines regularly collected from different European countries. Besides, its use aims to detect multiple types of fraud, including geographical identification.

Nevertheless, this thesis contains the development of a new analytical strategy to address wine's geographical origin. This study does not pretend to substitute the existing isotopic profile approach but to be used complementarily or as an alternative.

1.1.6.5. Nuts and seeds

Nuts and seeds encompass an extensive range of products, including almonds, Brazil nuts, cashew nuts, chestnuts, hazelnuts, macadamia nuts, pecans, pine nuts, pistachios, pumpkin seeds, sunflower seeds, and walnuts. Besides, although peanuts are botanically classified as legumes (grow in a pod), consumers often relate them to tree nuts. Therefore, in this thesis, they are included in the nut category.

Consumers usually eat them as a snack, although they can also be added to salads, specific sausages, stews, and bakery products. In addition, various kinds of nut oil and beverages (colloquially known as milks) are also commercialised. Indeed, incorporating some of these goods in the diet is highly recommended since their regular intake has beneficial health effects in humans (*i.e.*, prevention of cardiovascular diseases) [93]. In this line, they are rich in unsaturated fatty acids —monounsaturated fatty acids (MUFA) and polyunsaturated fatty acids (PUFA)— and phenolic and polyphenolic compounds, which play a crucial role in those effects [94,95].

In this context, nuts and seeds are among the most appreciated goods by society, leading to increasing demand and production in the last decades. Nevertheless, they are highly exposed to fraud practices causing economic deception and a threat to human health due to their allergenic properties (mainly because of specific proteins or glycoproteins) [96,97]. Thus, they turn a food fraud issue into a food safety one. Thereby, the most diffuse nuts deceits comprise the substitution with low-quality and cheaper products and the mislabelling of the variety, the geographical origin (in some cases including PDO or PGI labels), the production system, or the crop year [98]. Suman *et al*. [99] summarised the current approaches in industrial facilities to detect

some of these practices. In this regard, direct visual inspection and morphological evaluation are used for detecting defective nuts and recognising specific cultivars. Besides, enzyme-linked immunosorbent assays (ELISA) and lateral flow devices (LFD) are generally employed to detect allergens derived from specific nuts, sometimes followed by deoxyribonucleic acid (DNA) marker analysis for confirmation. Nevertheless, most nut frauds are committed in processed or semi-processed products, making their detection more complicated.

Therefore, this thesis addresses the authentication of nut flour and custard cream, with a special focus on almond products.

1.2. FOODOMICS

The term '*foodomics*' was first described in 2009 by Prof. Cifuentes as *a discipline that studies the Food and Nutrition domains through the application of omics technologies* [100]. In this context, in recent years, omics tools —comprising genomics, transcriptomics, proteomics, metabolomics, and elementomics/isotopollomics (see Figure 1.4)— have been applied to the food fraud field, along with biostatistics and chemometrics [101,102].

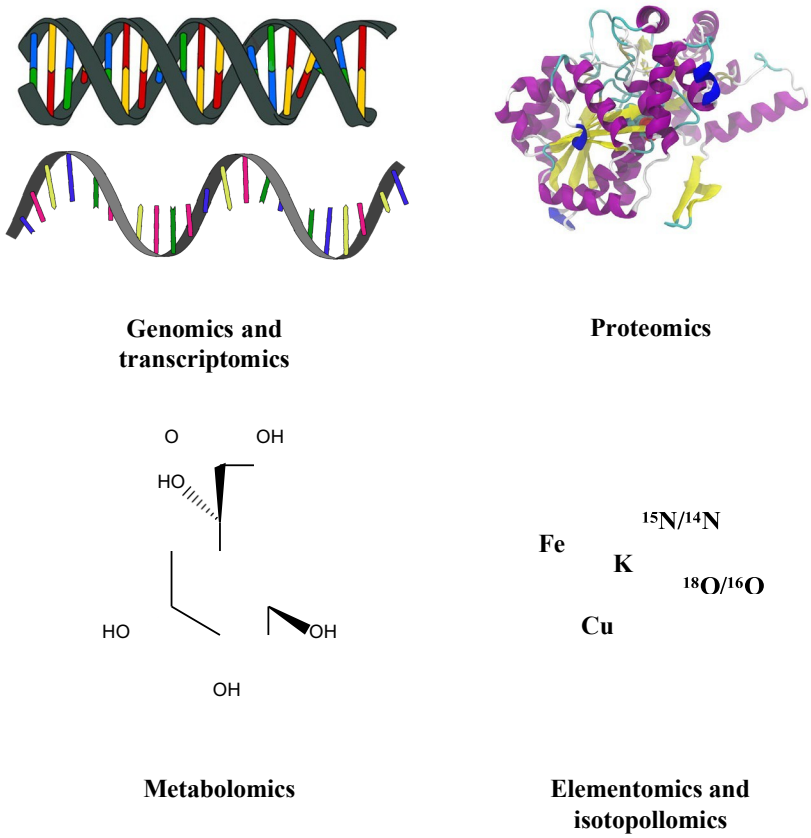


Figure 1.4. Omics tools.

1.2.1. TARGETED AND NON-TARGETED APPROACHES

Traditionally, the analytical methodologies proposed to detect fraudulent practices have relied upon targeted approaches, focusing on the detection and usually quantification of a specific analyte or group of analytes directly linked to the authenticity issue. These analytes are primary markers and often need to meet certain legal limits (*i.e.*, their presence is prohibited or has to be below an established value). For instance, primary markers encompass flavouring/colourant additives or DNA from a specific species, sometimes illegally added. Furthermore, these methods are well-established in routine analysis. This fact is mainly due to the available validation guidelines, which are based on the evaluation of several parameters —*i.e.*, linearity, selectivity, the limit of detection (LOD), the limit of quantification (LOQ), accuracy, robustness, and uncertainty—, and has validity as legal evidence [103].

Furthermore, targeted methods can also be used to determine qualitatively or quantitatively secondary markers whose content can provide helpful information to assess the authenticity of a product indirectly. In this line, '*profiling*' refers to simultaneously determining several targeted analytes (*i.e.*, at least three secondary markers). In this case, profiling results can be either employed to calculate a value to be tested to a reference threshold limit or, more commonly, to be compared to a database. For example, stable isotope ratios and multi-elemental profiles are widely used to classify food products according to their geographical origin.

Finally, non-targeted methods —also so-called '*fingerprinting*'— are based on the unspecific detection of instrumental responses without assuming any previous knowledge about the sample composition. Generally, fingerprinting methods are used in qualitative studies when primary or secondary markers are unknown [104]. In this context, this approach has been extensively used in both proteomics and metabolomics.

Thus, while detecting specific primary markers through targeted methods is typical for an official control purpose in food surveillance, the presence of profiling and

fingerprinting applications is currently limited since there still needs to be a consensus regarding how their validation procedure should be. Nevertheless, food scientists have become increasingly interested in these approaches in the last decade. Therefore, this thesis evaluates the potential of profiling (Chapter 2) and fingerprinting (Chapter 3) approaches to assess several food authentication issues focusing on metabolite detection.

1.2.2. GENOMICS AND TRANSCRIPTOMICS

Genomics involves the comprehensive evaluation of the structure and function of the complete set of DNA, including all of its genes (genome), in an organism or body. Unlike the other omics technologies, it is context-independent since DNA sequences are almost constant in front of exogenous agents. Hence, because of genetic similarity within a species, DNA-based methods have been mostly developed for detecting frauds requiring animal- or plant-species identification or discrimination. Besides, DNA presents good stability, allowing its detection in raw foodstuffs and their processed forms [105]. In this line, several techniques and approaches have been proposed to address food authentication from a genomics perspective.

In this context, most techniques employed in genomics are based on polymerase chain reaction (PCR), one of the most widespread analytical techniques in the food industry, with good sensitivity, accuracy, and ease of testing [106]. Briefly, it allows the exponential amplification of a targeted DNA small region in a thermal cycler, using the corresponding complementary sequence of oligonucleotides (primers), DNA polymerase, and deoxynucleotide triphosphates. Subsequently, DNA fragments are separated electrophoretically (*i.e.*, agarose gel electrophoresis), and the corresponding banding patterns are detected by UV (a fluorescent tag is usually added before the DNA amplification procedure). However, several more recent techniques, such as multiplex-PCR, real-time PCR —also known as quantitative PCR (qPCR)—

or digital droplet PCR (ddPCR), among others, have emerged from conventional PCR [107,108].

Hence, while conventional PCR only focuses on a specific DNA segment, multiplex-PCR simultaneously amplifies several of them using the appropriate primers for each one. For instance, this technique has been applied to identify the octopus species or the meat type in processed products [109,110]. However, despite the usefulness of qualitative species identification in a food product, sometimes accurate quantitation (in the case of matrix effect is rather an estimation) of its content by real-time PCR is required. Thereby, in this technique, and distinctly to the conventional PCR, the DNA amplification is monitored by adding a specific or non-specific substance marked with a fluorophore. Besides, nuclear DNA is preferred over mitochondrial or chloroplast DNA (generally selected for qualitative assays due to its high cell copy number) since the latter can vary between species or tissues, affecting the quantitative step. For example, Lopes *et al*. [111] applied real-time PCR to check that the content of *Pittosporum undulatum* Vent. pollen grains was over 30% in incense honey. Instead, Kim *et al*. [112] successfully employed it to determine the porcine content in processed goods. Recently, methods for on-site food authentication combining PCR-based techniques with microfluidics or biosensors technology have been developed. For instance, authentication of the Atlantic white shrimp (*Litopenaeus setiferus*) was achieved by Kwawukume *et al*. [113] through a developed multiplex PCR-LFD method.

DNA barcoding emerged two decades ago as an alternative to the already mentioned DNA-based techniques, aiming to identify biological species. It compares orthologous DNA regions (approximately 400 – 800 base pairs) after their amplification by PCR and sequencing, which traditionally has been carried out by the Sanger method and, more recently, by next generation sequencing (NGS) tools. Thus, ideally, the DNA barcode should contain low intraspecific and high interspecific genetic variation, with high taxonomic coverage and high resolution [114,115]. In this line, the mitochondrial gene cytochrome c oxidase 1 (COI) is usually employed

as the DNA barcode to identify animal and fungal species, whereas, in the case of plant species, the chloroplast genes carboxylase (rbcL) and maturase K (matK) are used [107]. Furthermore, computerisation and standardisation are among the strengths of this approach [116]. Indeed, the Barcode of Life Data System (BOLD) is an online digital library used as a reference for all living species [117]. For example, in the food fraud field, DNA barcoding has been used for porcini (*Boletus edulis* Bull.), Thai fish fillets, or honey authentication [118–120].

Finally, other techniques —such as high-resolution melting (HRM)— have also been used to assess food authenticity from a genomics perspective. Particularly, HRM is a post-PCR technique that compares the melting curve of amplified DNA segments [108]. For instance, this technique has been employed to address meat species identification of raw and cooked products [121] or the adulteration of commercial tea products with cashew nut husk [122].

Concerning transcriptomics, although it has been used in other food issues [101], its application in the food fraud field is limited.

1.2.3. PROTEOMICS

Studies in proteomics (and, to a lesser extent, peptidomics) have increased in the last years, mostly because of the spread of high-resolution mass spectrometry (HRMS) mass analysers. Particularly in the food field, proteomics comprehensively studies the composition, modifications, and functional properties of protein from products for human consumption [123]. In this line, since protein and peptide composition and content pivot on both endogenous (*i.e.*, genetically derived processes) and exogenous factors (*i.e.*, either from a natural or anthropogenic origin), several mass spectrometry (MS)-based methods have been developed either for biological origin identification or the geographical origin or production system determination [124].

Highly specific and targeted methods to detect a particular protein through ELISA are commonly used in the food industry (especially for meat product analysis) [106]. However, aside from limitations due to interfering compounds or the omission of protein post-translational modifications (PTMs) (*e.g.*, due to protein denaturation caused by thermal procedures), ELISA requires prior knowledge about the markers that need to be determined to solve the issue under study. In most cases, these markers are still unknown, and therefore, a non-targeted discovery phase step is demanded, followed by a targeted-driven phase. In this context, as previously mentioned, MS plays a fundamental role.

Bottom-up and top-down proteomics are the most spread workflows for evaluating protein content and composition [125,126]. Thereby, the peptides resulting from protein enzymatic digestion (*i.e.*, proteolytic cleavage), generally using trypsin, are determined in the bottom-up approach (peptide level). Within this approach, two procedures can be proposed, mainly differing when the separation step and the enzymatic digestion are carried out.

- On the one hand, the traditional bottom-up approach consists of a first protein fractionation by two-dimensional polyacrylamide gel electrophoresis, followed by the enzymatic digestion of each separated spot, and the subsequent peptide characterisation employing matrix-assisted laser desorption ionisation−mass spectrometry (MALDI−MS). Then, protein identification is commonly carried out through peptide mass fingerprinting (PMF), which compares the obtained MS spectra with proper databases.
- On the other hand, the so-called shotgun proteomic approach is usually preferred since liquid chromatography coupled to mass spectrometry (LC−MS) improves the proteome coverage of the analysis. In this case, the entire protein mixture is enzymatically digested, and the resulting peptides are subjected to LC−MS. Thus, peptide fragmentation fingerprinting (PFF) is commonly performed for protein identification and characterisation, implying the study of peptides tandem mass spectrometry (MS/MS) spectra,

which are compared with theoretical ones generated from in silico digestion and stored in protein databases. In this line, while in a non-targeted discovery phase, liquid chromatography coupled to high-resolution mass spectrometry (LC−HRMS) is generally used to identify peptide markers—either with the data-dependent acquisition (DDA) or data-independent acquisition (DIA) modes—, liquid chromatography coupled to low-resolution mass spectrometry (LC−LRMS) allows their targeted monitorisation for protein quantification through the multiple-reaction monitoring (MRM) mode [124,127].

In this context, multiple methods based on the shotgun proteomics approach that address food authenticity are described in the literature. Thus, as mentioned, most of them present a previous non-targeted step with LC−HRMS followed by targeted analysis with LC−LRMS. For instance, the bottom-up approach has been employed to quantify beef and pork meat in highly processed foods (*i.e.*, Bolognese sauce) [106], discriminate shrimp species [128], detect exogenous products such as meat or honey in leguminous-based products [129], or address honey botanical source [130].

In contrast, the top-down proteomics approach is based on the direct MS characterisation of intact proteins (protein level). It does not require enzymatic digestion and sometimes is preceded by LC separation. Unlike the bottom-up approach, it enables obtaining exhaustive structural information (*i.e.*, the complete characterisation of PTMs and isoforms) using HRMS and different dissociation techniques such as collision-induced dissociation (CID), high-energy collision dissociation (HCD), or electron transfer dissociation (ETD). However, this approach still presents limitations such as instrumental constraints or difficulty of data interpretation [123,127]. Therefore, it has yet to be applied for food authentication purposes. Instead, some authors have evaluated intact protein profiles or fingerprints obtained with other analytical techniques as chemical markers. For instance, Yue *et al*. [131] analysed different mono-floral honey types by capillary zone electrophoresis (CZE) to obtain their protein profile, which was subsequently

subjected to discriminant analysis (DA) for sample classification. In another case, Galindo *et al*. [132] developed a MALDI−MS protein fingerprinting method, using a time-of-flight (TOF) mass analyser, for quinoa grains discrimination according to their variety by partial least squares regression-discriminant analysis (PLS-DA).

Furthermore, peptidomics focuses on endogenous peptides in their native form. Thus, peptide determination is also frequently done by LC−MS. However, their low abundance in some matrices makes their determination difficult. Besides, hydrolysis caused by endogenous proteases may differ from the one caused by trypsin, on which traditional databases are based, hindering peptide identification. Therefore, alternatives such as *de novo* sequencing, suspected screening, or non-specific protease databases must be used. Nevertheless, these molecules, especially those of short amino acid chains, have proven to play a relevant role in biological organisms [133]. In this line, Zhao *et al*. [134] authenticated mountain-cultivated ginseng from cultivated one using non-targeted LC−HRMS, which allowed the identification of 52 peptides, of which 20 were discriminant between samples.

1.2.4. METABOLOMICS

The metabolome of a biological cell, tissue, organ, or organism describes its complete set of metabolites, which are organic compounds with a molecular weight of up to 1,500 Da. These compounds present great diversity, with distinct structures and physicochemical properties —being assigned to diverse classes (*i.e.*, amino acids, carbohydrates, lipids, organic acids, or nucleotides, among others)— and can be found in a wide range of concentrations. In this context, metabolomics is the closest omics technique to the phenotype since endogenous metabolites are the end product of multiple processes involving genes, transcripts, and proteins [123,135]. Moreover, external factors such as those of environmental or anthropological origin strongly affect the metabolome, altering its endogenous composition and contributing to exogenous metabolites (*e.g.*, additives, adulterants, or contaminants) [124].

Regarding food analysis, considering that the metabolome of a food product varies according to its biological nature and external conditions, metabolomics has shown excellent potential to assess several issues related to its authenticity and quality. For instance, it has been applied to address food biological identity, geographical origin, and cultivation or production method [136]. In this line, the techniques most used for metabolomic targeted or non-targeted studies are based on spectroscopic, separation, and MS-based techniques (with or without previous separation).

Molecular spectroscopy studies the interaction of different types of electromagnetic radiation with the molecules present in a sample, exciting them and producing an absorption spectrum that contains structural and compositional information [137]. Data obtained from different spectroscopic techniques are widely used in the food authentication field, frequently following a fingerprinting approach. In this line, although infrared (IR), Raman, and nuclear magnetic resonance (NMR) spectroscopies are predominantly chosen, other techniques, such as UV or fluorescence spectroscopies, have also been employed.

UV spectroscopy is based on the fact that molecules containing a chromophore group absorb ultraviolet-visible (UV-Vis) light, causing the excitation of an electron to an excited state. Hence, it evaluates the absorption spectrum after passing UV-Vis light from an incident beam —wavelength range from 1 to 780 nm— through a sample. Although UV spectroscopy has been widely used to estimate quality parameters such as the peroxide value of an oil or the antioxidant capacity of a product, it has also been proposed to solve more complex authenticity-related questions [138]. For instance, UV spectroscopic fingerprints, in combination with chemometrics, have been employed to detect adulteration of organic red pepper powders or authenticate honey from sugar syrups [139,140]. In contrast to UV spectroscopy, fluorescence spectroscopy focuses on the sample emission spectrum after excitation with a UV-Vis light incident beam. As an example of its application, excellent classification of fruit spirits was achieved by subjecting spectra collected through this technique to multivariate analysis [141].

As mentioned, vibrational spectroscopy, which embraces IR and Raman spectroscopies, plays an important role in developing methodologies to assess food quality and authenticity. Indeed, these techniques are suitable for this purpose mainly due to their non-destructive nor invasive nature, green technology, high speed, low cost, reliability, and capability to detect many compounds simultaneously [61,142]. IR spectroscopy measures the absorption spectrum of an IR beam within a sample. Thereby, IR radiation is absorbed when its frequency matches a molecule's vibrational frequency, inducing a molecular vibration (*i.e.*, stretching, bending, or twisting) and causing a change in the dipole moment. Moreover, the IR range is constituted of three regions: the far- (FIR), mid- (MIR), and near-infrared (NIR), with wavelengths ranging from 500,000 to 5,000,000 nm, 3,000 to 500,000 nm, and 780 to 3,000 nm, respectively.

For a long time, IR spectroscopy (especially MIR) was only used to characterise the chemical structure of purified compounds. Nevertheless, over the last decades, IR's relevance in the food science field has increased due to not only technical developments —such as using Fourier transformation (FT) or attenuated total reflectance (ATR)— but also its combination with multivariate analysis. Thus, although MIR was traditionally preferred over NIR for structural characterisation since it provides more sensitive data and narrower bands (making it easier to interpret the spectra), both have already been established in routine analysis to determine chemical properties such as moisture, fat, and protein content in food [143]. In addition, more recent applications have shown that they can also be useful to generate distinctive patterns to authenticate food products. In this context, MIR spectroscopy provides sensitive data related to the fundamental vibrations of the molecules contained in the sample under analysis, being suitable for detecting compounds at low concentrations or discriminating samples with subtle differences. In addition, fitting the MIR spectrophotometer with an ATR module has allowed overcoming its intense water absorption, which was its main drawback when analysing food [142]. Thus, for instance, it has been applied to quantify the adulteration of Australian tea

tree oil (*Melaleuca alternifolia*) with eucalyptus oil [144], to authenticate true cinnamon from false one [145], and to assess *Ternera de Navarra* PGI tenderness [146]. In contrast, although NIR spectroscopy provides broader spectral bands and, hence, less rich in terms of descriptive information, it lacks water interferences and can penetrate much further into the surface of samples, making it convenient for analytical analyses within the agri-food sector [138]. In this line, for example, NIR has been employed in the authentication of *Apulo-Calabrese* pork meat (used in different PDO products) [147], the study of rice freshness [148], and the determination of cocoa shell content in cocoa products [149].

Raman spectroscopy relies on the inelastic Raman scattering phenomenon, which consists of energy exchange between the incident beam (*i.e.*, UV-Vis or IR radiation) and matter. In this case, a photon interacts with the electron cloud of a molecule, causing its polarizability and molecular vibration. Besides, the photon loses energy (Stokes radiation) or, in the case that the molecule is excited, gains it (anti-Stokes radiation), and a shift of the incident light is observed and measured. Because of the inherent weakness of this effect, Raman spectroscopy was challenging to be applied. However, recent developments in terms of instrument design, detector sensitivity, laser construction, and sampling probes have spread its use, although it has also increased its price [137].

In this line, it provides more characteristic spectra (the corresponding bands are sharper) than IR spectroscopy in a faster way. Moreover, Raman spectroscopy is especially adequate for food analysis since it is not altered by moisture content (weak O-H polarizability) and allows analysing samples through their plastic or glass package [61]. Thus, for instance, it has been applied for botanical differentiation of honey varieties [150] and varietal, geographical, and vintage discrimination of white wines [151]. In addition, Arroyo-Cerezo *et al*. [152] addressed the animal origin authentication of plastic-packaged sliced cheeses without requiring the package opening.

Furthermore, beyond lab-based benchtop applications, both IR and Raman spectroscopies have excellent characteristics to be adapted for on-site analysis (*e.g.*, portability, no requirement of highly trained personnel, and provision of reproducible data). Indeed, several handheld spectroscopic devices have already been used in several food fraud applications [153].

Nevertheless, NMR spectroscopy is probably the most suitable spectroscopic technique for comprehensively detecting small molecules (metabolites). It consists of applying radio frequency (RF) radiation to nuclei with magnetic properties while subjected to a constant magnetic field. Briefly, in the presence of a constant magnetic field, the nuclear magnetic moments line up in parallel —spin state α (lower energy)— or antiparallel —spin state β (higher energy)— to it, generating an energy difference. This energy difference is characteristic of each type of nucleus (*i.e.*, ^{1}H, ^{13}C, ^{15}N, ^{19}F, or ^{31}P). Then, the absorption of the equivalent RF radiation produces the resonance of the active nuclei, which emit signals that are detected by the NMR spectrometer and provide structural information [154].

The resonant frequency, as well as the energy of the RF radiation absorbed and the intensity of the signal (sensitivity), are proportional to the strength of the magnetic field. In this line, the NMR instruments available in the market work with resonant frequencies ranging from 40 (low-resolution) to 900 MHz (high-resolution). Moreover, regardless of the device employed, NMR is characterised by its high-throughput and reproducibility. Thus, on the one hand, it allows non-invasive rapid analyses with minimal sample handling [155]. On the other hand, no significant instrumental drift is observed in NMR, and new technological developments have reduced its inter-laboratory variation. These facts have allowed the development of platforms such as Bruker's FoodScreener™, which is composed of four proton nuclear magnetic resonance (^{1}H-NMR) reference databases for honey, olive oil, juice, and wine authentication. Hence, this tool contains representative spectra of authentic samples that can be used for comparison after following a standardised operating

protocol [156,157]. However, its implementation in official controls is limited due to its intellectual property [158].

In general, food authentication studies found in the literature that use NMR rely on 400 MHz spectrometers (chosen as a compromise between the acquisition cost and the resolution) and follow a non-targeted approach followed by multivariate analysis. For instance, oregano and '*super-seeds*' authentication has been addressed in this way [159,160]. Instead, Haddad *et al*. [161] proposed a suspected targeted analysis for species, geographical origin, and variety authentication of cheese, focusing on triacylglycerols NMR signals. Moreover, 600 MHz NMR spectrometers have also been used for this purpose, providing a higher resolving power and, hence, increasing the amount of chemical information and the capacity to characterise unknown compounds [162]. In contrast, some authors have used low-resolution NMR spectrometers, whose purchase and maintenance costs are lower, to address food issues through less resolved spectra. For instance, Gunning *et al*. [163] used the NMR spectra acquired by a 40 MHz spectrometer to assess saffron authenticity in front of different adulterants.

Despite the usefulness of the above-mentioned techniques for food authentication through metabolomic analysis, this thesis exploits separation and/or MS-based techniques for this purpose. In this context, specific details of these techniques are further detailed in Chapter 2 and Chapter 3.

1.2.5. ELEMENTOMICS AND ISOTOPOLLOMICS

Elementomics and isotopollomics approaches comprehensively determine and characterise the sub-molecular composition, ergo metal and non-metal elements. Both have been widely proposed to assess food authentication, particularly geographical origin- or production system-related issues [124].

On the one hand, elementomics focuses on determining the elemental profile, which includes major, minor, trace, and rare-earth elements. In contrast to atomic absorption spectroscopy (AAS) techniques, almost all atomic emission spectroscopy (AES) ones allow performing multi-elemental analysis, which is advantageous for addressing food fraud. In fact, flame atomic emission spectroscopy (FAES) is the only one that does not. In this context, plasma-based systems —inductively coupled plasma atomic emission spectroscopy (ICP-OES) and inductively coupled plasma mass spectrometry (ICP−MS)—, where temperatures around 10,000 K are reached in the excitation and atomisation (and ionisation in ICP−MS) steps, allow the simultaneous determination of a broad range of chemical elements at trace levels, with high reproducibility and a wide dynamic range. Besides, lower LODs are obtained in ICP−MS compared to ICP-OES since *m/z* detection reduces potential interferences [164,165].

Therefore, considering that the elemental composition of agri-food products is strongly influenced by the soil of the growing area (among other factors), either quantitative or semi-quantitative approaches using AES plasma-based techniques, and combined with chemometrics, have been developed to ensure their authenticity [166]. For instance, Inaudi *et al*. [167] quantified 13 elements by ICP-OES to distinguish *Tonda Gentile Trilobata* PGI and Turkish hazelnuts through chemometric exploratory analysis. Instead, in Quinn *et al*. study [168], 40 elements were detected through ICP−MS in rice samples from different origins (China, India, and Vietnam), whose proper authentication was achieved by chemometrics. Furthermore, it is noteworthy that despite the dominance of plasma-based techniques, other techniques —such as energy dispersive-X ray fluorescence (ED-XRF)— have also been employed [169].

On the other hand, isotopollomics provides information regarding the abundance of the isotopic composition of certain elements by means of the stable isotope ratio (δ), which compares the obtained isotope ratio in a sample to the one in a reference material, and is expressed as indicated in Eqn. 1.1.

$$\delta^{i}X\ (‰) = \frac{\left(\frac{hX}{lX}\right)_{sample} - \left(\frac{hX}{lX}\right)_{standard}}{\left(\frac{hX}{lX}\right)_{standard}} \times 1000 \qquad \textbf{1.1}$$

where X is the corresponding element, i is the atomic mass number of the heavier isotope, hX is the concentration of the heavier isotope, and lX is the concentration of the lighter isotope.

According to their atomic mass, stable isotopes can be divided into light isotopes (*i.e.*, $^{13}C/^{12}C$, $^{15}N/^{14}N$, $^{18}O/^{16}O$, $^{2}H/^{1}H$, and $^{34}S/^{32}S$), usually determined by isotope ratio mass spectrometry (IRMS), or heavy isotopes (*i.e.*, mostly $^{87}Sr/^{86}Sr$ and $^{206}Pb/^{204}Pb$), determined using multi-collector-ICP−MS (MC-ICP−MS) [170]. However, generally, a single element's stable isotope ratio is insufficient to provide unambiguous information to assess the food fraud problem. Therefore, the combination of different stable isotopes is used. In this line, the combination of $^{13}C/^{12}C$ and $^{18}O/^{16}O$ or $^{13}C/^{12}C$ and $^{2}H/^{1}H$ has been traditionally used to check the presence of exogenous substances or additives in food products. More recently, changes in the stable isotope ratios of certain elements have been attributed to external factors related to the geographical area. Hence, $^{18}O/^{16}O$ and $^{2}H/^{1}H$ (present in the organic matter of food) are usually linked to the composition of water in the region of origin, $^{15}N/^{14}N$ and $^{13}C/^{12}C$ to the agricultural practices and the climate, and $^{34}S/^{32}S$ and $^{87}Sr/^{86}Sr$ to the soil geology [171,172]. Therefore, several applications of this technique to trace the geographical origin of food products have been described in the last years. For example, Chung *et al*. [173] classified rice samples through stable isotope ratios according to their geographical origin. Besides, some other authors have combined multi-elemental analysis results with stable isotope ratios to achieve sample geographical origin discrimination [174,175].

1.2.6. CHEMOMETRICS

As shown in Section 1.2.1, food fraud issues can be tackled differently depending on their complexity: from highly specific targeted assays to compositional profiles or instrumental fingerprints. In this context, while food scientists have widely employed univariate statistics to evaluate the significance of the results obtained for a certain primary marker in a targeted determination, their application in profiling and fingerprinting methodologies is inadequate. In this case, chemometrics (multivariate data analysis) allows handling the large amount of data generated in these approaches by modern analytical techniques, extracting as much valuable information as possible. Besides, unlike univariate statistics, it considers the complexity of the food product composition describing the information related to each variable individually and the interrelations amongst them, which cannot explicitly be detected [176,177].

Chemometrics was defined by Massart *et al*. [178] as '*the chemical discipline that uses mathematical, statistical, and other methods employing formal logic to design or select optimal measurement procedures and experiments, and to provide maximum relevant chemical information by analysing chemical data*'. Current chemometric methods can be classified as non-supervised (employed for exploratory data analysis) and supervised (used for prediction through regression or classification and discrimination) [179]. In this line, some of them are herein introduced, focusing on those applied in this thesis' research studies.

In general, multivariate data are arranged in a matrix structure. Thus, the so-called X-matrix contains the experimentally obtained data, with rows corresponding to the analysed samples and columns to each experimental variable. At this stage, and before the chemometric analysis, it is fundamental to pre-process the data in order to remove unwanted variability sources (*i.e.*, due to sample inhomogeneity or instrumental drifts) and focus on the information of interest. In this line, the selection of the proper data pre-treatment is strongly influenced by the experimental data type and the purpose of the analysis [180]. For instance, on the one hand, profiling

approaches usually involve simple data consisting of concentration values or instrumental signals used for abundance comparison (*e.g.*, chromatographic peak area or height). Thus, in this case, normalisation using an overall scaling factor —such as the total area or the total sum of intensities— can correct variations between samples. Moreover, since multivariate data analysis is scale dependent and tends to be attracted to larger variations, it is necessary to centre the mean (mean centring) and/or scale the standard deviation (unit scaling) of each variable to equalise their influence on the subsequent model [181]. The so-called autoscaling is the combination of both. On the other hand, data derived from fingerprinting approaches are generally more complex and require extensive pre-processing procedures (*e.g.*, smoothing for noise removal, baseline correction, peak alignment, or working window selection) aside from normalisation and autoscaling.

Non-supervised chemometric techniques are commonly used for exploratory analysis and, therefore, to identify natural trends (similarities and differences) between the analysed samples without previous knowledge. Essentially, they can be divided into principal component analysis (PCA) and cluster analysis —*e.g.*, hierarchical cluster analysis (HCA)—, being the former the most popular in food analysis. Briefly, PCA relies on linearly reducing the dimensionality of the original data (X-matrix) into a set of orthogonal components, known as principal components (PCs), while preserving the maximum variation. Because of their orthogonal nature, the amount of variation described by the PCs decreases consecutively from PC1 to the last. PCA decomposition follows Eqn. 1.2.

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{t} + \mathbf{E} \qquad \textbf{1.2}$$

where X (samples × variables) is the original data matrix, T (samples × PCs) is the scores matrix, P (variables × PCs) is the loadings matrix, and E (samples × variables) is the matrix of residuals.

As a result, representing the obtained PCs' scores in scatter plots (*i.e.*, plotting the scores of two PCs against each other) illustrates the dominant patterns present within

the samples. For instance, the samples' closeness indicates the similarity between them. Instead, loadings plots allow identifying the type of correlation between the original variables and the observed sample trends (direct, inverse, or none). Furthermore, it is noteworthy that apart from being used to detect sample patterns, PCA is also used for quality control (QC) checks and outlier detection. In this last case, the so-called Hotelling's T^2 (extension of the t-test applied to the PCA model) and the Q-statistics (the sum squared residuals) play a crucial role [182].

In contrast, supervised chemometric techniques attempt to predict a characteristic of the food samples under analysis. With this purpose, their models simultaneously consider experimental data (X-matrix) and sample information (Y-matrix), aiming to maximise the description of Y as a function of X. Depending on whether this sample information is quantitative (contiguous parameters such as percentages, concentrations, or time) or qualitative (any category or class), they can be classified as regression- or classification-based techniques. Nevertheless, a common strategy usually employed in both cases involves building the supervised model through a calibration dataset, which sets its parameters, and testing it by validation or unknown samples [183].

As mentioned in Section 1.2.1, standardised guidelines for method validation in multivariate analysis are currently unavailable. However, in this thesis, the validation scheme proposed by Riedl *et al*. has been followed [103]. This scheme (see Figure 1.5) distinguishes four main phases regardless of the supervised chemometric technique type: 1) data preparation, 2) model optimisation by internal validation, 3) model testing by external validation, and 4) stability testing of the model by system challenges.

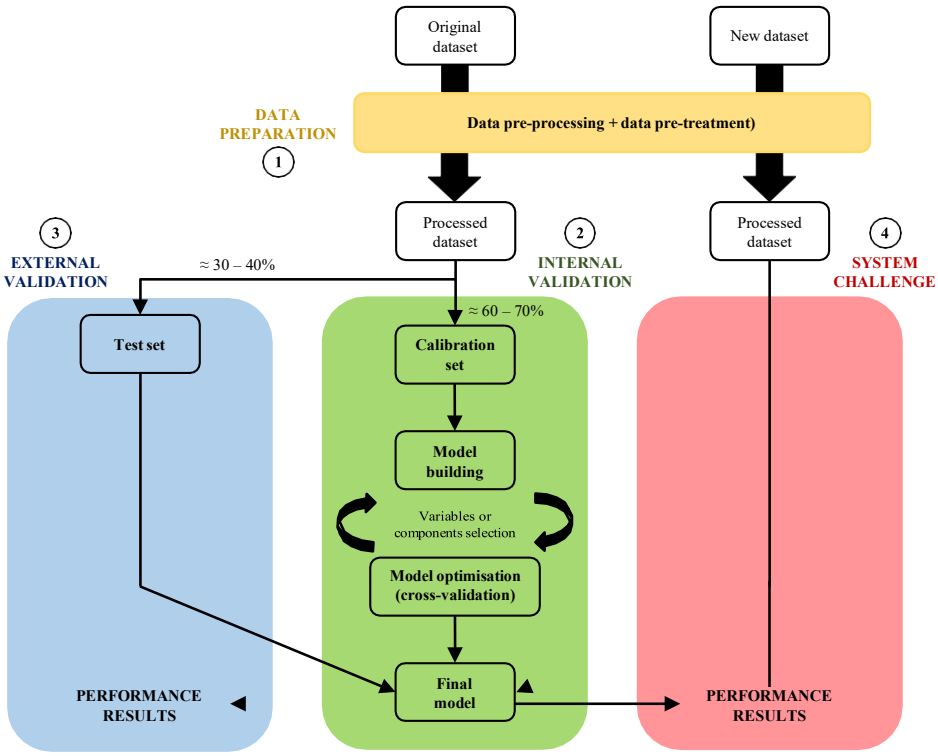


Figure 1.5. Validation scheme for chemometric methods. Adapted from Riedl *et al*. [103].

- Data preparation, which includes pre-processing and pre-treatment, must be identical for all the datasets subjected to the same chemometric model.
- Moreover, model optimisation by internal validation is usually carried out through cross-validation (CV). As shown in Figure 1.6, CV methods (*e.g.*, Venetian blinds, random subsets, or leave-one-out) divide the calibration dataset into diverse splits. Some are selected to construct the chemometric model (training set), and others to test it (test set). This procedure is done iteratively, and the obtained results allow determining the optimal model complexity (*i.e.*, the proper number of components or variables to build the model) and estimating its predictive performance. Therefore, CV prevents poor model prediction results because of an excessive amount of redundant information and noise content (overfitting) or a lack of captured variability in the dataset leading to an information loss (underfitting) [176].

- Instead, the whole dataset is separated into calibration and test sets to validate the model externally. Generally, this is done following ratios of 70:30 or 60:40 (training:test). Thus, while the calibration sample set is used to build and optimise (including the CV step) the supervised chemometric model, the test sample set assesses its prediction ability.
- Finally, to fully validate the chemometric model, the stability testing of the model by system challenges is required. This stage comprises the long-term evaluation of the model by analysing new samples and, therefore, increasing analytical and biological variance. However, this step has not been performed in this thesis due to limited resources.

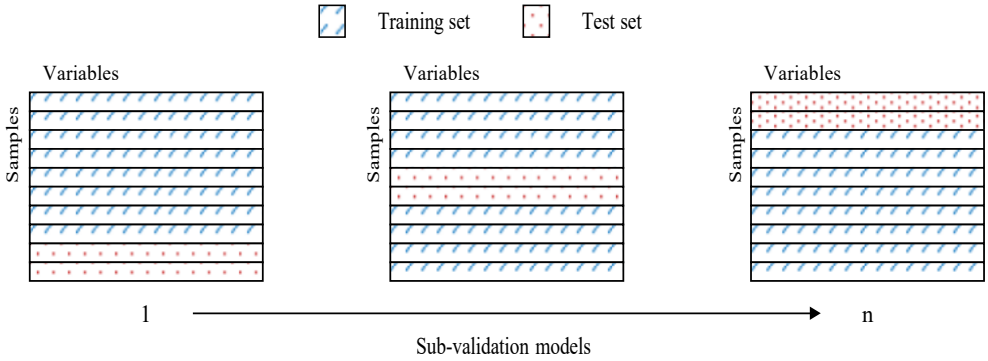


Figure 1.6. Basic operation of the CV procedure.

Focusing on partial least squares (PLS) regression, it is the most employed regression-based chemometric technique and, therefore, usually preferred over other options such as multiple linear regression (MLR). Similarly to PCA, PLS decomposes the original X-matrix following Eqn. 2. Nevertheless, in this case, the original data are transformed into orthogonal linear combinations, known as latent variables (LVs), constructed so that covariance between X and Y is maximum [184]. Different statistics are commonly employed to assess its model performance. For instance, on the one hand, the determination coefficient (R^2) indicates the correlation between the predicted parameter and the experimental data used when presenting values close to 1. On the other hand, root-mean-square errors of calibration

(RMSEC), cross-validation (RMSECV), and prediction (RMSEP) appraise the average uncertainty of potential predictions. Thus, these values are expected to be low in case of good model performance. Moreover, small differences between RMSEC and RMSECV demonstrate good internal consistency, while between RMSECV and RMSEP indicate good predictive capability [142,176].

Furthermore, PLS regression presents different methods to identify and select which are the most relevant variables related to the parameter under study. Hence, variable importance in projection (VIP), regression vector, and selectivity ratio are among the most employed. Besides, they are also used in PLS extensions such as PLS-DA, which is introduced below.

Alternatively, as previously mentioned, supervised classification-based techniques address qualitative issues such as whether a sample belongs to a specific class or not. Currently, they are classified in discriminant analysis and class-modelling techniques. Moreover, it is noteworthy to mention that although some applications have used supervised non-linear techniques —*e.g.*, support vector machine (SVM), artificial neuronal networks (ANN), and k-nearest neighbour (kNN)—, this thesis focuses on supervised linear techniques since their use is more widespread in the food fraud field.

Discriminant analysis techniques classify the samples into limited predefined classes. Although both linear discriminant analysis (LDA) and PLS-DA are widely used, PLS-DA is often preferred when dealing with profiling and, particularly, fingerprinting data since LDA requires datasets with a lower or equal number of variables than samples [142]. In this line, PLS-DA is an extension of PLS in which, instead of quantitative values, the Y-matrix contains categorical dummy variables describing sample class membership (*i.e.*, 1 for class 1, 2 for class 2, and so on). Thus, predicted samples are attributed to the class that presents the minimal distance between Y_{pred} and Y [185]. Nevertheless, PLS-DA results cannot be evaluated using PLS common statistics due to the dummy nature of the Y values. In this case,

parameters such as the overall classification performance (Eqn. 1.3), class sensitivity (Eqn. 1.4), and class specificity (Eqn. 1.5) are used. In addition, some authors also employ the receiver operating characteristic (ROC) curves since they express the balance between sensitivity and specificity [186].

$$\text{Classification accuracy (\%)} = \frac{\text{Well-classified samples}}{\text{Total samples}} \cdot 100 \quad \textbf{1.3}$$

$$\text{Sensitivity (\%)} = \frac{TP}{TP+FN} \cdot 100 \quad \textbf{1.4}$$

$$\text{Specificity (\%)} = \frac{TN}{TN+FP} \cdot 100 \quad \textbf{1.5}$$

where TP = true positive samples, TN = true negative samples, FP = false positive samples, and FN = false negative samples.

Although their great classification power, discriminant analysis techniques are highly influenced by the number of classes included in the model and the number of samples per class. In this regard, optimal discriminant analysis models include few classes, each with many samples. Therefore, some strategies to break down complex classification cases, which include many predefined classes, have been proposed (*i.e.*, classification decision trees).

The main limitation of discriminant analysis techniques concerns the classification of unknown samples that do not belong to one of the classes included in the developed model. In this case, the model would wrongly assign the sample to the class showing more similarities. Thus, this chemometric approach requires extensive coverage of the existing classes to decrease this risk. In contrast to discriminant analysis, class-modelling techniques generate individual models for each targeted class. In this line, the chemometric model is built using only samples belonging to that specific class to define a characteristic multivariate fingerprint; therefore, no information from other classes is used. Hence, each model aims to confirm and detect whether new samples belong to the corresponding class or not [187].

Currently, soft independent modelling of class analogies (SIMCA), the most applied class-modelling chemometric technique in food authentication, follows a two-step procedure. First, a PCA is carried out on the calibration samples of the targeted class to be modelled. Then, since SIMCA is a distance-based technique, the PCA results are used to calculate the class boundary by means of the linear combination of two relevant distances: the score distance (SD) —corresponds to the position of a sample within the score space— and the orthogonal distance (OD) —indicates the distance of a sample to the score space—. Indeed, while SD class limits are computed using Hotelling's T^2, the class boundaries for OD refer to the Q-statistics (in both cases, at a specific confidence limit). Therefore, when an unknown sample is projected, its SD and OD values are compared to the model ones through the so-called reduced distance (d), calculated as shown in Eqn. 1.6. Then, the sample's class membership is confirmed in case d is lower or equal than $\sqrt{2}$ [188,189].

$$\mathbf{1.6} \quad d = \sqrt{\left(\frac{OD_{i,j}}{OD_{j}}\right)^2 + \left(\frac{T^2_{i,j}}{T^2_{lim}}\right)^2} \leq \sqrt{2}$$

where i = index of each given unknown sample and j = class of the model.

In this case, the same figures of merit as in discriminant analysis-based techniques are used to evaluate the performance of the developed models. However, since it is a soft technique, SIMCA may sometimes assign a sample to several or no classes, which suggests that the used data is not sufficiently characteristic to each class. Moreover, although results are not comparable with discriminant techniques since they present different purposes, SIMCA is less powerful in classification. In this line, the PCs used to construct a SIMCA model are based on the largest variation and thus, they could not be related to the targeted class under evaluation [183,185].

1.3. REFERENCES

[1] F.L. Hart, A history of the adulteration of food before 1906, Food. Drug. Cosmet. Law J. 7 (1952) 5–22.

[2] A. Gonzalez, A history of food fraud, (2019). https://www.thermofisher.com/blog/food/a-history-of-food-fraud/ (accessed November 9, 2022).

[3] F. Accum, A treatise on adulteration of food, and culinary poisons, AB'M Small, London, 1820.

[4] The Parliament of the United Kingdom, Public Health Act 1875, UK Public Gen. Acts. 55 (1875). https://www.legislation.gov.uk/ukpga/Vict/38-39/55/enacted.

[5] The Parliament of the United Kingdom, Parliamentary Notes. Food and Drugs (Adulteration) Act 1928, Analyst. 53 (1928) 600–601. https://doi.org/10.1039/AN9285300600.

[6] République Française, Loi sur la répression des fraudes dans la vente des marchandises et les falsifications des denrées alimentaires et des produits agricoles, J. Off. La République Française. 210 (1905) 4813–4836.

[7] United States Congress, Federal Meat Inspection Act, Public Law. 34 (1906) 59–382.

[8] United States Congress, Pure Food and Drug Act, Public Law. 34 (1906) 59–384.

[9] S.P. Kennedy, History of food fraud and development of mitigation requirements and standards, in: R.S. Hellberg, K. Everstine, S.A. Sklare (Eds.), Food Fraud - a Glob. Threat with Public Heal. Econ. Consequences, Academic Press, London, 2021: pp. 9–22.

[10] B. Bréville, Consider the croissant: a history of food fraud, Le Monde Dipl. (2022). https://mondediplo.com/2022/06/12food.

[11] Generalitat de Catalunya. Departament d'agricultura, ramaderia, pesca i alimentació. Principals fraus alimentaris 1981-2014, 2017. https://agricultura.gencat.cat/web/.content/04-alimentacio/lluita-frau-alimentari/enllacos-documents/fitxers-binaris/cronograma_complet.pdf.

[12] P.H. and F.S. Committee on the Environment, Report on the food crisis, fraud in the food chain and the control thereof (2013/2091(INI)), (2013) 1–23. https://www.europarl.europa.eu/doceo/document/A-7-2013-0434_EN.pdf (accessed December 12, 2022).

[13] K. Robson, M. Dean, S. Haughey, C. Elliott, A comprehensive review of food fraud terminologies and food fraud mitigation guides, Food Control. 120 (2021) 107516. https://doi.org/10.1016/j.foodcont.2020.107516.

[14] C. Elliott, Elliott review into the integrity and assurance of food supply networks - final report, 2014.

[15] L. Manning, J.M. Soon, Food fraud vulnerability assessment: Reliable data sources

and effective assessment approaches, Trends Food Sci. Technol. 91 (2019) 159–168. https://doi.org/10.1016/j.tifs.2019.07.007.

[16] CBI Ministry of Foreign Affairs, European importers and EU authorities increase focus on the mislabelling of added water in seafood, (2019). https://www.cbi.eu/news/european-importers-eu-authorities-increase-focus-mislabelling-added-water-seafood (accessed November 20, 2022).

[17] European Union Food Integrity Project, FoodIntegrity, (2017). https://foodintegrity.fera.co.uk/ (accessed November 21, 2022).

[18] J. Spink, D.C. Moyer, Defining the public health threat of food fraud, J. Food Sci. 76 (2011) 157–163. https://doi.org/10.1111/j.1750-3841.2011.02417.x.

[19] L. Manning, Food fraud: policy and food chain, Curr. Opin. Food Sci. 10 (2016) 16–21. https://doi.org/10.1016/j.cofs.2016.07.001.

[20] The European Committee for Standardization, CWA 17369: Authenticity and fraud in the feed and food chain - Concepts, terms, and definitions, 2019.

[21] B. Popping, N. Buck, D. Bánáti, P. Brereton, S. Gendel, N. Hristozova, S.M. Chaves, S. Saner, J. Spink, C. Willis, D. Wunderlin, Food inauthenticity: Authority activities, guidance for food operators, and mitigation tools, Compr. Rev. Food Sci. Food Saf. 21 (2022) 4776–4811. https://doi.org/10.1111/1541-4337.13053.

[22] The European Parliament and the Council of the European Union, Regulation (EC) Nº 178/2002 of the European Parliament and of the Council of 28 January 2002, Off. J. Eur. Communities. L31 (2002) 1–24.

[23] B.M.J. van der Meulen, B.F.W. Wernaart, Food law and regulatory affairs, in: B.F.W. Wernaart, B.M.J. van der Meulen (Eds.), Appl. Food Sci., 1st ed., Wageningen Academic Publishers, Wageningen (The Netherlands), 2022: pp. 21–43. https://doi.org/10.3920/978-90-8686-933-6_2.

[24] The European Commission, White paper on food safety, Brussels, 2000.

[25] The European Parliament and the Council of the European Union, Regulation (EC) Nº 882/2004 of the European Parliament and of the Council of 29 April 2004, Off. J. Eur. Union. L165 (2004) 1–141.

[26] The European Parliament and the Council of the European Union, Regulation (EC) Nº 854/2004 of the European Parliament and of the Council of 29 April 2004, Off. J. Eur. Union. L155 (2004) 206.

[27] The European Parliament and the Council of the European Union, Regulation (EC) Nº 1333/2008 of the European Parliament and of the Council of 16 December 2008, Off. J. Eur. Union. L354 (2008) 16.

[28] The European Parliament and the Council of the European Union, Regulation (EU) Nº 1169/2011 of the European Parliament and of the Council of 25 October 2011, Off. J. Eur. Union. L304 (2011) 18–63.

[29] The European Parliament and the Council of the European Union, Directive

2000/13/EC of the European Parliament and of the Council of 20 March 2000, Off. J. Eur. Union. L109 (2000) 29–42.

[30] The European Parliament and the Council of the European Union, Regulation (EC) Nº 1924/2006 of the European Parliament and of the Council of 20 December 2006, Off. J. Eur. Union. L404 (2006) 9–25.

[31] The European Parliament and the Council of the European Union, Regulation (EU) Nº 2017/625 of the European Parliament and of the Council of 15 March 2017, Off. J. Eur. Union. L95 (2017) 1–142.

[32] The European Parliament and the Council of the European Union, Commission Implementing Regulation (EU) 2019/1715 of 30 September 2019, Off. J. Eur. Union. L261 (2019) 37.

[33] The European Parliament and the Council of the European Union, Regulation (EU) Nº 1151/2012 of the European Parliament and of the Council of 21 November 2012, Off. J. Eur. Union. L343 (2012) 1–39.

[34] The Council of the European Union, Council Regulation (EC) Nº 509/2006 of 20 March 2006, Off. J. Eur. Union. L 93 (2006) 1–11.

[35] The Council of the European Union, Council Regulation (EC) Nº 510/2006 of 20 March 2006, Off. J. Eur. Union. L93 (2006) 12–25.

[36] Directorate-General for Agriculture and Rural Development, eAmbrosia - The EU geographical indications register, (2022). https://ec.europa.eu/info/food-farming-fisheries/food-safety-and-quality/certification/quality-labels/geographical-indications-register/ (accessed December 13, 2022).

[37] The European Parliament and the Council of the European Union, Regulation (EU) Nº 2019/787 of the European Parliament and of the Council of 17 April 2019, Off. J. Eur. Union. L130 (2019) 1–68.

[38] The European Parliament and the Council of the European Union, Regulation (EU) Nº 2021/2117 of the European Parliament and of the Council of 2 December 2021, Off. J. Eur. Union. L435 (2021) 262–314.

[39] Directorate-General for Agriculture and Rural Development, eAmbrosia - Traditional Specialities Guaranteed, (2022). https://ec.europa.eu/info/food-farming-fisheries/food-safety-and-quality/certification/quality-labels/geographical-indications-register/tsg (accessed December 13, 2022).

[40] The European Parliament and the Council of the European Union, Regulation (EC) Nº 228/2013 of the European Parliament and of the Council of 13 March 2013, Off. J. Eur. Union. L78 (2013) 1–24.

[41] The European Parliament and the Council of the European Union, Council Regulation (EC) Nº 834/2007 of 28 June 2007, Off. J. Eur. Union. L189 (2007) 1–39.

[42] The European Parliament and the Council of the European Union, Regulation (EU) 2018/848 of the European Parliament and of the Council of 30 May 2018, Off. J. Eur. Union. L150 (2018) 1–134.

[43] A. Bērziņš, EFSA consolidated annual activity report 2021, 2022. https://doi.org/10.2805/226427.

[44] Directorate-General for Communication, Departments and executive agencies. Departments, (2023). https://commission.europa.eu/about-european-commission/departments-and-executive-agencies_en (accessed January 16, 2023).

[45] Directorate-General for Health and Food Safety, Annual activity report 2021. DG Health and Food Safety (SANTE), 2021.

[46] The European Commission, 2021 Annual Report. Alert and Cooperation Network, 2022. https://doi.org/10.2875/328358.

[47] Agencia Española de Seguridad Alimentaria y Nutrición, National plan for official control of the food chain 2021-2025 (V2-2022), 2022. https://doi.org/069-22-002X.

[48] Joint Research Centre, JRC Technical Report. Results of an EU wide coordinated control plan to establish the prevalence of fraudulent practices in the marketing of herbs and spices, 2021. https://doi.org/10.2760/309557.

[49] The European Commission, Knowledge Service. Knowledge Centre for Food Fraud and Quality, (2023). https://knowledge4policy.ec.europa.eu/food-fraud-quality_en (accessed January 17, 2023).

[50] Global Food Safety Initiative, Tackling food fraud through food safety management systems, 2018.

[51] National Advisory Committee on Microbiological Criteria for Foods, HACCP Principles and application guidelines, 1997.

[52] J. Spink, D.L. Ortega, C. Chen, F. Wu, Food fraud prevention shifts the food risk focus to vulnerability, Trends Food Sci. Technol. 62 (2017) 215–220. https://doi.org/10.1016/j.tifs.2017.02.012.

[53] S.M. van Ruth, W. Huisman, P.A. Luning, Food fraud vulnerability and its key factors, Trends Food Sci. Technol. 67 (2017) 70–75. https://doi.org/10.1016/j.tifs.2017.06.017.

[54] V. Barrere, K. Everstine, J. Théolier, S. Godefroy, Food fraud vulnerability assessment: Towards a global consensus on procedures to manage and mitigate food fraud, Trends Food Sci. Technol. 100 (2020) 131–137. https://doi.org/10.1016/j.tifs.2020.04.002.

[55] J.M. Soon, S.C. Krzyzaniak, Z. Shuttlewood, M. Smith, L. Jack, Food fraud vulnerability assessment tools used in food industry, Food Control. 101 (2019) 225–232. https://doi.org/10.1016/j.foodcont.2019.03.002.

[56] A.M. Pustjens, Y. Weesepoel, S.M. van Ruth, Food fraud and authenticity: Emerging issues and future trends, in: C.E. Leadley (Ed.), Innov. Futur. Trends Food Manuf. Supply Chain Technol., Elsevier Ltd, 2016: pp. 3–20. https://doi.org/10.1016/B978-1-78242-447-5.00001-0.

[57] A.G. Osman, V. Raman, S. Haider, Z. Ali, A.G. Chittiboyina, I.A. Khan, Overview

of analytical tools for the identification of adulterants in commonly traded herbs and spices, J. AOAC Int. 102 (2019) 376–385. https://doi.org/10.5740/jaoacint.18-0389.

[58] S. Barbosa, Desarrollo de metodologías analíticas para la caracterización y autenticación de fármacos y alimentos derivados de productos naturales de origen vegetal y ricos en polifenoles, University of Barcelona, 2020.

[59] European Statistical Office (Eurostat), Key figures on the European food chain. 2022 edition, 2022. https://doi.org/10.2785/500223.

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

[61] N. Modupalli, M. Naik, C.K. Sunil, V. Natarajan, Emerging non-destructive methods for quality and safety monitoring of spices, Trends Food Sci. Technol. 108 (2021) 133–147. https://doi.org/10.1016/j.tifs.2020.12.021.

[62] M.M. Oliveira, J.P. Cruz-Tirado, D.F. Barbin, Nontargeted analytical methods as a powerful tool for the authentication of spices and herbs: A review, Compr. Rev. Food Sci. Food Saf. 18 (2019) 670–689. https://doi.org/10.1111/1541-4337.12436.

[63] S. Réhault-Godbert, N. Guyot, Y. Nys, The golden egg: nutritional value, bioactivities, and emerging benefits for human health, Nutrients. 11 (2019) 684. https://doi.org/10.3390/nu11030684.

[64] Food and Agriculture Organization of the United Nations, FAOSTAT, (2022). https://www.fao.org/faostat/en/#data/QCL/ (accessed January 3, 2023).

[65] EFSA Panel on Biological Hazards, Scientific opinion on the public health risks of table eggs due to deterioration and development of pathogens, EFSA J. 12 (2014) 3782. https://doi.org/10.2903/j.efsa.2014.3782.

[66] The European Commission, Commission Regulation (EC) Nº 589/2008 of 23 June 2008, Off. J. Eur. Union. L163 (2008) 6.

[67] The European Parliament and the Council of the European Union, Regulation (EC) Nº 853/2004 of the European Parliament and of the Council of 29 April 2004, Off. J. Eur. Union. L139 (2004) 55.

[68] The Council of the European Union, Council Directive 1999/74/EC of 19 July 1999, Off. J. Eur. Union. L203 (1999) 53.

[69] M.C. Appleby, The European Union ban on conventional cages for laying hens: history and prospects, J. Appl. Anim. Welf. Sci. 6 (2010) 103–121. https://doi.org/10.1207/S15327604JAWS0602_03.

[70] The Commission of the European Communities, Commission Directive 2002/4/EC of 30 January 2002, Off. J. Eur. Union. L30 (2002) 44.

[71] The European Commission, Commission Implementing Regulation (EU) 2017/1185 of 20 April 2017, Off. J. Eur. Union. L171 (2017) 113–130.

[72] Directorate-General for Agriculture and Rural Development, Agriculture and rural development. Eggs, (2023). https://agriculture.ec.europa.eu/farming/animal-products/eggs_en (accessed January 9, 2023).

[73] The European Commission, Press release. European Citizens' Initiative: Commission to propose phasing out of cages for farm animals, Brussels, 2021.

[74] F. Leenstra, V. Maurer, F. Galea, M. Bestman, Z. Amsler-Kepalaite, J. Visscher, I. Vermeij, M. van Krimpen, Laying hen performance in different production systems; why do they differ and how to close the gap? Results of discussions with groups of farmers in the Netherlands, Switzerland and France, benchmarking and model calculations., Eur. Poult. Sci. 78 (2014) 1–10. https://doi.org/10.1399/eps.2014.53.

[75] G. Puertas, P. Cazón, M. Vázquez, A quick method for fraud detection in egg labels based on egg centrifugation plasma, Food Chem. 402 (2023) 134507. https://doi.org/10.1016/j.foodchem.2022.134507.

[76] M. Tarapoulouzi, S. Agriopoulou, A. Koidis, C. Proestos, H.A. El Enshasy, T. Varzakas, Recent advances in analytical methods for the detection of olive oil oxidation status during storage along with chemometrics, authenticity and fraud studies, Biomolecules. 12 (2022) 1180. https://doi.org/10.3390/biom12091180.

[77] The European Parliament and the Council of the European Union, Regulation (EU) Nº 1308/2013 of the European Parliament and of the Council of 17 December 2013, Off. J. Eur. Union. L347 (2013) 671.

[78] The European Commission, Commission Delegated Regulation (EU) 2022/2104 of 29 July 2022, Off. J. Eur. Union. L284 (2022) 1–22.

[79] The European Commission, Commission Implementing Regulation (EU) 2022/2105 of 29 July 2022, Off. J. Eur. Union. L284 (2022) 23–47.

[80] The Commission of the European Communities, Commission Regulation (EEC) Nº 2568/91 of 11 July 1991, Off. J. Eur. Communities. L248 (1991) 1–83.

[81] A. González-Pereira, P. Otero, M. Fraga-Corral, P. Garcia-Oliveira, M. Carpena, M.A. Prieto, J. Simal-Gandara, State-of-the-art of analytical techniques to determine food fraud in olive oils, Foods. 10 (2021) 484. https://doi.org/10.3390/foods10030484.

[82] A. Arrizabalaga-Larrañaga, F.J. Santos, E. Moyano, Liquid chromatography pigment profile for characterization and fraud detection in olive oils, in: O. Núñez, G. Campmajó (Eds.), Chromatogr. Relat. Sep. Tech. Food Integr. Authenticity. Vol. B Relev. Appl., World Scientific Publishing Europe Ltd., 2021: pp. 21–41. https://doi.org/10.1142/9781786349972_0002.

[83] L. Conte, A. Bendini, E. Valli, P. Lucci, S. Moret, A. Maquet, F. Lacoste, P. Brereton, D.L. García-González, W. Moreda, T. Gallina Toschi, Olive oil quality and authenticity: A review of current EU legislation, standards, relevant methods of analyses, their drawbacks and recommendations for the future, Trends Food Sci. Technol. 105 (2020) 483–493. https://doi.org/10.1016/j.tifs.2019.02.025.

[84] Università di Bologna, The Oleum Project, (2023). http://www.oleumproject.eu/ (accessed January 22, 2023).

[85] Federació de Cooperatives Agràries de Catalunya, AUTENFOOD. Autenticació, traçabilitat i seguretat alimentària., (2023). https://www.cooperativesagraries.cat/ca/projectes-assolits/l/10305-autenfood.html (accessed January 23, 2023).

[86] L.M. Buja, The history, science, and art of wine and the case for health benefits: perspectives of an oenophilic cardiovascular pathologist, Cardiovasc. Pathol. 60 (2022) 107446. https://doi.org/10.1016/j.carpath.2022.107446.

[87] International Organisation of Vine and Wine, International Organisation of Vine and Wine, (2022). https://www.oiv.int/ (accessed January 11, 2023).

[88] V. Merkyte, E. Longo, G. Windisch, E. Boselli, Phenolic compounds as markers of wine quality and authenticity, Foods. 9 (2020) 1785. https://doi.org/10.3390/foods9121785.

[89] C. Villano, M.T. Lisanti, A. Gambuti, R. Vecchio, L. Moio, L. Frusciante, R. Aversano, D. Carputo, Wine varietal authentication based on phenolics, volatiles and DNA markers: State of the art, perspectives and drawbacks, Food Control. 80 (2017) 1–10. https://doi.org/10.1016/j.foodcont.2017.04.020.

[90] The European Commission, Commission Implementing Regulation (EU) 2019/34 of 17 October 2018, Off. J. Eur. Union. L9 (2019) 46–76.

[91] The European Commission, Commission Delegated Regulation (EU) 2018/273 of 11 December 2017, Off. J. Eur. Union. L58 (2018) 1.

[92] The European Commission, Commission Implementing Regulation (EU) 2018/274 of 11 December 2017, Off. J. Eur. Union. L58 (2018) 60.

[93] E. Bitok, J. Sabaté, Nuts and cardiovascular disease, Prog. Cardiovasc. Dis. 61 (2018) 33–37. https://doi.org/10.1016/j.pcad.2018.05.003.

[94] E. Ros, Nuts and CVD, Br. J. Nutr. 113 (2015) S111–S120. https://doi.org/10.1017/S0007114514003924.

[95] D. Del Rio, A. Rodriguez-Mateos, J.P.E. Spencer, M. Tognolini, G. Borges, A. Crozier, Dietary (poly)phenolics in human health: Structures, bioavailability, and evidence of protective effects against chronic diseases, Antioxidants Redox Signal. 18 (2013) 1818–1892. https://doi.org/10.1089/ars.2012.4581.

[96] A. Valdés, A. Beltrán, C. Mellinas, A. Jiménez, M.C. Garrigós, Analytical methods combined with multivariate analysis for authentication of animal and vegetable food products with high fat content, Trends Food Sci. Technol. 77 (2018) 120–130. https://doi.org/10.1016/j.tifs.2018.05.014.

[97] T. Holzhauser, M. Röder, Allergens in tree nuts, sesame seeds, mustard, and celery, in: L.M.L. Nollet, A.J. van Hengel (Eds.), Food Allergens Anal. Intrumentation Methods, 1st ed., Taylor & Francis Group, 2011: pp. 77–128.

[98] G. Campmajó, O. Núñez, Authentication of nut and seed products, in: O. Núñez, G. Campmajó (Eds.), Chromatogr. Relat. Sep. Tech. Food Integr. Authenticity. Vol. B Relev. Appl., World Scientific Publishing Europe Ltd., 2021: pp. 129–155. https://doi.org/10.1142/9781786349972_0005.

[99] M. Suman, D. Cavanna, G. Sammarco, F. Lambertini, C. Loffi, Fighting food frauds exploiting chromatography-mass spectrometry technologies: Scenario comparison between solutions in scientific literature and real approaches in place in industrial facilities, TrAC - Trends Anal. Chem. 142 (2021) 116305. https://doi.org/10.1016/j.trac.2021.116305.

[100] A. Cifuentes, Food analysis and foodomics, J. Chromatogr. A. 1216 (2009) 7109. https://doi.org/10.1016/j.chroma.2009.09.018.

[101] A. Valdés, G. Álvarez-Rivera, B. Socas-Rodríguez, M. Herrero, E. Ibáñez, A. Cifuentes, Foodomics: analytical opportunities and challenges, Anal. Chem. 94 (2022) 366–381. https://doi.org/10.1021/acs.analchem.1c04678.

[102] M. Creydt, M. Fischer, Food authentication in real life: how to link nontargeted approaches with routine analytics?, Electrophoresis. 41 (2020) 1665–1679. https://doi.org/10.1002/elps.202000030.

[103] J. Riedl, S. Esslinger, C. Fauhl-Hassek, Review of validation and reporting of non-targeted fingerprinting approaches for food authentication, Anal. Chim. Acta. 885 (2015) 17–32. https://doi.org/10.1016/j.aca.2015.06.003.

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

[105] E. Haynes, E. Jimenez, M.A. Pardo, S.J. Helyar, The future of NGS (Next Generation Sequencing) analysis in testing food authenticity, Food Control. 101 (2019) 134–143. https://doi.org/10.1016/j.foodcont.2019.02.010.

[106] B. Prandi, F. Lambertini, A. Faccini, M. Suman, A. Leporati, T. Tedeschi, S. Sforza, Mass spectrometry quantification of beef and pork meat in highly processed food: Application on Bolognese sauce, Food Control. 74 (2017) 61–69. https://doi.org/10.1016/j.foodcont.2016.11.032.

[107] P. Kumar, A. Rani, S. Singh, A. Kumar, Recent advances on DNA and omics-based technology in food testing and authentication: A review, J. Food Saf. 42 (2022) e12986. https://doi.org/10.1111/jfs.12986.

[108] K. Böhme, P. Calo-Mata, J. Barros-Velázquez, I. Ortea, Review of recent DNA-based methods for main food-authentication topics, J. Agric. Food Chem. 67 (2019) 3854–3864. https://doi.org/10.1021/acs.jafc.8b07016.

[109] Y.M. Lee, G.Y. Lee, H.Y. Kim, Development of a multiplex PCR assay for the simultaneous detection of big blue octopus (Octopus cyanea), giant Pacific octopus (Enteroctopus dofleini), and common octopus (Octopus vulgaris), Food Sci. Biotechnol. 31 (2022) 497–504. https://doi.org/10.1007/s10068-022-01051-w.

[110] C. Yang, G. Zhong, S. Zhou, Y. Guo, D. Pan, S. Wang, Q. Liu, Q. Xia, Z. Cai, Detection and characterization of meat adulteration in various types of meat products by using a high-efficiency multiplex polymerase chain reaction technique, Front. Nutr. 9 (2022) 979977. https://doi.org/10.3389/fnut.2022.979977.

[111] A.R. Lopes, M. B.M.V. Moura, L. Grazina, J. Costa, J.S. Amaral, M. Alice Pinto, I. Mafra, Authentication of incense (Pittosporum undulatum Vent.) honey from the Azores (Mel dos Açores) by a novel real-time PCR approach, Food Chem. 411 (2023) 135492. https://doi.org/10.1016/j.foodchem.2023.135492.

[112] Y. Kim, H.-S. Lee, K.-G. Lee, Detection of porcine DNA in Korean processed foods by real-time PCR, Food Sci. Biotechnol. 32 (2023) 21–26. https://doi.org/10.1007/s10068-022-01169-x.

[113] S. Kwawukume, F.J. Velez, D. Williams, L. Cui, P. Singh, Rapid PCR-lateral flow assay for the onsite detection of Atlantic white shrimp, Food Chem. Mol. Sci. 6 (2023) 100164. https://doi.org/10.1016/j.fochms.2023.100164.

[114] P.D.N. Hebert, A. Cywinska, S.L. Ball, J.R. DeWaard, Biological identifications through DNA barcodes, Proc. R. Soc. B Biol. Sci. 270 (2003) 313–321. https://doi.org/10.1098/rspb.2002.2218.

[115] A. Galimberti, F. De Mattia, A. Losa, I. Bruni, S. Federici, M. Casiraghi, S. Martellos, M. Labra, DNA barcoding as a new tool for food traceability, Food Res. Int. 50 (2013) 55–63. https://doi.org/10.1016/j.foodres.2012.09.036.

[116] M. Casiraghi, M. Labra, E. Ferri, A. Galimberti, F. de Mattia, DNA barcoding: A six-question tour to improve users' awareness about the method, Brief. Bioinform. 11 (2010) 440–453. https://doi.org/10.1093/bib/bbq003.

[117] International Barcode of Life, Barcode of Life Data System, (2014). https://boldsystems.org/ (accessed February 21, 2023).

[118] M. Rivas-Ferreiro, A. Otero, P. Morán, It's what's inside that counts: DNA-barcoding of porcini (Boletus sp., Basidiomycota) commercial products reveals product mislabelling, Food Control. 144 (2023) 109346. https://doi.org/10.1016/j.foodcont.2022.109346.

[119] D. Panprommin, R. Manosri, DNA barcoding as an approach for species traceability and labeling accuracy of fish fillet products in Thailand, Food Control. 136 (2022) 108895. https://doi.org/10.1016/j.foodcont.2022.108895.

[120] D. Chavan, J.R.T. Adolacion, M. Crum, S. Nandy, K.H. Lee, B. Vu, K. Kourentzi, A. Sabo, R.C. Willson, Isolation and barcoding of trace pollen-free DNA for authentication of honey, J. Agric. Food Chem. 70 (2022) 14084–14095. https://doi.org/10.1021/acs.jafc.2c04309.

[121] P. Gholamnezhad, H. Ahari, G. Nikbakht Brujeni, S.A.A. Anvar, A. Motallebi, Evaluation of high resolution melting (HRM) analysis for meat species identification of raw and cooked meat, Separations. 8 (2021) 116. https://doi.org/10.3390/separations8080116.

[122] G. Lagiotis, E. Stavridou, I. Bosmali, M. Osathanunkul, N. Haider, P. Madesis, Detection and quantification of cashew in commercial tea products using high resolution melting (HRM) analysis, J. Food Sci. 85 (2020) 1629–1634. https://doi.org/10.1111/1750-3841.15138.

[123] V.G. Zaikin, R.S. Borisov, Mass spectrometry as a crucial analytical basis for omics sciences, J. Anal. Chem. 76 (2021) 1567–1587. https://doi.org/10.1134/S1061934821140094.

[124] M. Creydt, M. Fischer, Omics approaches for food authentication, Electrophoresis. 39 (2018) 1569–1581. https://doi.org/10.1002/elps.201800004.

[125] M. Valletta, S. Ragucci, N. Landi, A. Di Maro, P.V. Pedone, R. Russo, A. Chambery, Mass spectrometry-based protein and peptide profiling for food frauds, traceability and authenticity assessment, Food Chem. 365 (2021) 130456. https://doi.org/10.1016/j.foodchem.2021.130456.

[126] J.M. Gallardo, I. Ortea, M. Carrera, Proteomics and its applications for food authentication and food-technology research, TrAC - Trends Anal. Chem. 52 (2013) 135–141. https://doi.org/10.1016/j.trac.2013.05.019.

[127] H.J. Chien, Y.F. Zheng, W.C. Wang, C.Y. Kuo, Y.M. Hsu, C.C. Lai, Determination of adulteration, geographical origins, and species of food by mass spectrometry, Mass Spectrom. Rev. (2022) e21780. https://doi.org/10.1002/mas.21780.

[128] L. Hu, H. Zhang, X. Zhang, T. Zhang, Y. Chang, X. Zhao, J. Xu, Y. Xue, Z. Li, Y. Wang, C. Xue, Identification of peptide biomarkers for discrimination of shrimp species through SWATH-MS-based proteomics and chemometrics, J. Agric. Food Chem. 66 (2018) 10567–10574. https://doi.org/10.1021/acs.jafc.8b04375.

[129] G. Huschek, J. Bönick, D. Merkel, D. Huschek, H. Rawel, Authentication of leguminous-based products by targeted biomarkers using high resolution time of flight mass spectrometry, Lwt. 90 (2018) 164–171. https://doi.org/10.1016/j.lwt.2017.12.034.

[130] B. Silva, A. Carolina, O. Costa, S. Sagu, B. Josephine, G. Huschek, L. Valdemiro, R. Fett, S. Baldermann, H.M. Rawel, Comparative quantification and differentiation of bracatinga (Mimosa scabrella Bentham) honeydew honey proteins using targeted peptide markers identified by high-resolution mass spectrometry, Food Res. Int. 141 (2021) 109991. https://doi.org/10.1016/j.foodres.2020.109991.

[131] J. Yue, J. Fan, Y. Li, H. Ren, Rapid authentication of mono-floral honeys by capillary zone electrophoresis, J. Food Meas. Charact. 15 (2021) 3553–3560. https://doi.org/10.1007/s11694-021-00914-w.

[132] R. Galindo-Luján, L. Pont, V. Sanz-Nebot, F. Benavente, Protein profiling and classification of commercial quinoa grains by MALDI-TOF-MS and chemometrics, Food Chem. 398 (2023) 133895. https://doi.org/10.1016/j.foodchem.2022.133895.

[133] J. Ning, M. Li, W. Chen, H. Zhao, J. Chen, M. Yang, X. Cao, X. Yue, Peptidomics as a tool to analyze endogenous peptides in milk and milk-related peptides, Food Biosci. 50 (2022) 102199. https://doi.org/10.1016/j.fbio.2022.102199.

[134] N. Zhao, M. Cheng, W. Lv, Y. Wu, D. Liu, X. Zhang, Peptides as potential biomarkers for authentication of mountain-cultivated ginseng and cultivated ginseng of different ages using UPLC-HRMS, J. Agric. Food Chem. 68 (2020) 2263–2275. https://doi.org/10.1021/acs.jafc.9b05568.

[135] O. Fiehn, Metabolomics - The link between genotypes and phenotypes, Plant Mol. Biol. 48 (2002) 155–171. https://doi.org/10.1023/A:1013713905833.

[136] M. Creydt, M. Fischer, Food metabolomics: Latest hardware—developments for nontargeted food authenticity and food safety testing, Electrophoresis. 43 (2022) 2334–2350. https://doi.org/10.1002/elps.202200126.

[137] J.-L. Zinia Zaukuu, E. Benes, G. Bázár, Z. Kovács, M. Fodor, Agricultural potentials of molecular spectroscopy and advances for food authentication: An overview, Processes. 10 (2022) 214. https://doi.org/10.3390/pr10020214.

[138] R. Pandiselvam, R. Kaavya, S.I. Martinez Monteagudo, V. Divya, S. Jain, A.C. Khanashyam, A. Kothakota, V.A. Prasath, S. V. Ramesh, N.U. Sruthi, M. Kumar, M.R. Manikantan, C.A. Kumar, A.M. Khaneghah, D. Cozzolino, Contemporary developments and emerging trends in the application of spectroscopy techniques: A particular reference to coconut (Cocos nucifera L.), Molecules. 27 (2022) 3250. https://doi.org/10.3390/molecules27103250.

[139] M. Keskin, A. Arslan, Y. Soysal, Y.E. Sekerli, N. Celiktas, Feasibility of a chromameter and chemometric techniques to discriminate pure and mixed organic and conventional red pepper powders: A pilot study, J. Food Process. Preserv. 46 (2022) e15846. https://doi.org/10.1111/jfpp.15846.

[140] R. Ribeiro de Souza, D.D. de Sousa Fernandes, P.H. Gonçalves Dias Diniz, Honey authentication in terms of its adulteration with sugar syrups using UV–Vis spectroscopy and one-class classifiers, Food Chem. 365 (2021) 130467. https://doi.org/10.1016/j.foodchem.2021.130467.

[141] M. Tomková, J. Sádecká, K. Hroboňová, Synchronous fluorescence spectroscopy for rapid classification of fruit spirits, Food Anal. Methods. 8 (2015) 1258–1267. https://doi.org/10.1007/s12161-014-0010-9.

[142] S. Medina, R. Perestrelo, P. Silva, J.A.M. Pereira, J.S. Câmara, Current trends and recent advances on food authenticity technologies and chemometric approaches, Trends Food Sci. Technol. 85 (2019) 163–176. https://doi.org/10.1016/j.tifs.2019.01.017.

[143] D. Cozzolino, Foodomics and infrared spectroscopy: From compounds to functionality, Curr. Opin. Food Sci. 4 (2015) 39–43. https://doi.org/10.1016/j.cofs.2015.05.003.

[144] J.B. Johnson, P.R. Thani, J.S. Mani, D. Cozzolino, M. Naiker, Mid-infrared spectroscopy for the rapid quantification of eucalyptus oil adulteration in Australian tea tree oil (Melaleuca alternifolia), Spectrochim. Acta - Part A Mol. Biomol. Spectrosc. 283 (2022) 121766. https://doi.org/10.1016/j.saa.2022.121766.

[145] J.D.S. Lopes, A.B.S. de Lima, R.R. da C. Cangussu, M.V. da Silva, S.P.B. Ferrão,

L.S. Santos, Application of spectroscopic techniques and chemometric methods to differentiate between true cinnamon and false cinnamon, Food Chem. 368 (2022) 130746. https://doi.org/10.1016/j.foodchem.2021.130746.

[146] M.J. Beriain, M. Lozano, J. Echeverría, M.T. Murillo-Arbizu, K. Insausti, M. Beruete, Tenderness of PGI "ternera de Navarra" beef samples determined by FTIR-MIR spectroscopy, Foods. 11 (2022) 3426. https://doi.org/10.3390/foods11213426.

[147] M.P. Totaro, G. Squeo, D. De Angelis, A. Pasqualone, F. Caponio, C. Summo, Application of NIR spectroscopy coupled with DD-SIMCA class modelling for the authentication of pork meat, J. Food Compos. Anal. 118 (2023) 105211. https://doi.org/10.1016/j.jfca.2023.105211.

[148] S. Shi, J. Feng, L. Yang, J. Xing, G. Pan, J. Tang, J. Wang, J. Liu, C. Cao, Y. Jiang, Combination of NIR spectroscopy and algorithms for rapid differentiation between one-year and two-year stored rice, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 291 (2023) 122343. https://doi.org/10.1016/j.saa.2023.122343.

[149] A. Drees, J. Brockelt, L. Cvancar, M. Fischer, Rapid determination of the shell content in cocoa products using FT-NIR spectroscopy and chemometrics, Talanta. 256 (2023) 124310. https://doi.org/10.1016/j.talanta.2023.124310.

[150] D.A. Magdas, C. Berghian-Grosan, Botanical honey recognition and quantitative mixture detection based on Raman spectroscopy and machine learning, Spectrochim. Acta - Part A Mol. Biomol. Spectrosc. 293 (2023) 122433. https://doi.org/10.1016/j.saa.2023.122433.

[151] D.A. Magdas, F. Guyon, I. Feher, S.C. Pinzaru, Wine discrimination based on chemometric analysis of untargeted markers using FT-Raman spectroscopy, Food Control. 85 (2018) 385–391. https://doi.org/10.1016/j.foodcont.2017.10.024.

[152] A. Arroyo-Cerezo, A.M. Jiménez-Carvelo, A. González-Casado, I. Ruisánchez, L. Cuadros-Rodríguez, The potential of the spatially offset Raman spectroscopy (SORS) for implementing rapid and non-invasive in-situ authentication methods of plastic-packaged commodity foods – Application to sliced cheeses, Food Control. 146 (2023) 109522. https://doi.org/10.1016/j.foodcont.2022.109522.

[153] D.I. Ellis, H. Muhamadali, S.A. Haughey, C.T. Elliott, R. Goodacre, Point-and-shoot: Rapid quantitative detection methods for on-site food fraud analysis - moving out of the laboratory and into the food supply chain, Anal. Methods. 7 (2015) 9401. https://doi.org/10.1039/c5ay02048d.

[154] P.Y. Bruice, Química Orgánica, 5th ed., Pearson Educación, Mexico, 2008.

[155] M. Gallo, P. Ferranti, The evolution of analytical chemistry methods in foodomics, J. Chromatogr. A. 1428 (2016) 3–15. https://doi.org/10.1016/j.chroma.2015.09.007.

[156] A.P. Sobolev, F. Thomas, J. Donarski, C. Ingallina, S. Circi, F.C. Marincola, D. Capitani, L. Mannina, Use of NMR applications to tackle future food fraud issues, Trends Food Sci. Technol. 91 (2019) 347–353. https://doi.org/10.1016/j.tifs.2019.07.035.

[157] Bruker, FoodScreener, (2023). https://www.bruker.com/en/products-and-solutions/mr/nmr-food-solutions/food-screener.html (accessed April 4, 2023).

[158] S. Esslinger, J. Riedl, C. Fauhl-Hassek, Potential and limitations of non-targeted fingerprinting for authentication of food in official control, Food Res. Int. 60 (2014) 189–204. https://doi.org/10.1016/j.foodres.2013.10.015.

[159] F. Flügge, T. Kerkow, P. Kowalski, J. Bornhöft, E. Seemann, M. Creydt, B. Schütze, U.L. Günther, Qualitative and quantitative food authentication of oregano using NGS and NMR with chemometrics, Food Control. 145 (2023) 109497. https://doi.org/10.1016/j.foodcont.2022.109497.

[160] F.I. Brigante, M.E. García, A. López Radcenco, G. Moyna, D.A. Wunderlin, M. V. Baroni, Identification of chia, flax and sesame seeds authenticity markers by NMR-based untargeted metabolomics and their validation in bakery products containing them, Food Chem. 387 (2022) 132925. https://doi.org/10.1016/j.foodchem.2022.132925.

[161] L. Haddad, J. Francis, T. Rizk, S. Akoka, G.S. Remaud, J. Bejjani, Cheese characterization and authentication through lipid biomarkers obtained by high-resolution 1H NMR profiling, Food Chem. 383 (2022) 132434. https://doi.org/10.1016/j.foodchem.2022.132434.

[162] A. Rivera-Pérez, R. Romero-González, A. Garrido Frenich, A metabolomics approach based on 1H NMR fingerprinting and chemometrics for quality control and geographical discrimination of black pepper, J. Food Compos. Anal. 105 (2022) 104235. https://doi.org/10.1016/j.jfca.2021.104235.

[163] Y. Gunning, K.S. Davies, E.K. Kemsley, Authentication of saffron using 60 MHz 1H NMR spectroscopy, Food Chem. 404 (2023) 134649. https://doi.org/10.1016/j.foodchem.2022.134649.

[164] J. Zaldarriaga Heredia, M. Wagner, F.C. Jofré, M. Savio, S.M. Azcarate, J.M. Camiña, An overview on multi-elemental profile integrated with chemometrics for food quality assessment: toward new challenges, Crit. Rev. Food Sci. Nutr. (2022) 1–21. https://doi.org/10.1080/10408398.2022.2055527.

[165] K.H. Laursen, J.K. Schjoerring, S.D. Kelly, S. Husted, Authentication of organically grown plants - advantages and limitations of atomic spectroscopy for multi-element and stable isotope analysis, TrAC - Trends Anal. Chem. 59 (2014) 73–82. https://doi.org/10.1016/j.trac.2014.04.008.

[166] E.C. Mazarakioti, A. Zotos, A.A. Thomatou, A. Kontogeorgos, A. Patakas, A. Ladavos, Inductively coupled plasma-mass spectrometry (ICP-MS), a useful tool in authenticity of agricultural products' and foods' origin, Foods. 11 (2022) 3705. https://doi.org/10.3390/foods11223705.

[167] P. Inaudi, A. Giacomino, M. Malandrino, C. La Gioia, E. Conca, T. Karak, O. Abollino, The inorganic component as a possible marker for quality and for authentication of the hazelnut's origin, Int. J. Environ. Res. Public Health. 17 (2020) 447. https://doi.org/10.3390/ijerph17020447.

[168] B. Quinn, P. McCarron, Y. Hong, N. Birse, D. Wu, C.T. Elliott, R. Ch, Elementomics combined with DD-SIMCA and K-NN to identify the geographical origin of rice samples from China, India, and Vietnam, Food Chem. 386 (2022) 132738. https://doi.org/10.1016/j.foodchem.2022.132738.

[169] M. Ghidotti, Y. Fiamegos, C. Dumitrascu, M.B. de la Calle, Use of elemental profiles to verify geographical origin and botanical variety of Spanish honeys with a protected denomination of origin, Food Chem. 342 (2021) 128350. https://doi.org/10.1016/j.foodchem.2020.128350.

[170] K. Katerinopoulou, A. Kontogeorgos, C.E. Salmas, A. Patakas, A. Ladavos, Geographical origin authentication of agri-food products: A review, Foods. 9 (2020) 489. https://doi.org/10.3390/foods9040489.

[171] F. Camin, M. Boner, L. Bontempo, C. Fauhl-Hassek, S.D. Kelly, J. Riedl, A. Rossmann, Stable isotope techniques for verifying the declared geographical origin of food in legal cases, Trends Food Sci. Technol. 61 (2017) 176–187. https://doi.org/10.1016/j.tifs.2016.12.007.

[172] C. Li, H. Dong, D. Luo, Y. Xian, X. Fu, Recent developments in application of stable isotope and multi-element analysis on geographical origin traceability of cereal grains, Food Anal. Methods. 9 (2016) 1512–1519. https://doi.org/10.1007/s12161-015-0328-y.

[173] I.-M. Chung, J.-K. Kim, M. Prabakaran, J.-H. Yang, S.-H. Kim, Authenticity of rice (Oryza sativa L.) geographical origin based on analysis of C, N, O and S stable isotope ratios: A preliminary case report in Korea, China and Philippine, J. Sci. Food Agric. 96 (2015) 2433–2439. https://doi.org/10.1002/jsfa.7363.

[174] M. Perini, S. Pianezze, L. Ziller, M. Ferrante, F. Ferella, S. Nisi, M. Foschi, A.A. D'Archivio, Stable isotope ratio analysis combined with inductively coupled plasma-mass spectrometry for geographical discrimination between Italian and foreign saffron, J. Mass Spectrom. 55 (2020) e4595. https://doi.org/10.1002/jms.4595.

[175] D.A. Magdas, F. Guyon, R. Puscas, A. Vigouroux, L. Gaillard, A. Dehelean, I. Feher, G. Cristea, Applications of emerging stable isotopes and elemental markers for geographical and varietal recognition of Romanian and French honeys, Food Chem. 334 (2021) 127599. https://doi.org/10.1016/j.foodchem.2020.127599.

[176] J.J. Roberts, D. Cozzolino, An overview on the application of chemometrics in food science and technology — An approach to quantitative data analysis, Food Anal. Methods. 9 (2016) 3258–3267. https://doi.org/10.1007/s12161-016-0574-7.

[177] C. Buvé, W. Saeys, M.A. Rasmussen, B. Neckebroeck, M. Hendrickx, T. Grauwet, A. Van Loey, Application of multivariate data analysis for food quality investigations: An example-based review, Food Res. Int. 151 (2022) 110878. https://doi.org/10.1016/j.foodres.2021.110878.

[178] D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte, L. Kaufman, Chemometrics: a textbook, Elsevier, Amsterdam, 1988.

[179] R. González-Domínguez, A. Sayago, Á. Fernández-Recamales, An overview on the

application of chemometrics tools in food authenticity and traceability, Foods. 11 (2022) 3940. https://doi.org/10.3390/foods11233940.

[180] S. Sentellas, J. Saurina, The role of chemometrics in food integrity and authenticity, in: O. Núñez, G. Campmajó (Eds.), Chromatogr. Relat. Sep. Tech. Food Integr. Authenticity. Vol. A Adv. Chromatogr. Tech., World Scientific Publishing Europe Ltd., 2021: pp. 167–200. https://doi.org/10.1142/9781786349958_0007.

[181] B. Khakimov, G. Gürdeniz, S.B. Engelse, Trends in the application of chemometrics to foodomics studies, Acta Aliment. 44 (2015) 4–31. https://doi.org/10.1556/AAlim.44.2015.1.1.

[182] R. Bro, A.K. Smilde, Principal component analysis, Anal. Methods. 6 (2014) 2812–2831. https://doi.org/10.1039/c3ay41907j.

[183] M. Esteki, J. Simal-Gandara, Z. Shahsavari, S. Zandbaaf, E. Dashtaki, Y. Vander Heyden, A review on the application of chromatographic methods, coupled to chemometrics, for food authentication, Food Control. 93 (2018) 165–182. https://doi.org/10.1016/j.foodcont.2018.06.015.

[184] Eigenvector Research Incorporated, Webinar Series, (2023). https://eigenvector.com/resources/webinars/ (accessed March 24, 2023).

[185] D. Granato, P. Putnik, D. Bursać Kovačević, J. Sousa Santos, V. Calado, R. Silva Rocha, A. Gomes Da Cruz, B. Jarvis, O. Ye Rodionova, A. Pomerantsev, Trends in chemometrics: Food authentication, microbiology, and effects of processing, Compr. Rev. Food Sci. Food Saf. 17 (2018) 663–677. https://doi.org/10.1111/1541-4337.12341.

[186] M. Bevilacqua, R. Bro, F. Marini, Å. Rinnan, M.A. Rasmussen, T. Skov, Recent chemometrics advances for foodomics, TrAC - Trends Anal. Chem. 96 (2017) 42–51. https://doi.org/10.1016/j.trac.2017.08.011.

[187] O.Y. Rodionova, P. Oliveri, A.L. Pomerantsev, Rigorous and compliant approaches to one-class classification, Chemom. Intell. Lab. Syst. 159 (2016) 89–96. https://doi.org/10.1016/j.chemolab.2016.10.002.

[188] A.A. Saleh, L. Weigang, A new variables selection and dimensionality reduction technique coupled with SIMCA method for the classification of text documents, in: Manag. Intellect. Cap. Innov. Sustain. Incl. Soc., 2015: pp. 583–591.

[189] O.Y. Rodionova, A. V. Titova, A.L. Pomerantsev, Discriminant analysis is an inappropriate method of authentication, TrAC - Trends Anal. Chem. 78 (2016) 17–22. https://doi.org/10.1016/j.trac.2016.01.010.

CHAPTER 2

METABOLOMIC PROFILING APPROACHES

2.1. INTRODUCTION

Separation techniques such as gas chromatography (GC) and liquid chromatography (LC) are fundamental in developing metabolomic profiling approaches for food authentication. These techniques allow the separation of the metabolites found in complex mixtures (*e.g.*, food samples) according to their size, charge, solubility, or polarity. Alongside the applied sample treatment, selecting the chromatographic technique, the chosen conditions, and the detection method strongly influences the number and type of detectable metabolites. Nevertheless, these techniques generally provide more sensitive and selective results than the spectroscopic techniques detailed in Section 1.2.4.

Independently of the chosen chromatographic technique, two different profiling-based acquisition strategies have been commonly proposed in food authentication: targeted and suspected analysis. Both focus on determining a given group of known selected metabolites, generally sharing chemical similarities and even belonging to the same family. On the one hand, targeted analysis allows accurate compound quantification and confirmation by using the proper analytical reference standards. However, this approach is often not possible due to the lack of commercial standards, which is especially noticeable in the case of secondary metabolites because of their complexity and diversity. Alternatively, on the other hand, suspected analysis collects instrumental data related to specific compounds for which no analytical standards are usually available. For this purpose, metabolite inclusion lists are built based on knowledge in the scientific literature or experimental databases/libraries. Hence, although semi-quantification (*e.g.*, using the analytical reference standard of a structurally similar compound) or relative quantification are sometimes carried out, the suspected analysis provides more qualitative than quantitative benefits [1,2]. Therefore, although some authors comprise suspected analysis within the non-targeted analysis, in this thesis, it is considered more appropriate to include it within the profiling approach since it looks for differences between specific compounds rather than between all the detected metabolites.

In this context, GC has been commonly proposed for both targeted and suspected strategies to detect food fraud practices. This technique separates the compounds in the analysed sample after vaporisation. Thus, the analytes' migration through the GC column, under high pressure using an inert gas (usually He or N_2), depends on their interaction with the stationary phase. Therefore, GC —with a flame ionisation detector (GC-FID) or coupled to mass spectrometry (GC−MS)— is especially suitable for determining and detecting volatile and thermally stable compounds such as aldehydes, alcohols, or organic acids. Moreover, medium-volatile or thermolabile compounds can be determined by this technique after chemical derivatisation. For instance, amino acids, fatty acids, or carbohydrates need this previous step to be determined by GC [3,4].

Furthermore, focusing on GC−MS, 70 eV electron ionisation (EI) is commonly chosen in metabolomic profiling studies as the ionisation method. Its reproducibility allows compound identification by comparison of the experimental mass spectral fragmentation pattern to spectral libraries, such as of the National Institute of Standards and Technology (NIST) or Wiley, being especially interesting for suspected and non-targeted screening. Therefore, although some compounds present weak or absent molecular ions at 70 eV EI due to high fragmentation, it is still preferred over well-established soft ionisation methods —such as chemical ionisation (CI)— due to its compatibility with library-based identification [5].

Fatty acids (FAs) have been the most studied lipid compounds for food authentication and classification purposes by targeted GC methodologies. Their detection through GC-FID or GC−MS is well-established after a widespread sample treatment consisting of three main steps: 1) FAs extraction with a non-polar solvent (*e.g.*, *n*-hexane or petroleum ether), 2) compound derivatisation consisting of acid or basic methyl esterification, and 3) separation of the resulting fatty acid methyl esters (FAMEs) from the aqueous phase [6]. Thus, since FAs are abundantly found in edible oils, their profile has been evaluated as a potential chemical marker to discriminate cultivars of processed olives [7] or to detect VOO adulteration with other vegetable

oils [8]. Moreover, other applications involving these lipids' determination in different food products, such as almonds or green coffee beans, have also been developed to assess their geographical origin [9,10].

Moreover, profiling approaches focusing on specific volatile organic compounds (VOCs) have also been proposed in some food fraud applications. For instance, samples belonging to the *Pasta di Gragnano* PGI were authenticated through the targeted GC−MS analysis of 20 flavour compounds (derived from the Maillard reaction and lipid oxidation). In this case, all the compounds were identified by comparison of the experimental retention times and mass spectra with those of the reference standards [11]. Instead, the combination of targeted and suspected analysis by gas chromatography coupled to high-resolution mass spectrometry (GC−HRMS) of 36 hydrophilic compounds was proposed to address the geographical origin classification of adzuki beans (*Vigna angularis*). Thus, spectra information was obtained from both reference standards and in-house databases [12]. Nevertheless, it is noteworthy to mention that because of the great potential in compound identification of GC−MS, most of its current methods developed for food authentication or characterisation purposes are based on non-targeted analysis. In this line, some examples of the application of this approach are given in Section 3.1.

Alternatively, LC allows the separation of a mixture of analytes according to their interactions with the LC column stationary phase and the mobile liquid phase. Thus, while GC focuses on the volatile metabolome, LC offers a more comprehensive metabolomic coverage (higher detectability) comprising non-polar and polar compounds. In this line, liquid chromatography —with ultraviolet detection (LC-UV), fluorescent detection (LC-FLD), electrochemical detection (LC-ECD), or LC−MS— has been widely employed in the determination of metabolite profiles. Moreover, the recent development of ultra-high-performance liquid chromatography (UHPLC), which implies using sub-2 μm particle packed columns or sub-3 μm superficially porous particle columns, has substantially diminished the separation

time compared to high-performance liquid chromatography (HPLC), while keeping or improving the chromatographic resolution [13].

LC−MS has been the most employed technique in metabolomic targeted and non-targeted approaches (in this last case, the reader is referred to Chapter 3 of the thesis for further information). In this line, using MS as the detection system improves the method's sensitivity (noise reduction) and specificity, being adequate for both qualitative and quantitative studies. Hence, while LRMS mass analysers —mainly triple quadrupole (QqQ) and ion trap (IT)— are especially suitable for targeted analysis, HRMS ones —Orbitrap- and TOF-based mass analysers— present excellent capabilities for suspected analysis. For instance, HRMS mass analysers provide accurate mass measurements with high resolution; thus, comprehensive information about the exact molecular mass and the elemental composition can be obtained. Besides, they allow the elucidation of the chemical structure of a particular compound by studying the corresponding fragmentation pattern. Thereby, tentative compound identification or annotation can be carried out without analytical reference standards. Finally, these instruments usually acquire data under the full-scan mode, which increases the number of investigated analytes and enables retrospective analysis [2,14].

Lipids and phenolic compounds have been the most employed families of metabolites to assess food authentication through profiling LC methodologies. In general, lipids have been used as chemical markers in applications involving animal-derived products, fats, or oils, whereas polyphenols have been applied to plant-related foodstuffs. Thus, focusing on lipids, the triacylglycerols (TAGs) profile has been widely studied in food products for authentication purposes. The thermolability of these compounds, formed by three FAs linked to one glycerol, impedes their determination by GC. Therefore, they are typically determined by LC after extraction with a non-polar solvent [6]. In this line, lard content in cocoa butter was evaluated through the targeted determination of 16 TAGs, comprising unsaturated and saturated structures, by liquid chromatography with a refractive index detector (LC-RID). The

obtained results allowed the detection of admixtures ranging from 1% to 30% of lard [15]. Similarly, the content of 11 TAGs in different edible oils was assessed by relative quantification after liquid chromatography with charged aerosol detection (LC-CAD) analysis. Then, PCA allowed the distinction of EVOO samples in front of the other vegetable oils and their blends [16].

In addition, the suspected analysis of lipid compounds is encompassed within the '*lipidomics*' term. Thus, it is generally carried out by LC−HRMS, focusing on compounds belonging to one or more classes of lipids that are included in in-house or online databases. Currently, the LipidMaps online database is the most commonly used in this approach since it contains information related to 47,877 unique lipid structures [17]. For example, the discrimination of TSG hay milk against maize or grass silage milk was achieved using the profile composed of 232 suspected TAGs LC−HRMS signals. Among them, 14 TAG molecular species showed significant differences between the sample groups under study. Thus, after their DDA analysis, the discriminant TAGs were tentatively annotated based on their exact mass and fragmentation [18]. Besides, following a similar procedure, 53 TAGs were tentatively identified in milk by LC−HRMS to distinguish organic and conventional production. As a result, 10 TAG compounds were significantly lower in organic milk, while 11 were significantly higher [19].

Furthermore, as previously mentioned, phenolic and polyphenolic compounds are one of the main classes investigated by LC for food characterisation, classification, and authentication. These compounds are secondary metabolites with an aromatic structural skeleton. Moreover, they are ubiquitously spread through the plant kingdom and some present bioactive activity (*e.g.*, antioxidant or anticancer). Besides, depending on their chemical structure —the number of phenol rings and how they are bound—, phenolic compounds are classified into four main classes: phenolic acids, flavonoids, lignans, and stilbenes. Particularly, phenolic acids (*i.e.*, hydroxybenzoic and hydroxycinnamic acids) and flavonoids (*i.e.*, flavonols,

flavones, isoflavones, flavanones, anthocyanins, and flavanols) account for almost 90% of the total extractable phenolic and polyphenolic content in the diet [20].

Since this chapter will describe the application of a phenolic liquid chromatography coupled to tandem mass spectrometry (LC−MS/MS) profiling method for the classification of paprika samples according to their geographical origin (Publication I), especial attention to developed methodologies has been paid. In this context, Table 2.1 presents a compilation of recent studies in which phenolic and polyphenolic LC profiles have been used to address different plant-related foodstuff authentication issues. Hence, as reported in Table 2.1, soluble phenolic and polyphenolic compounds are commonly investigated as chemical markers instead of non-extractable ones (polymeric polyphenols or individual polyphenols linked to macromolecules), whose determination in this field is limited [21]. Therefore, considering that these compounds are more soluble in polar organic solvents than water, their extraction is usually carried out through solid-liquid extraction (SLE) or liquid-liquid extraction (LLE) procedures, using extracting mixtures of water with organic solvents such as methanol, ethanol, or acetone. Moreover, in most cases, this extraction step is enhanced by applying ultrasound waves to the sample [20].

Moreover, as shown in Table 2.1, reversed-phase chromatography predominates in the chromatographic separation of polyphenolic compounds over other options, such as hydrophilic interaction chromatography (HILIC). In this context, it usually relies on using C_{18} stationary phase columns. However, in some applications, other reversed-phase stationary phases have been proposed. For instance, Rivera-Pérez *et al*. [22] employed a reversed-phase pentafluorophenyl (PFP) column to separate phenolic compounds within a suspected LC−HRMS analysis to authenticate thyme. In any case, the mobile phase used in the reversed-phase separation of polyphenols is commonly composed of water and methanol or acetonitrile as the organic solvent. Besides, a low amount of acid (*i.e.*, formic or acetic acids) is commonly added to the mobile phase since it improves the chromatographic peak shapes, resulting in better chromatographic resolution and sensitivity.

Regarding the detection system, it is strongly dependent on the profiling approach used. Thus, similarly to the previously described TAGs determination, while polyphenolic targeted analysis is generally performed using LC-UV or LC−MS, suspected analysis is exclusively carried out by LC−HRMS.

Therefore, as detailed in Table 2.1, several targeted LC-UV methods have been developed for different food authentication purposes. In this context, phenolic acids display absorption bands in the 300 – 400 nm range, whereas more complex polyphenolics structures —such as stilbenes, flavanols, flavonols, or anthocyanins— also do it in the visible range [23]. For instance, González-Domínguez *et al*. [24] evaluated the effect of cultivar and cultivation conditions in strawberry samples by combining different determinations: sugars, organic acids, phenolic compounds, and mineral elements. In the case of targeted phenolics, LC-UV was used for their detection at different wavelengths: 260 nm for ellagic acid and derivatives, 280 nm for benzoic acids and flavanols, 320 nm for cinnamic acids, 360 nm for flavonols, and 520 nm for anthocyanins. As a result, 13 phenolic compounds were quantified and, combined with the other determined components, allowed clear chemometric discrimination according to the sample cultivar. Similarly, Kalogiouri *et al*. [25] used the profile of 18 phenolic compounds to address the geographical classification of walnut samples. In this case, the phenolic compounds were detected by LC-UV, using a diode array detector (DAD), which allowed quantitating the analytes at their maximum absorption wavelengths. Therefore, for instance, *p*-coumaric acid (hydroxycinnamic acid) was detected at 270 nm and myricetin (flavonol) at 370 nm.

Table 2.1. Compilation of some LC profiling methodologies, focusing on phenolic and polyphenolic compounds as chemical markers, to address different food authentication issues.

Sample	Compounds and extraction	Chromatographic separation	Detection system	Data analysis	Ref.
Cultivar/botanical origin					
Edible flowers	18 phenolics (phenolic acids and flavonoids) SLE with methanol:water (50:50, *v/v*)	LC Agilent Poroshell EC-C_{18} column (150 × 4.6 mm, 2.7 µm) Solvents: A) water with 0.1% phosphoric acid (*v/v*) and B) acetonitrile	UV λ_{exc}: 217 and 327 nm	Heat map, HCA, and PLS-DA	[26]
Honey	18 phenolics (phenolic acids and flavonoids) SPE with Chromabond C_{18} SPE column and elution with methanol	LC Nucleosil 100-5 C_{18} column (250 × 4.6 mm, 5.0 µm) Solvents: A) water with 2% acetic acid (*v/v*) and B) water:acetonitrile (50:50, *v/v*) with 0.5% acetic acid (*v/v*)	UV λ_{exc}: 280, 320, and 350 nm	PCA	[27]
Honey	5 phenolics (phenolic acid and flavonoids) SLE with ethyl acetate	LC Eclipse XDB C_{18} column (150 × 4.5 mm, 5.0 µm) Solvents: A) water with 2% acetic acid (*v/v*) and B) acetonitrile	UV λ_{exc}: 280 and 330 nm	LDA	[28]
Lemon and Persian lime juices	11 phenolic derivatives (coumarins and psoralens) LLE with ethyl acetate	LC Phenomenex Kinetex C_{18} column (150 × 4.6 mm, 2.6 µm) Solvents: A) water:acetonitrile:*t*BME (85:13:2, *v/v/v*) and B) acetonitrile:methanol:*t*BME (65:30:5, *v/v/v*)	UV λ_{exc}: 190 to 370 nm LRMS (IT) Full-scan (*m/z* 100 – 1500) and MS^2 and MS^3 (*m/z* 50 – 400) ESI (+)	PCA	[29]

Fruit extracts	26 phenolics (phenolic acids, flavonoids, and derivatives) SLE with acetone:water (70:30, *v/v*) with 0.1% hydrochloric acid	LC Kinetex C_{18} column (100 × 4.6 mm, 2.6 µm) Solvents: A) water with 0.1% formic acid (*v/v*) and B) methanol	LRMS (QqQ) MRM H-ESI (-)	PCA	[30]
Chia, sesame, and flax seeds bakery products	44 phenolics (phenolic acids, flavonoids, and lignans) SLE with water:methanol (50:50, *v/v*)	LC Phenomenex Luna C_{18} column (250 × 5.0 mm, 4.6 µm) Solvents: A) water and B) methanol, both with 0.5% formic acid (*v/v*)	UV λ_{exc}: 200 to 600 nm HRMS (Q-TOF) Full-scan (*m/z* 80 – 1500) and DDA H-ESI (±)	PCA, CA, and DA	[31]
Turmeric and curry	53 phenolics (phenolic acids, flavonoids, stilbenes, and others) and suspected phenolics (curcuminoids) SLE with DMSO	LC Ascentis Express C_{18} column (150 × 2.1 mm, 2.7 µm) Solvents: A) water and B) acetonitrile, both with 0.1% formic acid (*v/v*)	HRMS (Q-Orbitrap) Full-scan (*m/z* 100 – 1500) and DDA H-ESI (-)	PCA and PLS-DA	[32]
Cranberry-based extracts	53 phenolics (phenolic acids, flavonoids, stilbenes, and others) SLE with acetone:water (70:30, *v/v*) with 0.1% hydrochloric acid	LC Ascentis Express C_{18} column (150 × 2.1 mm, 2.7 µm) Solvents: A) water and B) acetonitrile, both with 0.1% formic acid (*v/v*)	HRMS (Q-Orbitrap) Full-scan (*m/z* 100 – 1500) and DDA H-ESI (-)	PCA and PLS	[33]
Berry fruit juices	43 phenolics (phenolic acids and flavonoids) Dilution with water	LC Phenomenex C_{18} column (100 × 2.1 mm, 2.6 µm) Solvents: A) water and B) methanol, both with 0.1% formic acid (*v/v*)	HRMS (Q-TOF) Full-scan (*m/z* 50 – 1000) and DDA H-ESI (+)	OPLS-DA	[34]
EVOO	70 phenolics (phenolic acids, flavonoids, lignans, others) and suspected phenolics LLE with methanol:water (80:20, *v/v*)	LC Thermo Fisher Scientific Acclaim C_{18} column (100 × 2.1 mm, 2.2 µm) Solvents: A) water:methanol (90:10, *v/v*) and B) methanol, both with 5 mM ammonium acetate	HRMS (Q-TOF) Full-scan (*m/z* 50 – 1000), DIA, and DDA H-ESI (-)	ANOVA	[35]

Berry seeds	Suspected phenolics SLE with methanol:water (70:30, *v/v*) with 0.1% hydrochloric acid (*v/v*)	LC Syncronis C_{18} column (100 × 2.1 mm, 1.7 µm) Solvents: A) water and B) acetonitrile, both with 0.1% formic acid (*v/v*)	HRMS (LTQ-Orbitrap) Full-scan (*m/z* 100 – 1500) and DDA H-ESI (-)	-	[36]
Geographical origin					
Walnuts	18 phenolics (phenolic acids and flavonoids) SLE with methanol:water (60:40, *v/v*) with 0.05% TFA (*v/v*)	LC C_{18} Fortis UniverSil column (250 × 4.6 mm, 5.0 µm) Solvents: A) water with 1% acetic acid (*v/v*) and B) acetonitrile	UV λ_{exc}: 250 to 400 nm	PLS-DA	[25]
Green coffee beans	11 chlorogenic acids SLE with methanol:water (95:5, *v/v*)	LC BEH C_{18} column Solvents: A) water with 1% formic acid (*v/v*) and B) acetonitrile	UV λ_{exc}: 200 to 400 nm HRMS (Q-TOF) H-ESI (-)	PCA and LDA	[37]
Harvesting time					
Hangbaiju (*Chrysanthemum morifolium* Ramat.)	19 flavonoids and suspected flavonoids SLE with methanol:water (50:50, *v/v*)	LC Phenomenex C_{18} column (150 × 2.0 mm, 3.0 µm) Solvents: A) water and B) acetonitrile, both with 0.1% formic acid (*v/v*)	HRMS (Q-TOF) Full-scan (*m/z* 100 – 1250) and DIA H-ESI (-)	PCA and OPLS-DA	[38]
Processing					
Zingiberis Rhizoma	7 gingerols and suspected gingerols SLE with methanol:water (80:20, *v/v*)	LC Phenomenex Kinetex C_{18} column (100 × 3.0 mm, 2.6 µm) Solvents: A) water with 0.1% formic acid (*v/v*) and B) acetonitrile	HRMS (Q-TOF) Full-scan (*m/z* 100 – 800) and DIA H-ESI (+)	PCA and OPLS-DA	[39]

Adulteration and geographical origin					
Saffron	Suspected phenolics (Phenol-Explorer) SLE with methanol:water (80:20, *v/v*) with 1% formic acid (*v/v*)	LC Knauer Blue Orchid C_{18} column (100 × 2.0 mm, 1.8 µm) Solvents: A) water and B) methanol, both with 0.1% formic acid (*v/v*) and ammonium formate 5 mM	HRMS (Q-TOF) Full-scan (*m/z* 100 – 1000) H-ESI (+)	HCA and OPLS-DA	[40]
Cultivar/botanical origin and harvesting time					
EVOO	19 phenolics (phenolic acids, flavonoids, lignans, and derivatives) LLE with methanol:water (70:30, *v/v*)	LC Phenomenex Luna C_{18} column (150 × 2.0 mm, 3.0 µm) Solvents: A) water and B) acetonitrile, both with 0.1% formic acid (*v/v*)	HRMS (Q-TOF) Full-scan (*m/z* 50 – 1200) and MS/MS	PCA	[41]
Cultivar/botanical origin and processing					
Strawberry and blueberry fruits and jams	36 phenolics (phenolic acids and flavonoids) SLE with ethanol:water (70:30, *v/v*) with 1.5% hydrochloric acid (*v/v*)	LC Phenomenex Synergi Polar-RP C_{18} column (250 □ 4.6 mm, 4.0 µm) Solvents: A) water and B) methanol, both with 0.1% formic acid (*v/v*)	LRMS (QqQ) Dynamic-MRM H-ESI (±)	-	[42]
Geographical origin and processing					
Thyme	Suspected phenolics (Phenol-Explorer) SLE with methanol:water (80:20, *v/v*) with 0.1% formic acid (*v/v*)	LC Agilent Poroshell PFP column (100 × 2.1 mm, 1.9 µm) Solvents: A) water and B) acetonitrile, both with 0.1% formic acid (*v/v*)	HRMS (Q-TOF) Full-scan (*m/z* 100 – 1200) and DDA H-ESI (+)	HCA, PLS-DA, and OPLS-DA	[22]

Geographical origin and variety					
Paprika	53 phenolics (phenolic acids, flavonoids, stilbenes, and others) SLE with water:acetonitrile (20:80, *v/v*)	LC Ascentis Express C_{18} column (150 × 2.1 mm, 2.7 µm) Solvents: A) water and B) acetonitrile, both with 0.1% formic acid (*v/v*)	HRMS (Q-Orbitrap) Full-scan (*m/z* 100 – 1500) and DDA H-ESI (-)	PCA and PLS-DA	[43]
Adulteration, geographical origin, and variety					
EVOO	Suspected phenolics (Phenol-Explorer) LLE with methanol:water (80:20, *v/v*) with 0.1% formic acid (*v/v*)	LC Agilent Zorbax Eclipse plus C_{18} column (50 × 2.1 mm, 1.8 µm) Solvents: A) water and B) acetonitrile	HRMS (Q-TOF) Full-scan (*m/z* 100 – 1200) H-ESI (+)	OPLS-DA and ANN	[44]
Cultivar/botanical origin, harvesting time, and production system					
Strawberry	13 phenolics (coloured[a]: phenolic acids and flavonoids / colourless[b]: anthocyanins) SLE with methanol	LC Ultrabase C_{18} column (100 × 2.5 mm, 4.6 µm) Solvents[a]: A) water:methanol (95:5, *v/v*) and B) methanol, both with 2% acetic acid (*v/v*) Solvents[b]: A) water with 10% formic acid (*v/v*) and B) methanol	UV λ_{exc}[a]: 260, 280, 320, and 360 nm λ_{exc}[b]: 520 nm	PCA, SIMCA, LDA, and PLS-DA	[24]

Cluster analysis (CA); dimethyl sulfoxide (DMSO), electrospray ionisation (ESI), heated-electrospray ionisation (H-ESI), linear trap quadrupole-Orbitrap (LTQ-Orbitrap), orthogonal partial least squares regression-discriminant analysis (OPLS-DA), quadrupole-Orbitrap (Q-Orbitrap), quadrupole-time-of-flight (Q-TOF), solid-phase extraction (SPE).

Instead, targeted analysis of phenolic and polyphenolic compounds can also be done by LC−MS. In this context, as reflected in Table 2.1, heated-electrospray ionisation (H-ESI) has been established as the most common option for the ionisation step of these compounds, especially in the negative mode. Moreover, several targeted LC−MS methodologies have been developed using low-resolution mass spectrometers. Thus, while the MRM mode is typically applied in QqQ instruments, full-scan or product-ion scan are employed in IT ones. Therefore, for instance, the quantification of 26 phenolic compounds in fruit extracts was carried out through LC−MS/MS, using H-ESI in the negative mode and a QqQ as the mass analyser. The obtained phenolic profile was then subjected to PCA to discriminate cranberry-based extracts from grape ones [30]. Alternatively, positive H-ESI ionisation can improve the ionisation efficiency of specific phenolic classes. For example, Mustafa *et al*. [42] quantified 36 phenolic compounds to evaluate significant differences between blueberry and strawberry and their corresponding jams. Hence, compound detection was carried out using a QqQ, and negative or positive H-ESI ionisation was chosen depending on the phenolic class. In this line, nine compounds —six anthocyanins, two flavonols, and one flavanone— were detected under the positive ionisation mode. Moreover, similarly, the identification of targeted coumarins and psoralens, which were previously quantified through LC-UV to authenticate lemon samples, was confirmed by LC−MS through MS^1, MS^2, and MS^3 data acquired in the positive mode by the IT mass analyser [29].

Furthermore, targeted analysis has also been carried out through LC−HRMS, which allows the monitorisation of more compounds than LC−LRMS through the full-scan mode. For instance, 53 different targeted phenolic compounds, whose analytical reference standards were available, were determined by LC−HRMS in fruit-based extracts. In this case, the obtained phenolic profile allowed the authentication of cranberry-based extracts through PCA and PLS [33]. Moreover, LC−HRMS has also allowed the suspected analysis of these compounds, often in combination with targeted analysis. In this context, the suspected analysis of phenolic compounds is

frequently based on in-house databases built according to the existing literature data or the Phenol-Explorer database, an online tool containing information regarding phenolic and polyphenolic compounds in food products [45]. For instance, the phenolic LC−HRMS profile comprising targeted phenolic compounds and suspected curcuminoids was employed to assess the authenticity of turmeric and curry samples by PLS-DA [32]. In contrast, the *Tagiasca Ligure* EVOO authentication was addressed by means of a suspected LC−HRMS analysis using the entire Phenol-Explorer database. Besides, in this case, sterol data from the LipidMaps database was also combined in the subsequent chemometric analysis [44].

Aside from lipid and phenolic compounds, other metabolite families have also been exploited to address specific food fraud cases. In this line, Section 2.2 of this chapter contains the determination of the capsaicinoid and carotenoid profile through LC−HRMS to classify paprika samples according to their geographical origin and variety (Publication II). In this context, as shown in Table 2.2, some studies have already focused on these compounds for authentication purposes.

Capsaicinoids are the compounds responsible for the spicy flavour of red peppers. Thus, since red pepper and its derived products are commonly used as food additives worldwide, LC−MS/MS detection of targeted capsaicinoids was proposed to assess the authentication of vegetable oils in front of waste cooking oil [46,47]. Moreover, Barbosa *et al*. [43] carried out the suspected LC−HRMS analysis of several capsaicinoids and capsinoids (combined with suspected phenolic analysis) in paprika samples. As a result, these compounds allowed the discrimination of paprika varieties for all the geographical origins under study. Furthermore, although these studies have proposed LC−MS-based methods to detect these compounds, their determination can also be performed by LC-UV with a maximum absorption level at 280 nm [48].

Table 2.2. Compilation of some LC profiling methodologies, focusing on capsaicinoids and carotenoids as chemical markers, to address different food authentication issues.

Sample	Compounds and extraction	Chromatographic separation	Detection system	Data analysis	Ref.
CAPSAICINOIDS					
Adulteration					
Edible and crude vegetable oils	3 capsaicinoids LLE with methanol:water (44:56, *v/v*), defatting step with hexane, and SPE with C_{18} cartridges and eluting with methanol:water (80:20, *v/v*)	LC Agilent Zorbax Eclipse Plus C_{18} column (50 × 2.1 mm, 1.8 µm) Solvents: A) water and B) acetonitrile, both with 0.1% formic acid (*v/v*)	LRMS (QqQ) MRM ESI (+)	-	[47]
Vegetable oils	2 capsaicinoids LLE with methanol, IAC, and elution with methanol	LC Hypersil Gold C_{18} column (100 × 2.1 mm, 3.0 µm) Solvents: A) water with 0.1% formic acid (*v/v*) and B) acetonitrile	LRMS (QqQ) MRM ESI (+)	-	[46]
Geographical origin and variety					
Paprika	Suspected capsaicinoids SLE with water:acetonitrile (20:80, *v/v*)	LC Ascentis Express C_{18} column (150 □ 2.1 mm, 2.7 µm) Solvents: A) water and B) acetonitrile, both with 0.1% formic acid (*v/v*)	HRMS (Q-Orbitrap) Full-scan (*m/z* 100 – 1500) and DDA H-ESI (-)	PCA and PLS-DA	[43]

CAROTENOIDS					
Adulteration					
Sea buckthorn oil food supplement	3 carotenoids (lutein, lycopene, and β-carotene) LLE with acetone:ethanol (40:60, *v/v*)	LC Agilent Poroshell 120 column (100 × 2.1 mm, 2.7 µm) Solvents: A) water and B) acetonitrile	UV λ_{exc}: 450 nm	-	[49]
Cultivar/botanical origin					
Olive oil	3 carotenoids LLE with hexane and SPE with C_{18} cartridges eluting with acetone	LC Waters Spherisorb ODS-2 C_{18} reversed-phase column (250 × 4.6 mm, 5.0 µm) Solvents: A) 0.05 M tetrabutylammonium and 1 M ammonium acetate aqueous solution:methanol (2:8, *v/v*) and B) acetone:methanol (1:1, *v/v*)	UV λ_{exc}: 430 nm	PCA	[50]
Geographical origin					
Saffron	Suspected carotenoids and flavonoids SLE with water:methanol (50:50, *v/v*)	LC Phenomenex Synergi C_{18} column (250 × 4.6 mm, 4.0 µm) Phenomenex Kinetex C_{18} column (250 × 4.6 mm, 5.0 µm) Solvents: A) water and B) acetonitrile	UV λ_{exc}: 250, 310, and 441 nm	LDA	[51]
EVOO	SPE with Supelclean™ LC-Si SPE cartridges [a]Chlorophylls and xanthophylls (4 carotenoids) Elution with acetone [b]β-carotene (carotenoid) Elution with hexane, saponification with 10% KOH (*v/v*) in ethanol, and cleaning with water	LC Accucore C_{18} column (100 × 2.1 mm, 2.6 µm) Solvents[a]: A) water, B) methanol, C) acetonitrile, and D) acetone Solvents[b]: A) acetonitrile and B) acetone	HRMS (Q-Orbitrap) Full-scan (*m/z* 200 – 1000) and AIF APCI (+)	PCA	[52]

Production system						
Hen eggs	9 targeted carotenoids and suspected carotenoids LLE with hexane	LC Agilent Zorbax Eclipse XDB-C_{18} column (150 × 4.6 mm) Solvents: A) methanol:water (90:10, *v/v*) and B) acetonitrile:2-propanol (63:37, *v/v*)	UV λ_{exc}: 445 nm	PCA and kNN		[53]

All-ion fragmentation (AIF), atmospheric pressure chemical ionisation (APCI).

Instead, carotenoids are natural pigments strongly involved in the final colour of several food products. As reported in Table 2.2, LC-UV has been extensively used to determine the carotenoid profile in different food matrices. Thus, for instance, van Ruth *et al*. [53] carried out the targeted and suspected LC-UV analysis of hen eggs, focusing on nine specific carotenoids and some reported in the literature, respectively. Since chickens do not produce carotenoids naturally, their content in hen eggs comes from hen feed, which depends on the production system (*i.e.*, organic or conventional). In this line, the carotenoid profile allowed excellent chemometric sample classification according to the production system. Moreover, D'Archivio *et al*. [51] tentatively identified several carotenoid-related compounds in saffron using LC-UV at different wavelengths. Subsequently, the obtained profile was subjected to LDA for sample discrimination according to geographical origin. Similarly, Arrizabalaga-Larrañaga *et al*. [52] evaluated EVOOs' carotenoid profile and other pigments (*i.e.*, chlorophylls) to evaluate through PCA potential sample trends linked to their geographical origin. In this case, five targeted carotenoids were determined by LC−HRMS, using atmospheric pressure chemical ionisation (APCI) in the positive mode as the ionisation source.

Finally, it is noteworthy to mention that capillary electrophoresis (CE) has also been used in some specific applications [54]. CE separates charged chemical analytes based on their migration through a silica capillary, filled with a background electrolyte and under a high electric field strength. However, it still presents some drawbacks compared to GC and LC (*i.e.*, low sensitivity and robustness and relatively complex coupling to MS) that have limited their use in these metabolomic studies [55].

2.2. RESULTS

This chapter includes two scientific studies in which metabolomic profiling approaches, based on targeted LC−MS methods, were proposed to address the geographical origin authenticity of paprika samples by multivariate analysis. On the one hand, 36 phenolic compounds determined by LC−MS/MS were evaluated as chemical markers to address this issue. On the other hand, the capsaicinoid and carotenoid profile acquired by LC−HRMS, using APCI as the ionisation source, was proposed with the same purpose.

The resulting scientific articles of these studies are herein presented as Publications I and II.

Publication I: Scientific article

Determination of phenolic compounds in paprika by ultrahigh performance liquid chromatography-tandem mass spectrometry: Application to product designation of origin authentication by chemometrics.

Barbosa, S.; **Campmajó, G.**; Saurina, J.; Puignou, L.; Núñez, O.

Journal of Agricultural and Food Chemistry. 2020, 68, 591.

Publication II: Scientific article

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

Arrizabalaga-Larrañaga, A.; **Campmajó, G.**; Saurina, J.; Núñez, O.; Santos, F. J.; Moyano, E.

LWT. 2021, 139, 110533.

2.2.1. PUBLICATION I

Determination of phenolic compounds in paprika by ultrahigh performance liquid chromatography-tandem mass spectrometry: Application to product designation of origin authentication by chemometrics.

Barbosa, S.; Campmajó, G.; Saurina, J.; Puignou, L.; Núñez, O.

Journal of Agricultural and Food Chemistry. 2020, 68, 591.

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

pubs.acs.org/JAFC Article

Determination of Phenolic Compounds in Paprika by Ultrahigh Performance Liquid Chromatography−Tandem Mass Spectrometry: Application to Product Designation of Origin Authentication by Chemometrics

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

Cite This: *J. Agric. Food Chem.* 2020, 68, 591−602 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

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

KEYWORDS: *Paprika, UHPLC-ESI-MS/MS, Phenolic compounds, Food classification, Principal component analysis (PCA), Partial least squares regression-discriminant analysis (PLS-DA), Food authentication*

INTRODUCTION

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

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

Food manufacturers, as well as the public in general, are increasingly concerned about food quality attributes, and therefore, the demand for food products of a specific geographical origin grows. Within this context and with the aim of preserving the reputation of the products and supporting good practices in rural and agricultural activities, the European legislation has established several quality parameters related to the protection of geographical indications and appellations of origin of agricultural and food products (Council Regulation, EEC No. 510/2006[21]): Protected Designation of Origin (PDO) that links the products with the defined geographical area where

Received: September 25, 2019
Revised: December 19, 2019
Accepted: December 20, 2019
Published: December 20, 2019

https://dx.doi.org/10.1021/acs.jafc.9b06054
J. Agric. Food Chem. 2020, 68, 591−602

they are produced; Protected Geographical Indication (PGI) that links products to a geographical area where at least one step of production occurred; and Traditional Specialties Guaranteed (TSG) that protects traditional production methods.[22]

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

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

In this work, an ultrahigh performance liquid chromatography-electrospray-tandem mass spectrometry (UHPLC-ESI-MS/MS) method using a triple quadrupole (QqQ) analyzer has been developed for the determination and quantification of 36 phenolic and polyphenolic compounds in paprika and subsequent characterization, classification, and authentication of paprika samples by multivariate chemometric methodologies. Chromatographic and electrospray ion source conditions were optimized, and the method performance was established by determining quality parameters such as linearity, limits of detection, limits of quantitation, run-to-run and day-to-day precisions, and trueness. 111 paprika samples belonging to La Vera PDO and Murcia PDO (Spain) and to Czech Republic were analyzed with the proposed methodology after applying a simple extraction method using acetonitrile/water (80:20 v/v) solution as the extracting agent. Then the contents of the 36 phenolic and polyphenolic compounds were employed as chemical descriptors of the analyzed paprika samples to their classification and authentication by principal component analysis (PCA) and partial least squares regression-discriminant analysis (PLS-DA).

MATERIALS AND METHODS

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

Instrumentation. The determination of polyphenols and phenolic acids was carried out on an Open Accela UHPLC instrument (Thermo Fisher Scientific, San Jose, CA, USA), equipped with a quaternary pump and a CTC autosampler. The separation was performed by reversed-phase chromatography using an Ascentis Express C18 fused-core (100 × 2.1 mm, 2.7 μm partially porous particle size) column from Supelco (Bellefonte, PA, USA) and gradient elution using 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B) as mobile phase components, with a mobile phase flow rate of 300 μL/min. The elution gradient program was as follows: 0−5.5 min, isocratic elution at 5% solvent B; 5.5−6.5 min, linear gradient up to 10% solvent B; 6.5−12 min, isocratic elution at 10% followed by a 1 min increase to 20% solvent B; 13−18 min, isocratic elution at 20% solvent B; 18−19 min, linear gradient raising up to 50% solvent B and then 2 min elution at this percentage; 21−22 min, linear gradient to 95% solvent B and 3 min keeping this composition of the mobile phase. Afterward, initial conditions for a 5 min column re-equilibration were returned, completing a total elution program time of 30 min. The chromatographic column was kept at room temperature, and an injection volume of 10 μL, full loop mode, was employed.

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

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

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

https://dx.doi.org/10.1021/acs.jafc.9b06054
J. Agric. Food Chem. 2020, 68, 591−602

Journal of Agricultural and Food Chemistry pubs.acs.org/JAFC Article

Table 1. Instrumental MRM Acquisition Parameters

segment	time (min)	number	compound	tube lens offset voltage (V)	precursor ion		quantifier product ion			qualifier product ion			quantifier/qualifier ion ratio
					m/z	assignment	*m/z*	NCE (eV)	assignment	*m/z*	NCE (eV)	assignment	
1	0.7	1	D-(−)-quinic acid	−50	190.9	$[M-H]^-$	85.2	20	$[M-H-C_3H_6O_4]^-$	93.2	20	$[M-H-C_4H_2O_3]^-$	2.6
1	1.0	2	arbutin	−124	271.0	$[M-H]^-$	161.3	10	$[M-H-C_6H_{10}O_5]^-$	108.7	25	$[M-H-C_6H_6O_2]^-$	10.0
1	1.4	3	gallic acid	−95	169.0	$[M-H]^-$	125.1	15	$[M-H-CO_2]^-$	79.0	23	$[M-H-C_2H_2O_4]^-$	17.5
1	2.3	4	homogentisic acid	−94	167.2	$[M-H]^-$	123.0	13	$[M-H-CO_2]^-$	121.9	23	$[M-H-CHO_2]^{-\bullet}$	2.0
1	4.9	5	protocatechuic aldehyde	−72	137.0	$[M-H]^-$	135.9	20	$[M-H-H]^{-\bullet}$	92.1	25	$[M-H-CHO_2]^{-\bullet}$	3.9
1	5.3	6	4-hydroxybenzoic acid	−90	136.9	$[M-H]^-$	93.2	20	$[M-H-CO_2]^-$	65.0	35	$[M-H-C_2O_3]^-$	20.2
1	5.7	7	gentisic acid	−87	153.0	$[M-H]^-$	109.0	20	$[M-H-CO_2]^-$	81.4	20	$[M-H-C_2O_3]^-$	15.6
1−2	8.4	8	chlorogenic acid	−148	353.0	$[M-H]^-$	190.9	21	$[M-H-C_9H_6O_3]^-$	85.1	44	$[M-H-C_{12}H_{12}O_7]^-$	23.7
2	8.5	9	(+)-catechin	−73	288.9	$[M-H]^-$	244.8	15	$[M-H-C_2H_4O]^-$	203.2	20	$[M-H-C_4H_6O_2]^-$	2.0
2	8.8	10	caffeic acid	−63	179.0	$[M-H]^-$	134.9	16	$[M-H-CO_2]^-$	133.8	25	$[M-H-CHO_2]^{-\bullet}$	6.2
2	9.2	11	homovanillic acid	−97	181.1	$[M-H]^-$	137.2	10	$[M-H-CO_2]^-$	122.0	16	$[M-H-C_2H_3O_2]^{-\bullet}$	6.3
2	9.4	12	syringic acid	−83	196.9	$[M-H]^-$	182.0	14	$[M-H-CH_3]^{-\bullet}$	123.1	24	$[M-H-C_2H_2O_3]^-$	2.8
2	10.3	13	vanillin	−77	151.2	$[M-H]^-$	136.0	15	$[M-H-CH_3]^{-\bullet}$	91.9	20	$[M-H-C_2H_3O_2]^{-\bullet}$	7.4
2−3	11.6	14	(−)-epicatechin	−95	289.1	$[M-H]^-$	244.9	16	$[M-H-C_2H_4O]^-$	203.0	20	$[M-H-C_4H_6O_2]^-$	1.9
2−3	11.9	15	ethyl gallate	−97	197.2	$[M-H]^-$	123.9	22	$[M-H-C_3H_5O_2]^{-\bullet}$	169.0	15	$[M-H-C_2H_4]^-$	1.6
2−3	12.2	16	*p*-coumaric acid	−85	163.1	$[M-H]^-$	118.8	17	$[M-H-CO_2]^-$	93.1	35	$[M-H-C_3H_2O_2]^-$	14.8
2−3	12.3	17	(−)-epigallocatechin gallate	−125	457.0	$[M-H]^-$	169.0	19	$[M-H-C_{15}H_{12}O_6]^-$	125.2	39	$[M-H-C_{16}H_{12}O_8]^-$	3.1
3	12.5	18	syringaldehyde	−57	181.1	$[M-H]^-$	166.2	13	$[M-H-CH_3]^{-\bullet}$	150.9	21	$[M-H-CH_2O]^-$	1.3
3	12.9	19	umbelliferone	−94	160.9	$[M-H]^-$	133.1	20	$[M-H-CO]^-$	105.1	23	$[M-H-C_3H_4O]^-$	2.9
3	14.2	20	procyanidin C1	−151	864.8	$[M-H]^-$	407.0	40	$[M-H-C_{23}H_{22}O_{10}]^-$	286.9	30	$[M-H-C_{30}H_{26}O_{12}]^-$	1.4
3	14.4	21	veratric acid	−86	181.0	$[M-H]^-$	136.9	15	$[M-H-CO_2]^-$	107.0	25	$[M-H-C_2H_2O_3]^-$	5.9
3	14.4	22	ferulic acid	−91	193.2	$[M-H]^-$	134.1	18	$[M-H-C_2H_3O_2]^{-\bullet}$	178.1	14	$[M-H-CH_3]^{-\bullet}$	1.3
3	14.7	23	sinapic acid	−91	223.0	$[M-H]^-$	207.9	15	$[M-H-CH_3]^{-\cdot}$	163.8	18	$[M-H-C_2H_3O_2]^{-\cdot}$	1.9
3	14.9	24	polydatin	−144	389.1	$[M-H]^-$	227.0	20	$[M-H-C_6H_{10}O_5]^-$	185.2	38	$[M-H-C_8H_{12}O_6]^-$	9.8
3	15.0	25	rutin	−139	609.0	$[M-H]^-$	300.1	35	$[M-H-C_{12}H_{21}O_9]^{-\bullet}$	270.9	60	$[M-H-C_{13}H_{22}O_{10}]^-$	1.8
3	15.7	26	procyanidin A2	−155	575.1	$[M-H]^-$	284.9	22	$[M-H-C_{15}H_{14}O_6]^-$	449.0	23	$[M-H-C_6H_6O_3]^-$	1.3
3−4	15.7	27	nepetin-7-glucoside	−135	477.1	$[M-H]^-$	315.2	25	$[M-H-C_6H_{10}O_5]^-$	299.7	35	$[M-H-C_7H_{13}O_5]^{-\bullet}$	1.4
3−4	16.8	28	hesperidin	−139	608.8	$[M-H]^-$	301.0	20	$[M-H-C_{12}H_{20}O_9]^-$	325.1	35	$[M-H-C_{13}H_{16}O_7]^-$	21.8
3−4	17.2	29	homoplantaginin	−163	461.0	$[M-H]^-$	283.1	34	$[M-H-C_7H_{14}O_5]^-$	297.1	35	$[M-H-C_9H_8O_3]^-$	4.2
3−4	17.7	30	fisetin	−108	285.1	$[M-H]^-$	135.1	23	$[M-H-C_8H_6O_3]^-$	120.9	27	$[M-H-C_8H_4O_4]^-$	1.9
3−4	17.8	31	rosmarinic acid	−115	358.7	$[M-H]^-$	161.0	18	$[M-H-C_9H_{10}O_5]^-$	133.2	40	$[M-H-C_{10}H_{10}O_6]^-$	4.3
4	19.8	32	morin	−91	301.0	$[M-H]^-$	151.1	21	$[M-H-C_8H_6O_3]^-$	148.9	29	$[M-H-C_7H_4O_4]^-$	1.8
4	20.2	33	quercetin	−121	300.9	$[M-H]^-$	151.1	25	$[M-H-C_7H_2O_4]^-$	179.1	20	$[M-H-C_6H_2O_3]^-$	1.8
4	20.5	34	kaempferol	−107	285.0	$[M-H]^-$	185.2	25	$[M-H-C_4H_4O_3]^-$	117.2	43	$[M-H-C_7H_4O_5]^-$	1.1
4	21.6	35	asiatic acid	−126	487.3	$[M-H]^-$	409.1	35	$[M-H-C_2H_6O_3]^-$	379.4	45	$[M-H-C_3H_8O_4]^-$	1.3

https://dx.doi.org/10.1021/acs.jafc.9b06054
J. Agric. Food Chem. 2020, 68, 591−602

Table 1. continued

segment	time (min)	number	compound	tube lens offset voltage (V)	precursor ion		quantifier product ion			qualifier product ion			quantifier/qualifier ion ratio
					m/z	assignment	*m/z*	NCE (eV)	assignment	*m/z*	NCE (eV)	assignment	
4	24.2	36	betulinic acid	−123	455.0	$[M-H]^-$							

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

Samples were randomly analyzed with the proposed UHPLC-ESI-MS/MS method. Moreover, a QC and an instrumental chromatographic blank of acetonitrile were also injected every 10 analyzed samples.

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

X-data matrices in both PCA and PLS-DA consisted of the concentration levels of the 36 phenolic and polyphenolic compounds quantified in the set of paprika samples and QCs, whereas the Y-data matrix in PLS-DA defined the membership of each sample in the corresponding class. Data was autoscaled to equalize the influence of major and minor compounds on the descriptive models. Scatter plots of scores and loadings from principal components (PCs), in PCA, and from latent variables (LVs), in PLS-DA, were employed to study the distribution of samples and variables (quantified compounds), revealing patterns that could be correlated with their characteristics.

RESULTS AND DISCUSSION

UHPLC Chromatographic Separation. As commented in the Introduction section, one of the objectives of the present work is the development of an LC−MS/MS method for the determination of a total of 36 phenolic and polyphenolic compounds, which belong to different phenolic classes, in paprika samples. The separation of polyphenols and phenolic acids in food products by LC−MS techniques is normally addressed by reversed-phase chromatography under gradient elution conditions using acidified water and methanol or acetonitrile as mobile phase components.[18] For that purpose, as a first attempt in this work, the separation was carried out with an Ascentis Express C18 fused-core (100 × 2.1 mm, 2.7 μm partially porous particle size) column, using water and acetonitrile, both acidified with 0.1% formic acid, as mobile phase components, and applying a universal gradient elution profile from 0 to 90% acetonitrile in 25 min. Under these conditions, multiple co-elutions were observed, and almost all the analyzed compounds eluted within the first 5 min, showing that, when acetonitrile was used as the organic mobile phase modifier, low elutropic strength was needed for the elution of this family of compounds by reversed-phase chromatography. Therefore, the separation of the studied compounds was optimized by combining isocratic and linear gradient elution steps at low acetonitrile contents (5 to 50%) to improve separation among the more polar phenolic acids, increasing then the acetonitrile content to elute all the compounds. It should be noted that, due to the high number of compounds under study, a compromise between chromatographic resolution and analysis time was considered. Figure 1 shows the proposed UHPLC chromatographic separation for the 36 studied phenolic and polyphenolic compounds (see the elution program in the Instrumentation section). As can be seen, an acceptable chromatographic separation was obtained in less than 26 min, although still some partial and total co-elutions were found for some compounds, such as for homovanillic and syringic acids (peaks 11 and 12), *p*-coumaric acid, (−)-epigallocatechin gallate, and syringaldehyde (peaks 16, 17, and 18), and veratric and ferulic acids (peaks 21 and 22). However, the use of MS detection under MRM acquisition mode allowed us to overcome problems dealing with partial and total co-elutions for the

https://dx.doi.org/10.1021/acs.jafc.9b06054
J. Agric. Food Chem. 2020, 68, 591−602

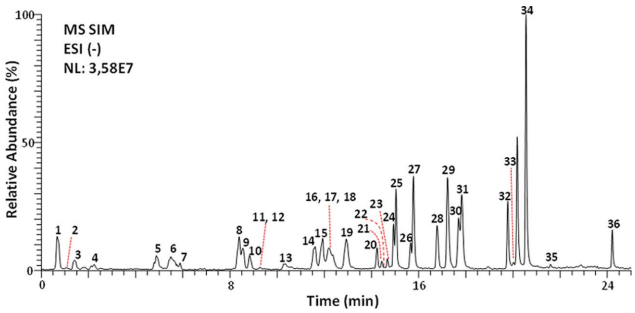


Figure 1. UHPLC-ESI-MS chromatographic separation of the 36 studied compounds (standard of 500 μg/L in water) under the proposed elution program (see Instrumentation). MS acquisition was performed in secondary ion monitoring (SIM) mode by following the $[M-H]^-$ ion for each compound. Peak identification: 1, D-(−)-quinic acid; 2, arbutin; 3, gallic acid; 4, homogentisic acid; 5, protocatechuic aldehyde; 6, 4-hydroxybenzoic acid; 7, gentisic acid; 8, chlorogenic acid; 9, (+)-catechin; 10, caffeic acid; 11, homovanillic acid; 12, syringic acid; 13, vanillin; 14, (−)-epicatechin; 15, ethyl gallate; 16, *p*-coumaric acid; 17, (−)-epigallocatechin gallate; 18, syringaldehyde; 19, umbelliferone; 20, procyanidin C1; 21, veratric acid; 22, ferulic acid; 23, sinapic acid; 24, polydatin; 25, rutin; 26, procyanidin A2; 27, nepetin-7-glucoside; 28, hesperidin; 29, homoplantaginin; 30, fisetin; 31, rosmarinic acid; 32, morin; 33, quercetin; 34, kaempferol; 35, asiatic acid; 36, betulinic acid.

correct determination of the studied compounds. In this regard, different SRM transitions were monitored for the co-eluting compounds, and no ion-suppression effects within ESI were present (that will be addressed in the next section).

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

Full-scan MS spectra (*m*/*z* 50−1000) of individual solutions of all the studied compounds in negative ionization mode were also registered. As an example, Figure S1a (Supporting Information) shows the obtained MS spectra of syringaldehyde and ethyl gallate. As can be seen, the most abundant ion (base peak) in both spectra is the deprotonated molecule, $[M-H]^-$, at *m*/*z* 181.1 and 197.2 for syringaldehyde and ethyl gallate, respectively. Similar results were obtained for most of the studied phenolic and polyphenolic compounds, with the deprotonated molecule being the spectrum base peak. Moreover, adduct formation with the mobile phase components was not observed. In general, no ion in-source fragmentation was obtained, except some particular compounds. For instance, in the case of polydatin, the spectrum base peak was not the deprotonated molecule but the $[M-H-C_6H_{10}O_5]^-$ ion at *m*/*z* 227.0, although the $[M-H]^-$ was also very abundant. In the case of syringaldehyde (Figure S1a) and gentisic and 4-hydroxybenzoic acids, ion source fragmentations with relative intensities lower than 40 and 60%, respectively, were observed. Finally, it should be mentioned that, in most of the MS spectra obtained, a signal at *m*/*z* 91.2 was also observed due to the dimer formation of the formic acid present in the mobile phase ($[HCOOH-HCOO]^-$). After the study of the MS spectra, the deprotonated ion was then proposed as the precursor ion for further fragmentation studies (Table 1).

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

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

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

Limits of detection (LODs), based on a signal-to-noise ratio of 3:1, were assessed by analyzing standard solutions at low concentration levels, obtaining values in a wide range depending on the compound (from 0.01 μg/L for D-(−)-quinic acid to 1.4 mg/L for kaempferol). Limits of quantitation (LOQs), based on a signal-to-noise ratio of 10:1, in the range of 0.03 μg/L−4.5 mg/L were then established. Of those, 7 compounds showed

https://dx.doi.org/10.1021/acs.jafc.9b06054
J. Agric. Food Chem. 2020, 68, 591−602

Journal of Agricultural and Food Chemistry pubs.acs.org/JAFC Article

Table 2. Method Performance: Instrumental Quality Parameters

					run-to-run precision (RSD, %)				day-to-day precision (RSD, %)				trueness (relative error, %)			
number	compound	LOD (μg/L)	LOQ (μg/L)	linearity (R^2)	5 μg/L	50 μg/L	500 μg/L	10 mg/L	5 μg/L	50 μg/L	500 μg/L	10 mg/L	5 μg/L	50 μg/L	500 μg/L	10 mg/L
1	D-(−)-quinic acid	0.01	0.03	0.998	7.0	4.9	4.7	2.2	14.6	6.3	6.0	4.2	5.0	0.1	5.2	3.6
2	arbutin	0.22	0.73	0.996	9.5	8.2	4.5	1.9	16.5	9.7	5.2	4.9	4.6	0.9	2.3	2.1
3	gallic acid	964	3214	0.995				3.0				3.7				4.1
4	homogentisic acid	770	2566	0.996				2.5				3.0				3.8
5	protocatechuic aldehyde	2	7	0.995	6.4	5.3	3.2	2.8	8.9	6.1	5.6	4.4	5.6	2.7	1.5	3.0
6	4-hydroxybenzoic acid	21	71	0.999		8.0	3.8	1.4		9.5	4.8	3.9		5.3	4.1	3.1
7	gentisic acid	11	35	0.998		1.5	0.4	0.2		9.4	7.9	5.0		15.0	6.8	3.3
8	chlorogenic acid	0.87	3	0.999		4.0	2.9	2.1		6.7	4.8	3.7		7.9	0.3	4.7
9	(+)-catechin	161	537	0.999				2.2				5.0				1.2
10	caffeic acid	18	60	0.995		5.1	2.6	2.5		8.9	6.1	2.9		9.7	6.0	3.1
11	homovanillic acid	425	1417	0.995				2.3				4.4				1.5
12	syringic acid	18	59	0.998		7.9	4.0	1.9		15.9	8.6	3.3		14.0	12.6	0.8
13	vanillin	10	33	0.998		5.4	1.5	1.1		6.6	2.3	1.7		3.4	4.5	1.7
14	(−)-epicatechin	1282	4272	0.999				2.8				4.3				3.6
15	ethyl gallate	262	872	0.998			3.1	1.3			6.7	2.6			9.4	6.9
16	*p*-coumaric acid	4	12	0.995		5.8	2.9	2.0		14.1	4.3	4.1		13.7	5.5	8.2
17	(−)-epigallocatechin gallate	770	2565	0.999				1.1				2.0				0.9
18	syringaldehyde	2	8	0.999		9.9	5.0	0.7		10.6	5.5	1.9		4.5	0.3	1.5
19	umbelliferone	0.37	1	0.998	12.3	6.4	3.4	2.3	15.4	8.3	5.6	5.4	9.4	5.3	3.4	4.0
20	procyanidin C1	359	1196	0.998				1.0				2.6				3.9
21	veratric acid	281	936	0.999				3.9				8.5				1.4
22	ferulic acid	2	6	0.997	6.2	4.4	5.5	4.0	18.5	9.3	7.8	6.0	13.3	0.2	0.9	0.4
23	sinapic acid	25	84	0.995			5.9	2.5			6.5	5.3			1.4	0.7
24	polydatin	0.14	0.48	0.999	5.9	2.3	2.6	2.2	19.2	13.6	5.8	2.8	11.1	4.4	2.1	4.3
25	rutin	3	9	0.996		5.0	4.8	4.4		15.7	7.3	6.4		2.6	3.0	4.4
26	procyanidin A2	170	566	0.998				1.1				2.5				2.1
27	nepetin-7-glucoside	0.06	0.21	0.998	6.2	5.0	3.0	1.9	12.6	7.7	5.8	5.1	12.3	8.0	4.0	3.8
28	hesperidin	0.31	1	0.999	7.9	4.1	2.7	2.1	16.3	5.3	5.2	3.8	10.5	0.7	2.3	1.6
29	homoplantaginin	0.19	0.63	0.998	9.1	3.5	1.9	1.3	11.4	4.7	5.7	3.3	11.8	1.2	4.0	1.9
30	fisetin	759	2529	0.995				1.0				1.4				4.4
31	rosmarinic acid	12	41	0.999		2.7	1.2	0.8		5.2	4.4	2.0		7.3	0.6	1.9
32	morin	209	696	0.999		4.5	2.1	1.2		9.6	6.2	3.5		8.6	2.6	2.2
33	quercetin	89	296	0.996			1.8	1.2			7.1	2.0			1.4	7.0
34	kaempferol	1357	4522	0.998				1.6				2.7				3.4
35	asiatic acid	210	700	0.995				6.7				13.2				0.9
36	betulinic acid	265	885	0.998				7.3				10.3				5.6

https://dx.doi.org/10.1021/acs.jafc.9b06054
J. Agric. Food Chem. 2020, 68, 591−602

Journal of Agricultural and Food Chemistry pubs.acs.org/JAFC Article

Table 3. Concentrations of Studied Compounds in the Analyzed Paprika Samples[a]

		La Vera PDO paprika (mg/L)						Murcia PDO paprika (mg/L)				Czech Republic paprika (mg/L)					
		spicy		bittersweet		sweet		spicy		sweet		spicy		sweet-smoked		sweet	
number	compound	range	mean ± SD	range	mean ± SD	range	mean ± SD	range	mean ± SD	range	mean ± SD	range	mean ± SD	range	mean ± SD	range	mean ± SD
3	gallic acid	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ
4	homogentisic acid	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ
5	protocatechuic aldehyde	0.1−0.6	0.3 ± 0.1	0.2−0.7	0.3 ± 0.1	0.1−0.6	0.3 ± 0.1	0.1−0.5	0.4 ± 0.1	0.4−0.8	0.7 ± 0.1	0.6−0.8	0.7 ± 0.1	0.5−0.6	0.5 ± 0.1	0.6−0.8	0.7 ± 0.1
8	chlorogenic acid	5.0−10.1	7.4 ± 1.5	5.3−10.4	6.9 ± 1.9	5.1−9.7	6.8 ± 1.2	5.2−6.8	6.0 ± 0.6	7.1−12.4	10.4 ± 1.5	6.6−7.5	7.1 ± 0.4	5.6−8.3	7.0 ± 1.0	5.8−8.4	7.0 ± 1.0
10	caffeic acid	1.4−2.8	1.8 ± 0.3	1.2−2.0	1.7 ± 0.2	1.0−2.4	1.6 ± 0.3	1.8−2.6	2.2 ± 0.3	1.6−2.0	1.8 ± 0.1	2.3−5.2	3.2 ± 1.2	2.3−7.4	4.0 ± 2.2	1.9−5.5	3.0 ± 1.5
11	homovanillic acid	2.8−8.9	6.1 ± 1.7	1.6−10.3	5.6 ± 2.3	0.5−12.9	5.9 ± 2.8	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	4.23−5.0	4.6 ± 0.3	n.d.	n.d.
13	vanillin	1.9−11.2	3.9 ± 2.0	0.9−6.4	3.3 ± 1.4	1.4−7.2	3.5 ± 1.5	3.2−4.4	3.7 ± 0.3	2.3−3.3	2.7 ± 0.3	2.1−3.3	2.7 ± 0.5	1.9−2.7	2.3 ± 0.3	1.7−2.8	2.2 ± 0.5
16	*p*-coumaric acid	1.6−5.6	3.3 ± 1.0	2.1−4.5	3.3 ± 0.7	1.4−5.0	3.1 ± 0.9	3.9−10.3	8.2 ± 1.8	1.3−3.7	2.8 ± 0.8	9.7−10.7	10.1 ± 0.4	7.4−8.7	8.0 ± 0.6	6.2−8.6	7.3 ± 1.0
18	syringaldehyde	4.5−10.8	7.7 ± 1.7	4.0−14.3	8.5 ± 2.9	1.2−14.8	8.0 ± 3.0	0.8−1.8	1.2 ± 0.3	0.8−1.8	1.2 ± 0.4	1.4−1.8	1.6 ± 0.2	3.0−4.1	3.6 ± 0.5	1.6−1.9	1.7 ± 0.1
19	umbelliferone	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.1−0.2	0.1 ± 0.1	n.d.	n.d.	n.d.	n.d..
22	ferulic acid	1.4−16.3	7.1 ± 3.4	2.6−8.1	5.2 ± 1.6	1.3−9.9	5.7 ± 2.4	2.8−19.1	10.3 ± 5.2	4.7−16.1	9.6 ± 3.6	11.8−18.4	14.9 ± 2.7	12.8−19.2	15.2 ± 2.8	5.4−12.1	9.5 ± 3.2
23	sinapic acid	2.0−5.7	3.4 ± 1.0	1.9−5.0	3.2 ± 1.0	1.6−4.8	3.1 ± 1.0	1.2−6.8	2.6 ± 1.5	1.7−4.5	3.2 ± 0.9	1.9−4.6	3.7 ± 1.1	2.1−5.5	3.6 ± 1.4	3.1−4.5	4.1 ± 0.6
25	rutin	1.3−6.1	3.0 ± 1.2	1.4−5.0	2.9 ± 0.8	1.3−4.5	2.8 ± 0.9	1.0−1.8	1.4 ± 0.3	3.8−5.4	4.5 ± 0.5	3.4−3.8	3.6 ± 0.2	2.8−3.3	3.0 ± 0.2	2.5−3.0	2.6 ± 0.2
27	nepetin-7-glucoside	n.d. − 0.2	0.1 ± 0.1	n.d.−0.2	0.1 ± 0.1	n.d.−0.1	0.1 ± 0.0	0.1−0.2	0.2 ± 0.0	0.1−0.2	0.1 ± 0.0	0.3−0.4	0.3 ± 0.0	0.2−0.2	0.2 ± 0.0	0.6−0.7	0.6 ± 0.0
28	hesperidin	<LOQ−0.4	0.4 ± 0.0	<LOQ−5.1	2.6 ± 3.6	<LOQ−5.8	5.8 ± 0.0	0.1−0.5	0.2 ± 0.1	3.8−6.1	4.6 ± 0.7	7.1−9.2	7.9 ± 1.0	1.2−2.4	1.9 ± 0.5	<LOQ−0.2	0.1 ± 0.10
29	homoplantaginin	0.001−0.008	0.004 ± 0.002	0.002−0.009	0.005 ± 0.002	0.001−0.008	0.004 ± 0.002	0.004−0.011	0.008 ± 0.002	0.008−0.017	0.011 ± 0.003	0.011−0.040	0.031 ± 0.012	0.004−0.024	0.012 ± 0.010	0.011−0.038	0.027 ± 0.012
31	rosmarinic acid	1.1−5.5	2.3 ± 1.8	1.4−8.3	3.8 ± 2.5	1.1−5.6	2.5 ± 1.6	1.5−2.1	1.7 ± 0.2	1.2−4.4	3.6 ± 0.9	5.7−10.4	7.4 ± 1.8	3.7−7.5	4.7 ± 1.6	2.9−4.8	3.7 ± 0.7
33	quercetin	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ
34	kaempferol	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ	<LOQ
36	betulinic acid	1.2−6.4	3.2 ± 1.6	1.4−9.1	3.6 ± 1.5	1.0−8.3	3.0 ± 1.6	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.

[a]n.d.: not detected; SD: standard deviation.

https://dx.doi.org/10.1021/acs.jafc.9b06054
J. Agric. Food Chem. 2020, 68, 591−602

LOQ values equal to or below 1 μg/L, 9 compounds were in the range of 1−50 μg/L, 12 compounds were in the range 50 μg/L−1 mg/L, and only 8 compounds provided LOQ values higher than 1 mg/L. Taking into account that these compounds are naturally occurring secondary metabolites in plant-based products and the huge variety of compounds and concentration levels that can be found (usually at the relatively low to high mg/L level), these values are acceptable for the quantitation of this family of compounds in paprika samples.

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

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

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

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

Sample Analysis. The applicability of the proposed UHPLC-ESI-MS/MS method for the determination of the 36 studied compounds in paprika was evaluated. Paprika samples were extracted by solid−liquid extraction with water:acetonitrile (20:80 v/v) as described in the Materials and Methods. The obtained extracts were then analyzed in triplicate with the proposed analytical method, and targeted compounds were quantified by external calibration. Quantitation results for all the 111 paprika samples analyzed are provided in the Supporting Information. As an overview, Table 3 shows, for each compound, the concentration ranges and the mean values ± standard deviations found in the analyzed paprika samples depending on the production region (La Vera PDO, Murcia PDO, and Czech Republic) and the paprika flavors. Gallic acid, quercetin, and kaempferol were always detected below the LOQ value. 16 of the studied compounds (D-(−)-quinic acid, arbutin, 4-hydroxybenzoic acid, gentisic acid, (+)-catechin, syringic acid, (−)-epicatechin, ethyl gallate, (−)-epigallocatechin gallate, procyanidin C1, veratric acid, polydatin, procyanidin A2, fisetin, morin, and asiatic acid) were not detected in any of the 111 paprika samples (these compounds were not included in Table 3). Anyway, these compounds were preliminarily selected for this study because of their presence in other similar matrices such as Serbian red spice paprika, Italian red sweet pepper, or in red pepper fruits and seed oils.[11,32−34]

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

In more details, some compounds were only found in some specific paprika samples depending on the production region, so they could be considered as putative markers with high selectivity with respect to origins. For example, homogentisic acid was only detected in Czech Republic samples, although always below the LOQ. Umbelliferone was only found, at low concentrations, in the spicy flavor paprika from Czech Republic, while betulinic acid was only found in La Vera PDO samples.

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

The significance of the differences in the concentration values among classes was evaluated using statistical tests. As a result, most of the previous considerations regarding the occurrence of quite featured compounds of different classes could be confirmed. Results commented here have been limited to various illustrative cases since a comprehensive analysis dealing with all variables seems to be excessive. Data given as follows corresponds to the probability (p values) of Student's t test for the comparison of means of two classes before a Fisher test of variances. We assume a confidence level of 0.99, so when $p < 0.01$, differences in the analyte concentrations among the classes are significant. Results reveal the existence of several compounds

https://dx.doi.org/10.1021/acs.jafc.9b06054
J. Agric. Food Chem. 2020, 68, 591−602

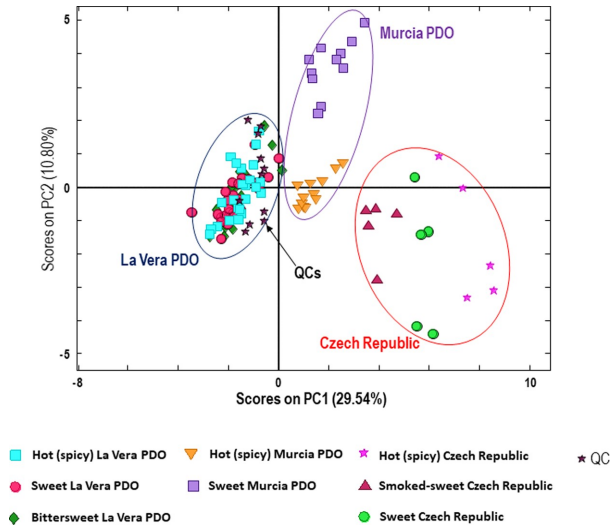


Figure 2. PCA score plot of PC1 vs PC2 when using the 36 compound concentrations found in the analyzed paprika samples as chemical descriptors.

such as syringaldehyde (at least, $p < 0.0006$), caffeic acid (at least, $p < 0.0042$), and homoplantaginin (at least, $p < 0.0016$) with statistically relevant differences in the concentration levels depending on the origin. Other species such as ferulic acid and nepetin-7-glucoside show no significant differences among Murcia and Czech Republic ($p = 0.02$ and 0.048, respectively). Finally, compounds such as chlorogenic acid are unspecific, so their role in class description and discrimination is quite irrelevant ($p = 0.04$, 0.04, and 0.91 for La Vera/Murcia, Murcia/Czech Republic, and La Vera/Czech Republic, respectively).

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

The study of the PCA loading plot allows us to see which variables (concentration) are defining the separation observed in the score plot. Figure S4 (Supporting Information) shows the obtained PCA loading plot of PC1 versus PC2. Thus, the separation of Czech Republic samples is achieved mainly by the presence of homoplantaginin, nepetin-7-glucoside, *p*-coumaric acid, and kaempferol among other compounds. Chlorogenic acid, rutin, and hesperidin are more discriminating compounds for the Murcia PDO samples. In contrast, vanillin, homovanillic acid, syringaldehyde, and quercetin seem to be the more characteristic compounds to separate La Vera PDO samples from the other two groups. Although more studies will be necessary, a priori, these compounds would be good candidates as potential biomarkers for the authentication of paprika.

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

The first study was focused on the classification of paprika samples according to geographical origin into La Vera, Murcia, and Czech Republic types. In this case, the calibration set was composed of 48 La Vera, 16 Murcia, and 10 Czech Republic samples randomly selected, which approximately corresponded to 70% of the analyzed samples. The other ~30% of the samples were used as the test set for prediction purposes. The optimum number of LVs established by cross validation using Venetian

https://dx.doi.org/10.1021/acs.jafc.9b06054
J. Agric. Food Chem. 2020, 68, 591−602

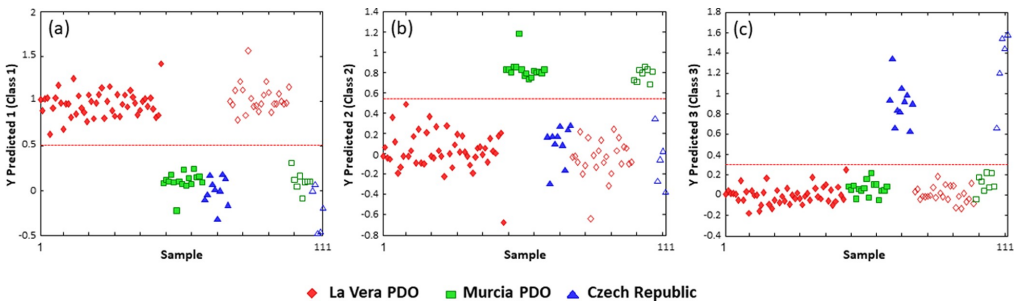


Figure 3. PLS-DA classification plots according to the production region. (a) La Vera versus other classes; (b) Murcia versus other classes; (c) Czech Republic versus other classes. Sample assignment: rhombus = La Vera PDO, square = Murcia, triangle = Czech Republic. The dashed line means the classification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

blinds was 4, providing the minimum of the root mean square error of cross validation (RMSECV) function. The analysis of scores and loadings of LV1 versus LV2 (not shown here) revealed that the three classes were perfectly separated, and relevant compounds for their discrimination were similar to those annotated for PCA.

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

Figure 3 shows the classification plots corresponding to (a) La Vera (rhombus symbols) versus the other samples, (b) Murcia (square symbols) versus the other samples, and (c) Czech Republic (triangle symbols) versus the other samples. The dashed line indicates the classification boundary, so samples belonging to the targeted class were located to the top while those belonging to the other types were to the bottom. Samples to be used for calibration were on the left, and those for prediction were on the right side. Results indicated that the classification rate was 100%, so all the samples were correctly assigned to the corresponding classes in both calibration and prediction steps (confusion matrix was [24, 0, 0; 0, 8, 0; 0, 0, 5] for La Vera, Murcia, and Czech Republic, respectively).

Table S2 (Supporting Information) shows the validation results for both calibration and prediction. The obtained validation results are satisfactory. Calibration sensitivity and specificity are 1, and the RMSECV and the bias showed values tending to zero, ensuring a good calibration model.

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

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

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jafc.9b06054.

Structures, family groups, CAS numbers, and suppliers of the 36 polyphenolic compounds under study; quality indicators for the calibration and prediction of the proposed PLS-DA model when dealing with the classification of paprika samples regarding the production region; full-scan spectra, collision energy curves, and product-ion spectra for syringaldehyde and ethyl gallate; boxplots with whiskers representing the concentrations of

https://dx.doi.org/10.1021/acs.jafc.9b06054
J. Agric. Food Chem. 2020, 68, 591−602

selected compounds in samples; bar plots showing concentration levels found in the analyzed paprika samples for syringaldehyde, rutin, and nepetin-7-glucoside; PCA loading plot of PC1 versus PC2; plots of La Vera PDO class qualitative parameters for the PLS-DA model; PLS-DA score plot of LV1 versus LV2 for the classification of each production region according to flavor varieties; plots of hot Murcia PDO and sweet Czech Republic paprika qualitative parameters for the PLS-DA models (PDF)

Concentration levels (mg/L) of the 36 determined compounds in the 111 paprika samples analyzed (XLSX)

AUTHOR INFORMATION

Corresponding Author

Oscar Núñez − *University of Barcelona, Barcelona, Spain;* orcid.org/0000-0001-5850-8972; Phone: 34-93-403-3706; Email: oscar.nunez@ub.edu; Fax: 34-93-402-1233

Other Authors

Sergio Barbosa − *University of Barcelona, Barcelona, Spain*

Guillem Campmajó − *University of Barcelona, Barcelona, Spain*

Javier Saurina − *University of Barcelona, Barcelona, Spain*

Lluis Puignou − *University of Barcelona, Barcelona, Spain*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jafc.9b06054

Author Contributions

#S.B. and G.C. contributed equally in the research of this work.

Funding

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

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

O.N. is a Serra Hunter Fellow. G.C. thanks the University of Barcelona for the APIF grant.

REFERENCES

(1) Serrano, N.; Cetó, X.; Núñez, O.; Agaó, M.; Ganz, A.; Ariño, C.; Díaz-Cruz, J. M. Characterization and Classification of Spanish Paprika (*Capsicum Annuum* L.) by Liquid Chromatography Coupled to Electrochemical Detection with Screen-Printed Carbon-Based Nanomaterials Electrodes. *Talanta* 2018, *189*, 296−301.

(2) Barboza, G. E.; García, C. C.; Gonzalez, S. L.; Scaldaferro, M.; Reyes, X. Four New Species of Capsicum (Solanaceae) from the Tropical Andes and an Update on the Phylogeny of the Genus. *PLoS One* 2019, *14*, No. e0209792.

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

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

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

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

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

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

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

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

(11) Mudric, S. Z.; Gasic, U. M.; Dramicanin, A. M.; Ciric, I. Z.; Milojkovic-Opsenica, D. M.; Popovic-Đorđevic, J. B.; Momirovic, N. M.; Tesic, Ž. L. The Polyphenolics and Carbohydrates as Indicators of Botanical and Geographical Origin of Serbian Autochthonous Clones of Red Spice Paprika. *Food Chem.* 2017, *217*, 705−715.

(12) Han, X.; Shen, T.; Lou, H. Dietary Polyphenols and Their Biological Significance. *Int. J. Mol. Sci.* 2007, *8*, 950−988.

(13) Monago-Maraña, O.; Muñoz De La Peña, A.; Galeano-Díaz, T. Isocratic LC-DAD-FLD Method for the Determination of Flavonoids in Paprika Samples by Using a Rapid Resolution Column and Post-Column PH Change. *Talanta* 2016, *152*, 15−22.

(14) Monago-Maraña, O.; Durán-Merás, I.; Galeano-Díaz, T.; Muñoz De La Peña, A. Fluorescence Properties of Flavonoid Compounds. Quantification in Paprika Samples Using Spectrofluorimetry Coupled to Second Order Chemometric Tools. *Food Chem.* 2016, *196*, 1058−1065.

(15) Vallejos, S.; Moreno, D.; Ibeas, S.; Muñoz, A.; García, F. C.; García, J. M. Polymeric Chemosensor for the Colorimetric Determination of the Total Polyphenol Index (TPI) in Wines. *Food Control* 2019, *106*, 106684.

(16) Fernández de Córdova, M. L.; Ruiz Medina, M. Chapter 29 - Analytical Methods for Determination of Polyphenols in Beer. In *Processing and Impact on Antioxidants in Beverages*; 2014; p Pages 289−299.

(17) Pardo-Mates, N.; Vera, A.; Barbosa, S.; Hidalgo-Serrano, M.; Núñez, O.; Saurina, J.; Hernández-Cassou, S.; Puignou, L. Characterization, Classification and Authentication of Fruit-Based Extracts by Means of HPLC-UV Chromatographic Fingerprints, Polyphenolic Profiles and Chemometric Methods. *Food Chem.* 2017, *221*, 29.

(18) Lucci, P.; Saurina, J.; Núñez, O. Trends in LC-MS and LC-HRMS Analysis and Characterization of Polyphenols in Food. *TrAC, Trends Anal. Chem.* 2017, *88*, 1−24.

(19) Morales-Soto, A.; Gómez-Caravaca, A. M.; García-Salas, P.; Segura-Carretero, A.; Fernández-Gutiérrez, A. High-Performance Liquid Chromatography Coupled to Diode Array and Electrospray Time-of-Flight Mass Spectrometry Detectors for a Comprehensive Characterization of Phenolic and Other Polar Compounds in Three Pepper (*Capsicum Annuum* L.) Samples. *Food Res. Int.* 2013, *51*, 977−984.

(20) Chen, L.; Kang, Y. H. Anti-Inflammatory and Antioxidant Activities of Red Pepper (Capsicum Annuum L.) Stalk Extracts: Comparison of Pericarp and Placenta Extracts. *J. Funct. Foods* 2013, *5*, 1724−1731.

(21) Consejo de la union europea. RCOUNCIL REGULATION (EC) No 510/2006 of 20 March 2006 on the Protection of

https://dx.doi.org/10.1021/acs.jafc.9b06054
J. Agric. Food Chem. 2020, 68, 591−602

Geographical Indications and Designations of Origin for Agricultural Products and Foodstuffs, Off. J. European Union L93 12−25., 2006.

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

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

(24) COMMISSION REGULATION (EC) No 982/e Registering Certain Names in the Register of Protected Designations of Origin and Protected Geographical Indications (Pimentó de La Vera (PDO) □ KarlovarskýSuchar (PGI) □ Riso Di Baraggia Biellese.

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

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

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

(28) Cetó X.; Serrano, N.; Aragó M.; Gámez, A.; Esteban, M.; Díaz-Cruz, J.; Núñez, O. Determination of HPLC-UV Fingerprints of Spanish Paprika (*Capsicum Annuum* L.) for Its Classification by Linear Discriminant Analysis. *Sensors* 2018, *18*, 4479.

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

(30) Massart, J.; Vandeginste, D. L.; Buydens, B. G. M.; de Jong, L. M. C.; Lewi, S.; Smeyers-Verbeke, P. J. Handbook of Chemometrics and Qualimetrics. *J. Chem. Inf. Comput. Sci.* 1997,

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

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

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

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

https://dx.doi.org/10.1021/acs.jafc.9b06054
J. Agric. Food Chem. 2020, 68, 591−602

Supporting Information for:

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

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

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

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

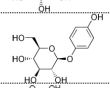
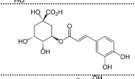
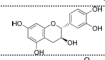
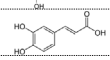
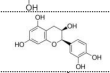
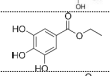
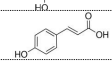
[c] Serra Hunter Fellow. Generalitat de Catalunya (Spain).

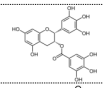
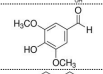
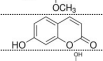
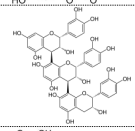
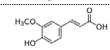
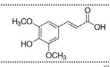
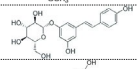
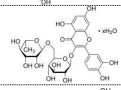
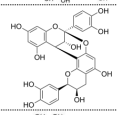
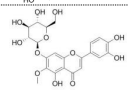
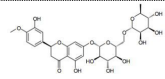
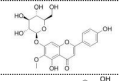
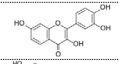
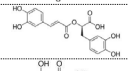
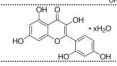
* Corresponding author: Oscar Núñez

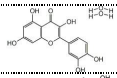
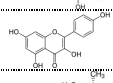
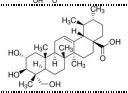
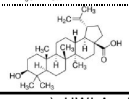
(Phone: 34-93-403-3706, Fax: 34-93-402-1233, e-mail: oscar.nunez@ub.edu)

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

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

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

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

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

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

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

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

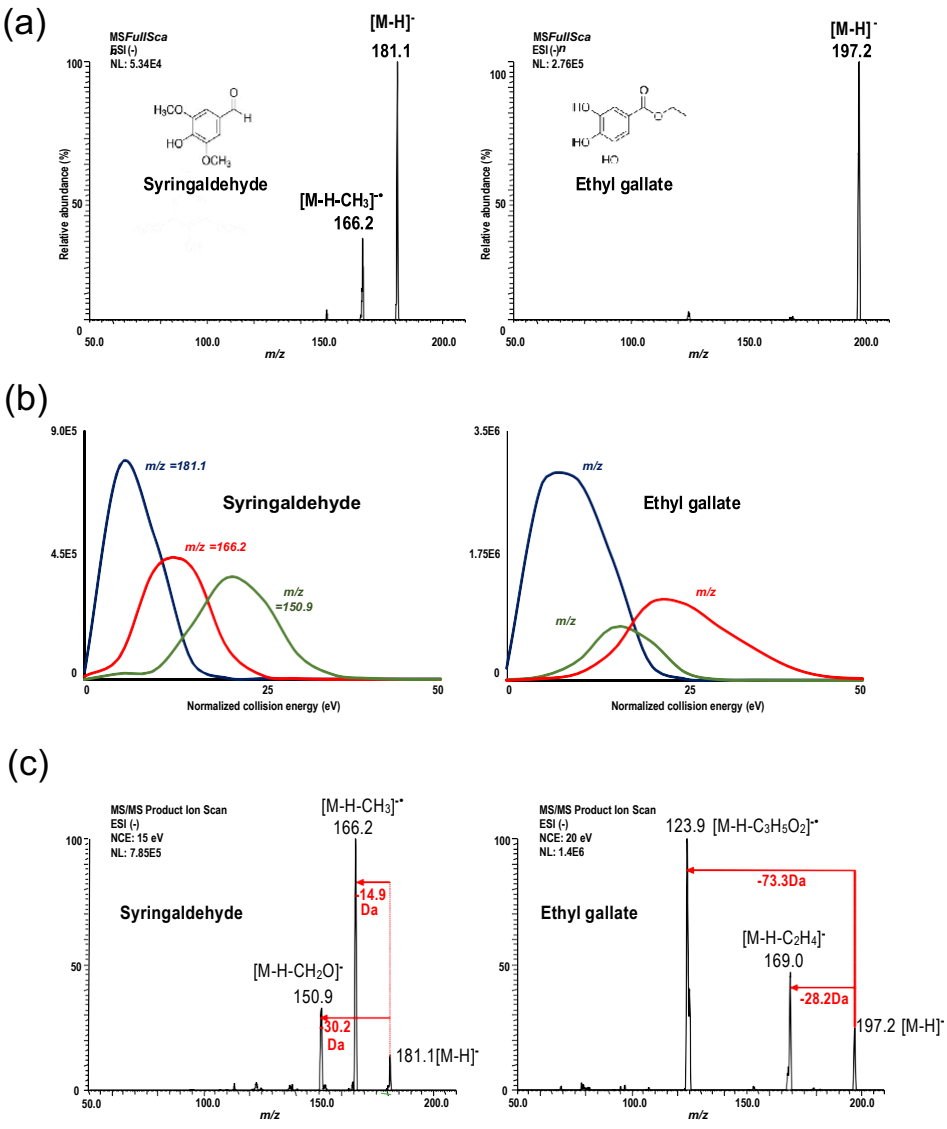


Figure S1. (a) MS full scan spectra, (b) collision energy curves, and (c) MS/MS product ion scan spectra for Syringaldehyde and Ethyl gallate.

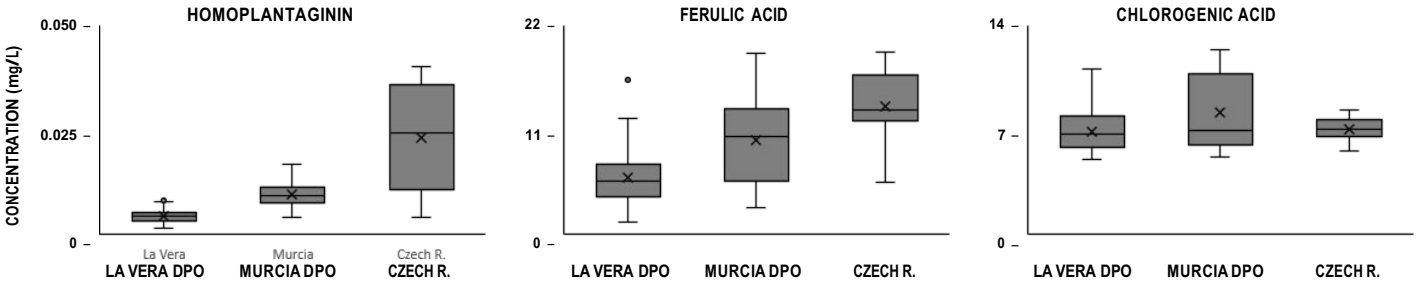


Figure S2. Boxplots with whiskers representing the concentration of selected compounds in the set of La Vera, Murcia and Czech Republic samples. From the left to the right: homoplantaginin, ferulic acid and chlorogenic acid.

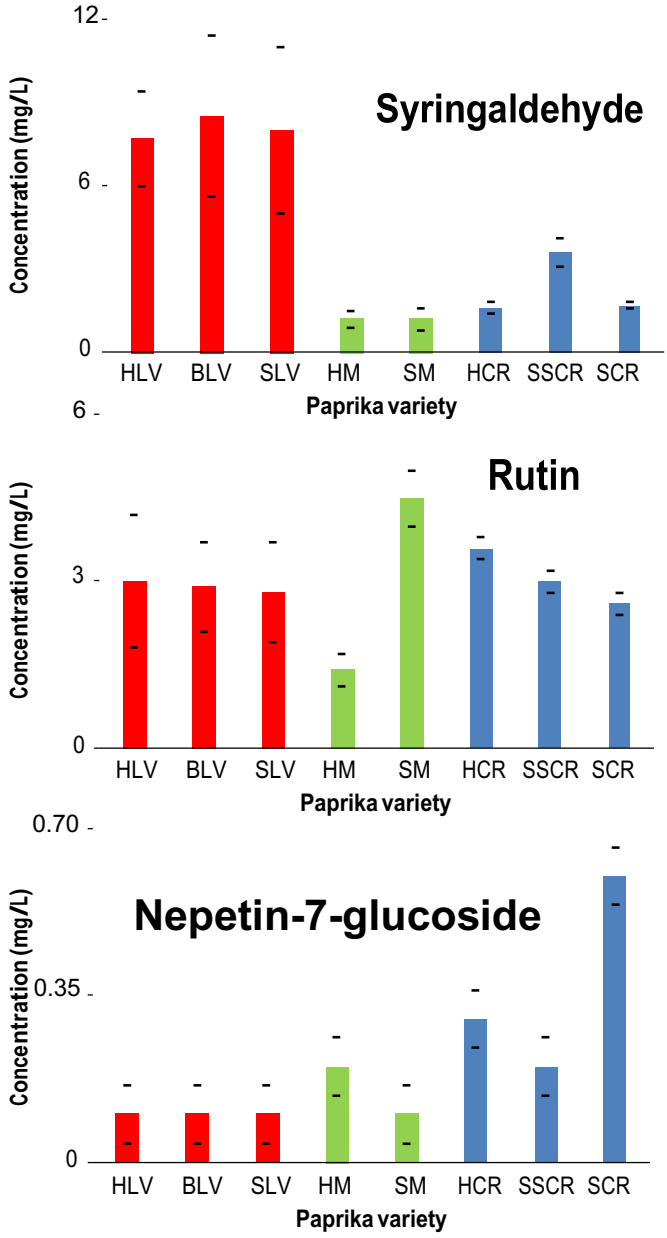


Figure S3. Bar plots showing concentration levels found in the analyzed paprika samples for syringaldehyde, rutin and nepetin-7-glucoside. HLV: hot (spicy) La Vera PDO; BLV: Bittersweet La Vera PDO; SLV: Sweet La Vera PDO; HM: Hot (spicy) Murcia PDO; SM: Sweet Murcia PDO; HCR: Hot (spicy) Czech Republic;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

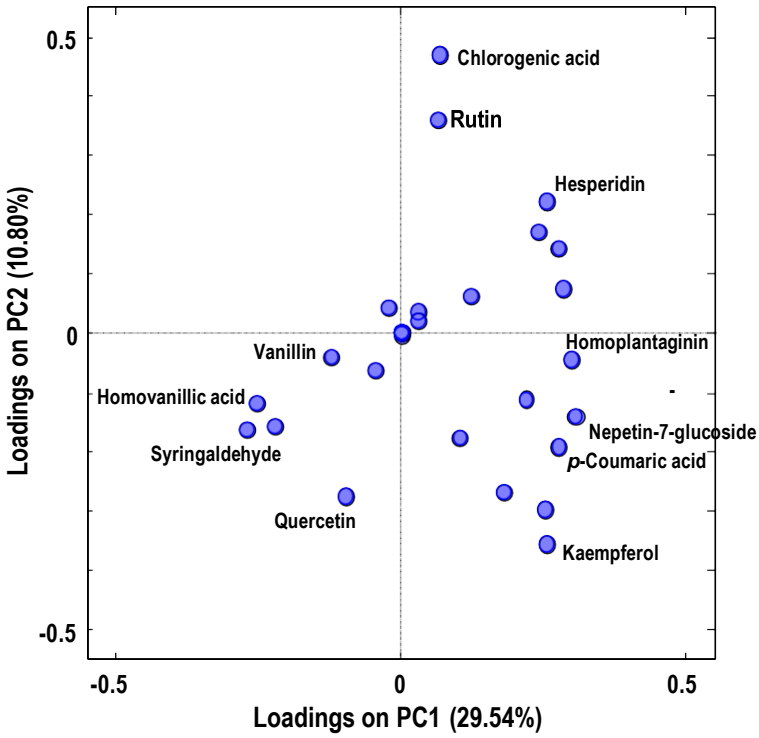


Figure S4. PCA loadings plot of PC1 *vs* PC2 when using the 36 compound concentrations found in the analysed paprika samples as chemical descriptors.

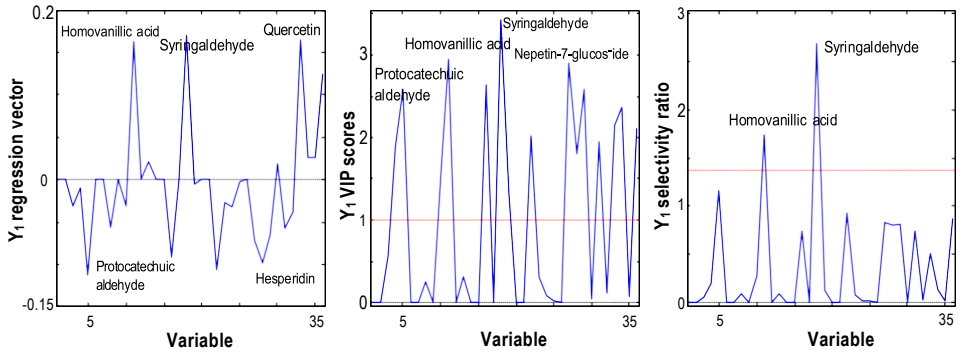


Figure S5. Plots of La Vera PDO class qualitative parameters (regression vector, the variable importance in projection (VIP) and the selectivity ratio) for the PLS-DA model obtained for the classification of paprika samples according to the production region (La Vera PDO, Murcia PDO and Czech Republic).

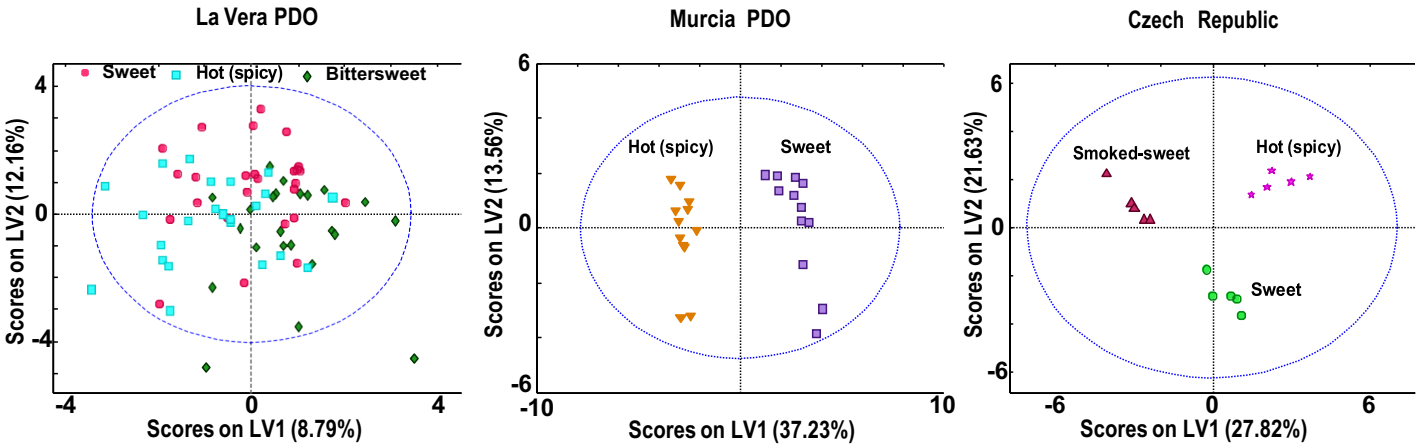


Figure S6. PLS-DA score plots of LV1 vs LV2 when using the 36 compound concentrations as chemical descriptors for the classification of each production regions (La Vera PDO, Murcia PDO and Czech Republic samples) according to their different flavor varieties.

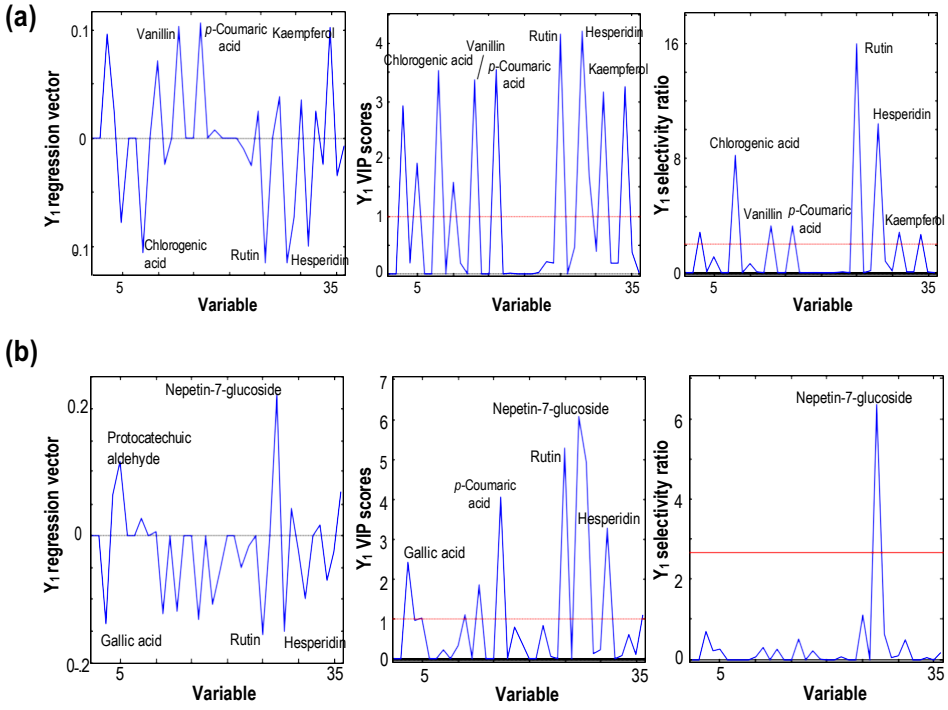


Figure S7. Plots of a) hot Murcia PDO and b) sweet Czech Republic paprika qualitative parameters (regression vector, the variable importance in projection (VIP) and the selectivity ratio) for the PLS-DA models obtained for the classification according to the different flavor varieties.

2.2.2. PUBLICATION II

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

Arrizabalaga-Larrañaga, A.; Campmajó, G.; Saurina, J.; Núñez, O.; Santos, F. J.; Moyano, E.

LWT. 2021, 139, 110533.

LWT - Food Science and Technology 139 (2021) 110533

Contents lists available at ScienceDirect

LWT

journal homepage: www.elsevier.com/locate/lwt

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

Ane Arrizabalaga-Larrañaga [a,1], Guillem Campmajó [a,b,1], 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

ARTICLE INFO

ey ords
Paprika
Capsaicinoids
Carotenoids
UHPLC-HRMS
Food authentication

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 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 .

I. 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 Goncalves, 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), *alocsai fűszerpaprika-őrlemény* (Hungary), *Szegedi fűszerpaprika-őrlemény* (Hungary), *Piment d'Espelette* (France), and *Paprika Žitava* (Slovakia). The presence of the PDO label ensures the

- Corresponding author. Department of Chemical Engineering and Analytical Chemistry University of Barcelona, Av. Diagonal 645, E-08028, Barcelona, Spain.
E-mail address encarna.moyano@ub.edu (E. Moyano).
[1] These authors contributed equally in the research of this work.

https //doi.org/10.1016/j.lwt.2020.110533
Received 6 August 2020; Received in revised form 4 November 2020; Accepted 5 November 2020
Available online 10 November 2020
0023-6438/

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 et al., 2018; Cetó, Sánchez, Serrano, Díaz-Cruz, & Núñez, 2020), and ultra-high-performance liquid chromatography coupled to high-resolution mass spectrometry (UHPLC–HRMS) (Barbosa, Saurina, Puignou, & Nuñez, 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 (1000 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

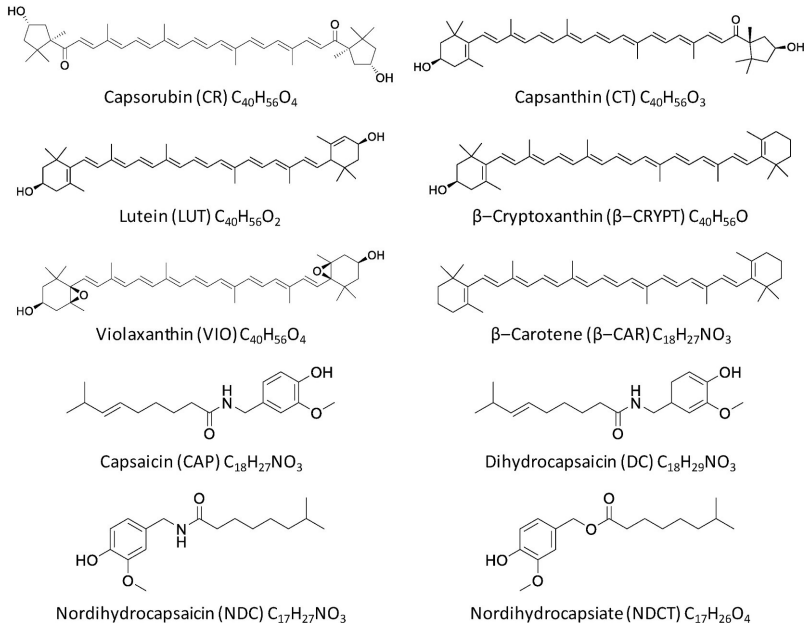


Fig. I. Chemical structures, acronyms, and chemical formula of the studied capsaicinoids and carotenoids.

subsequently used to obtain calibration solutions (0.001–10 mg L^{-1}) for quantification. All stock solutions were stored at − 20 °C 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 μL 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 3×10^6 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 and the Czech Republic; regarding types, hot, bittersweet, and sweet paprika were considered. Table 1 summarizes sample details such as the acronyms used for each region and the number of samples analyzed for each type of sample.

A simple solid-liquid extraction of target analytes from paprika samples was carried out as follows 0.05 g of paprika were extracted with 4 mL of methanol acetone (1 1, *v/v*) solution in a 15 mL PTFE tube. Subsequently, the sample was stirred in a Stuart Vortex for 0.5 min (Staffordshire, United Kingdom) and sonicated for 10 min (5510 Branson ultrasonic bath, Hampton, NH, USA). Afterward, the extract was centrifuged for 15 min at 4500 rpm (ROTANTA 460 HR Centrifuge, Hettich, Germany). Finally, the supernatant was filtered through 0.22 μm Nylon membrane filters and stored at 4 °C in 2 mL glass injection vials until the analysis by UHPLC-HRMS.

2 4 Instrumental and quality parameters

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

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

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

Table 1
Description of the samples analyzed 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]	es	2017
	Murcia	M	15	15	–	es	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.

2 5 Data analysis

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

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

. e ult and di cu ion

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, β−CR PT; β-carotene, β−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 *miz* 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 *miz* 170.1536 (CAP), *miz* 172.1693 (DC), and *miz* 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 *miz* 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, *miz* 583.4137; VIO, *miz* 583.4137; CT, *miz* 567.4186; β−CR PT, *miz* 535.4291; LUT, *miz* 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 *miz* 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 *miz* 137.0597, *miz* 122.0362, *miz* 94.0413 and *miz* 66.0464 (Fig. 2). The fragment ion at *miz* 122.0362 $[C_7H_6O_2]^{+\cdot}$ was produced by the α−cleavage of the C-O bond, generating the dissociation of the methylene moiety from the fragment ion at *miz* 137.0597 $[C_8H_9O_2]^+$ (Wolf, Huschka, Raith, Wohlrab, & Neubert, 1999). Moreover, the ion at *miz* 122.0362 $[C_7H_6O_2]^{+\cdot}$ can be further fragmented through neutral losses of CO (27.9943 Da) to form both fragment ions at *miz* 94.0413 $[C_7H_6O_2-CO]^{+\cdot}$ and *miz* 66.0464 $[C_7H_6O_2-C_2O_2]^{+\cdot}$. 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 $[C_{11}H_{13}]^+$ (*miz* 145.1012), $[C_9H_{11}]^+$ (*miz* 119.0855), and $[C_8H_9]^+$ (*miz* 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 $[C_{15}H_{21}O_2]^+$ (*miz* 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 $[C_8H_{13}]^+$ (*miz* 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

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 *miz* (Rel. Ab.)	Ion Assigment	Accurate mass error (ppm)	Fragment ion (*miz*)	Ion Assigment	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	$[C_7H_6O_2]^{+\cdot}$	0.0
					94.0417	$[C_7H_6O_2-CO]^{+\cdot}$	4.2
					66.0465	$[C_7H_6O_2-C_2O_2]^{+\cdot}$	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	$[C_7H_6O_2]^{+\cdot}$	0.0
		137.0595 (75)	$[C_8H_9O_2]^+$	−1.4	94.0417	$[C_7H_6O_2-CO]^{+\cdot}$	4.2
					66.0465	$[C_7H_6O_2-C_2O_2]^{+\cdot}$	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	$[C_7H_6O_2]^{+\cdot}$	0.0
					94.0417	$[C_7H_6O_2-CO]^{+\cdot}$	4.2
					66.0465	$[C_7H_6O_2-C_2O_2]^{+\cdot}$	1.5
NDCT	5.32	137.0596 (100)	$[C_8H_9O_2]^+$	−0.7	137.0595	$[C_8H_9O_2]^+$	−1.5
					122.0362	$[C_7H_6O_2]^{+\cdot}$	0.8
					94.0417	$[C_7H_6O_2-CO]^{+\cdot}$	4.2
					66.0465	$[C_7H_6O_2-C_2O_2]^{+\cdot}$	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-CR PT	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

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, *viv*) 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, *viv*). Moreover, the mobile phase eluotropic strength had to be increased at the end of the chromatographic run using acetonitrile acetone (50 50, *viv*) 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 min, 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 *miz*. 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 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 the concentration range, 0.001–10 mg kg^{-1} for most of the compounds and 0.1–10 mg kg^{-1} for β–CR PT and LUT, was satisfactory and showed correlation coefficients (R^2) higher than 0.998. ILODs ranged from 0.001 to 0.025 mg kg^{-1} for most of the target compounds, although for β–CR PT and LUT values were slightly higher (0.1 and 0.25 mg kg^{-1}, respectively). In terms of RSD and based on concentration values, run-to-run and day-to-day precision were always lower than 15 and 10 , respectively. Moreover, the trueness, based on the same concentration values, showed relative errors below 10 . These results demonstrated the good instrumental performance of the developed UHPLC-APCI-HRMS method for the determination of capsaicinoids and carotenoids.

Besides, before the determination of capsaicinoids and carotenoids by UHPLC–APCI–HRMS in paprika, samples were submitted to a solid-liquid extraction. Because of the differences in the physicochemical properties of both families of compounds, three commonly used solvents, methanol, acetonitrile, and acetone, as well as mixtures of them, were evaluated to achieve the most effective simultaneous extraction of target compounds. It was found that acetonitrile had less effectiveness in

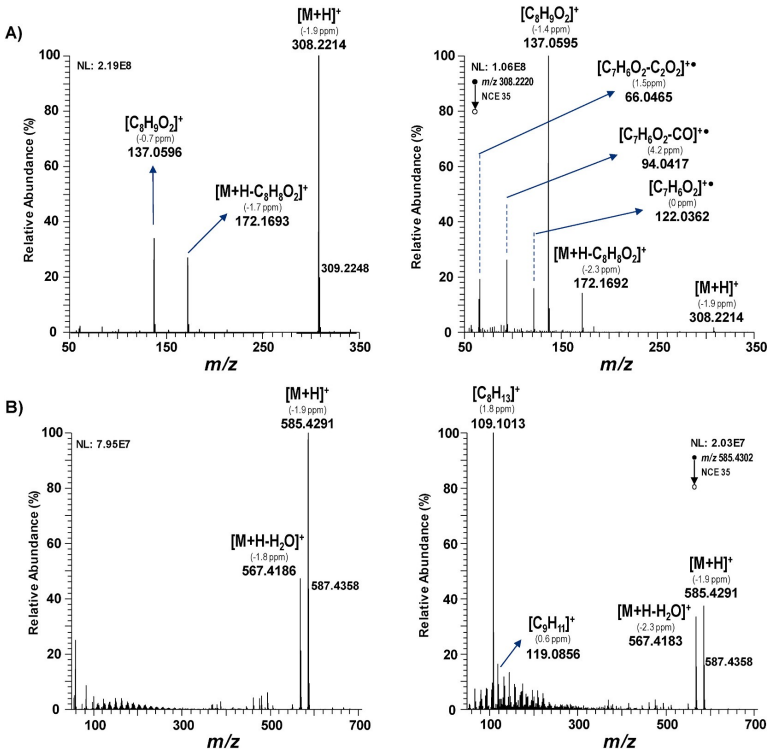


Fig. 2. HRMS spectrum and AIF (HRMS) spectrum of (A) DC and (B) CT.

extracting carotenoids than both pure acetone and the mixture methanol acetone. Moreover, pure methanol extracted more efficiently capsaicinoids, than pure acetonitrile or acetone. Nevertheless, the combination of both methanol and acetone seemed to improve the solubility of these compounds, and thus, as a compromise, a mixture methanol acetone (1 1, *viv*) was chosen as the most effective solvent for the simultaneous extraction of both capsaicinoids and carotenoids (*section 2 3*.) in agreement with Nagy et al. who proposed a similar solvent mixture (Nagy et al., 2017). Using the proposed extraction procedure, estimated MLODs ranged from 0.06 to 1.5 mg kg^{-1} for most of the analytes, except for β−CR PT and LUT, which were 6.1 and 15.3 mg kg^{-1}, respectively. While, MLOQs were comprised between 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 kg^{-1}, and hence a higher capsaicinoid ratio (40−90), than sweet and bittersweet samples, 9 ± 5 and 31 ± 32 mg kg^{-1}, respectively. This result was expected since these target compounds are responsible for the characteristic hot taste (de Sá Mendes & Branco de Andrade Goncalves, 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 kg^{-1}, respectively; whereas hot Hungary samples contained 719 ± 192 mg kg^{-1}. Besides, in accordance to Giuffrida et al. (Giuffrida et al., 2013), β−CAR was found to be the most predominant carotenoid (15−510 mg kg^{-1}) in all samples, followed by β−CR PT (25−360 mg kg^{-1}), and CT (6−270 mg kg^{-1}). Intead, VIO and CR occurred at lower concentrations (4.2−42 mg 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, usof, ahaya, 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

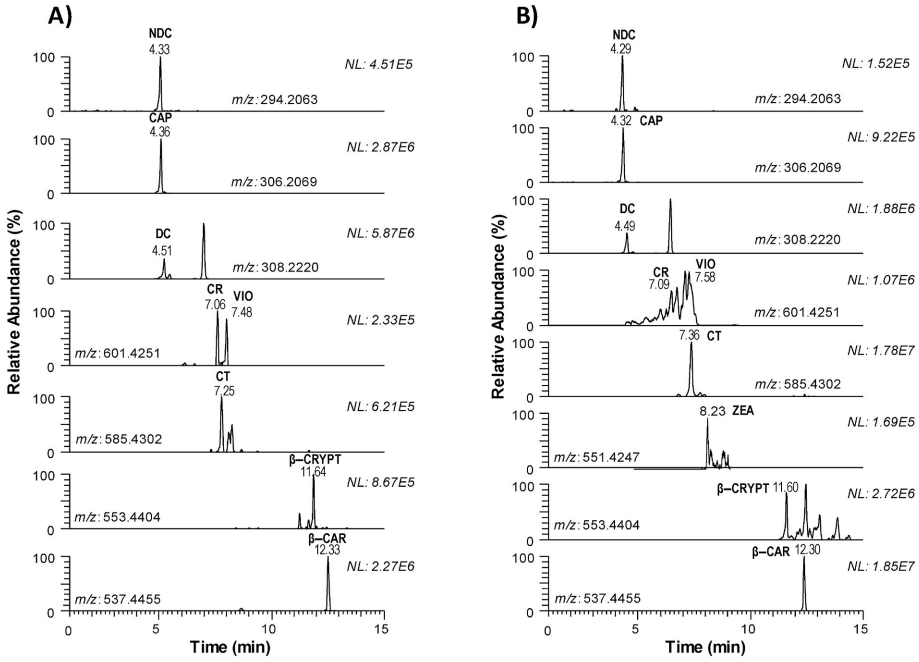


Fig. . UHPLC−APCI−HRMS capsaicinoid and carotenoid profile chromatograms of sweet paprika samples from (A) *Murcia,* sample MS9, and (B) Hungary, sample HS5.

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

PLS−DA was chosen as the chemometric technique to conduct the classificatory analysis. A first PLS−DA model was built, which included all the paprika samples under study, according to both origins, and type. Thus, a 136×10 X-data matrix and a -data matrix, assigning samples to nine classes, were used. Fig. 4 shows the corresponding scores plot of LV1 *vs* LV2 (two LVs, explaining the 18.29 -variance, were chosen for constructing the PLS−DA model), where remarkable discrimination between types could be seen. In this line, sweet samples were located on the upper side of the plot, whereas the hot ones on the bottom. Variable importance in projection (VIP) values indicated that this separation was mainly because of CAP, NDC, and DC contents. However, bittersweet *La Vera* samples did not present significant differences with *La Vera* sweet ones, so they were considered both as sweet in the following chemometric studies. Regarding the production area, Hungary paprika samples were clearly distinguished in the right part of the plot (displaying positive LV1 scores values) from the 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 -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 ×0, 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. Conclu ion

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,

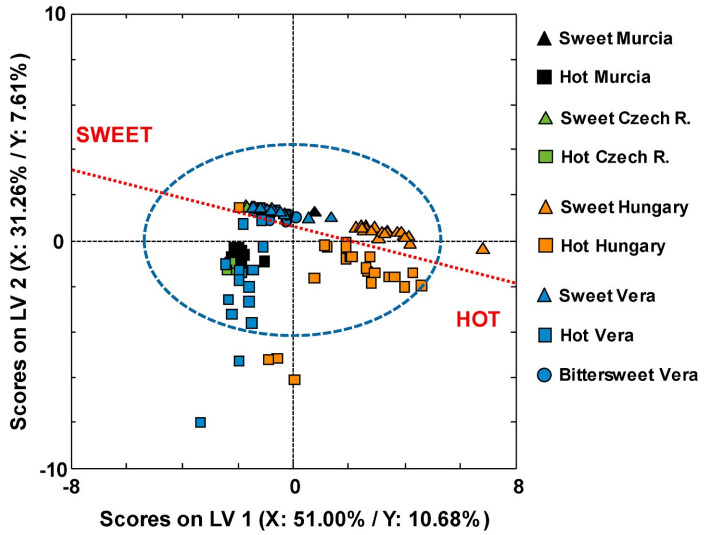


Fig. 4. PLS-DA Scores plot of LV1 *vs*. LV2, using the UHPLC-HRMS capsaicinoid and carotenoid profiling for the classification of all the paprika samples tested.

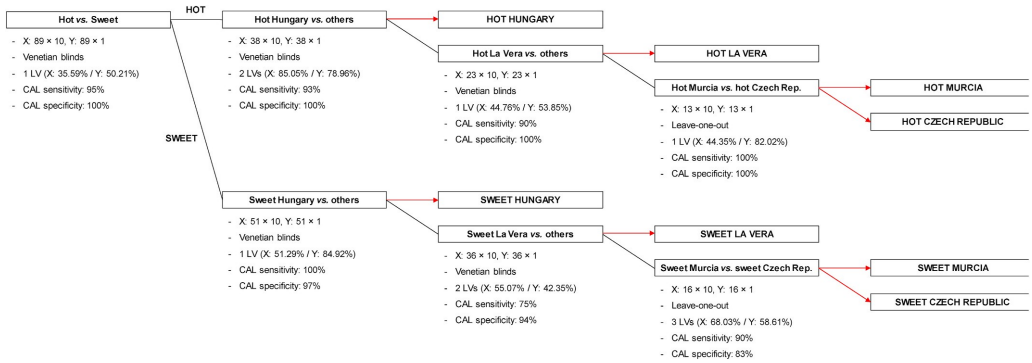


Fig. 5. Classification decision tree built by HMB for paprika geographical origin authentication by means of PLS-DA models. Dimensions, CV used method, LVs, and sensitivity and specificity of the model are detailed.

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.

C ediT author hip contribution tatement

Ane Arrizabalaga-Larrañaga: Conceptualization, Methodology, Investigation, Validation, Formal analysis, Writing - original draft, preparation. **Guillem Campmajó:** Conceptualization, Methodology, Investigation, Validation, Formal analysis, Writing - original draft, preparation. **Javier Saurina:** Writing - review & editing. **O car Núñez:** Methodology, Supervision, Writing - review & editing. **Franci co Javier Santo :** Methodology, Supervision, Writing - review & editing. **Encarnación Moyano:** Funding acquisition, Project administration, Methodology, Supervision, Writing - review & editing.

Declaration of competing intere t

D The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ac no ledgement

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https //doi.org/10.1016/j.lwt.2020.110533.

eference

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 & 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 & 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., et al. (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/10.5772/57353

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

Cetó, X., Serrano, N., Agagó, M., Gámez, A., Esteban, M., Díaz-Cruz, J. M., et al. (2018). Determination of HPLC-UV fingerprints of Spanish paprika (Capsicum annuum L.) for its classification by linear discriminant analysis. *Sensors, 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

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

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/.

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/.

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., et al. (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., usof, N. A., ahaya, 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. ., Jeong, J. ., Park, J. M., Kim, B. H., Kwon, K., et al. (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, ., & 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 *Hand ook of chemometrics and qualimetrics* (1st ed.). Elsevier. 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 & 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-Dorđević, J. B., et al. (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 & 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, 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

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

Reilly, C. A., Ehlhardt, W. J., Jackson, D. A., Kulanthaivel, P., Mutlib, A. E., Espina, R. J., et al. (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

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

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., Agagó, M., Gámez, A., Ariño, C., et al. (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

Štursa, V., & 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 & Related Technologies, 22*(4), 531–539. https //doi.org/10.1081/JLC-100101678

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	CRYPT	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·kg^{-1}) of capsaicinoids and carotenoids determined in paprika samples.

Sample	NDC	CAP	DC	NDCT	CR	VIO	CT	ZEA	CRYPT	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·kg^{-1}) of capsaicinoids and carotenoids determined in paprika samples.

Sample	NDC	CAP	DC	NDCT	CR	VIO	CT	ZEA	CRYPT	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·kg^{-1}) of capsaicinoids and carotenoids determined in paprika samples.

Sample	NDC	CAP	DC	NDCT	CR	VIO	CT	ZEA	CRYPT	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·kg^{-1}) of capsaicinoids and carotenoids determined in paprika samples.

Sample	NDC	CAP	DC	NDCT	CR	VIO	CT	ZEA	CRYPT	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	CRYPT	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 (L CAPS), total carotenoid content (L CAR), and their respective sum (L CAPS + L CAR), expressed as mean ± standard deviation, obtained for the analyzed paprika samples according to their geographical origin and flavor variety.

	Hot			Sweet			Bittersweet		
	L CAPS	L CAR	L CAPS + CAR	L CAPS	L CAR	L CAPS + CAR	L CAPS	L CAR	L 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

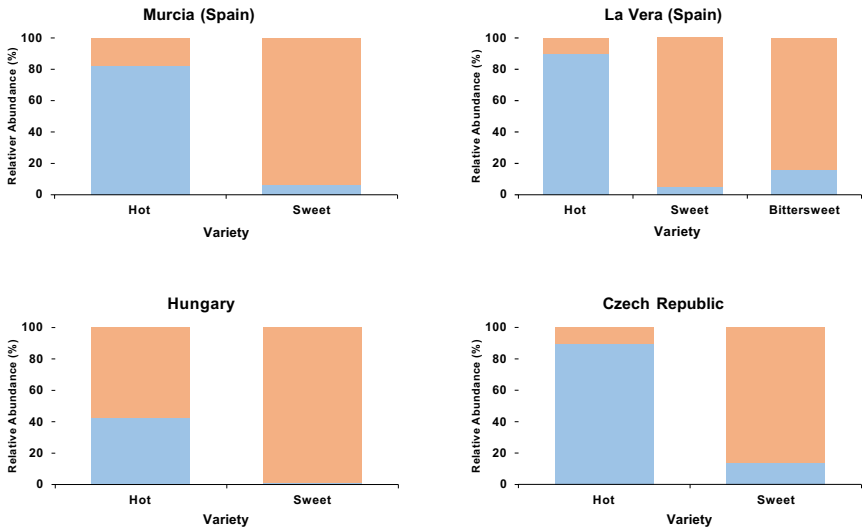


Figure S1: Capsaicinoid (blue) and carotenoid (orange) distribution of Paprika from different origins and varieties.

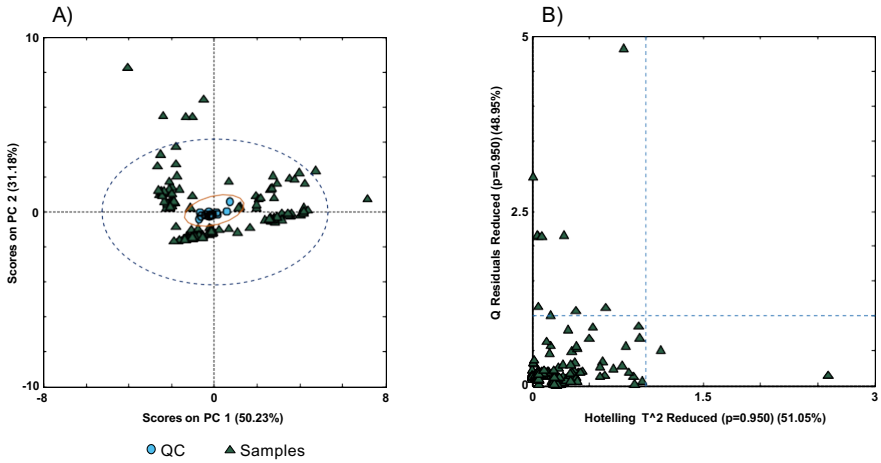


Figure S2: (A) PCA Scores plot of PC1 *vs*. PC2, showing a correct behavior of QC samples. (B) Hotelling T2 *vs*. Q residual values plot for the detection of outlier samples.

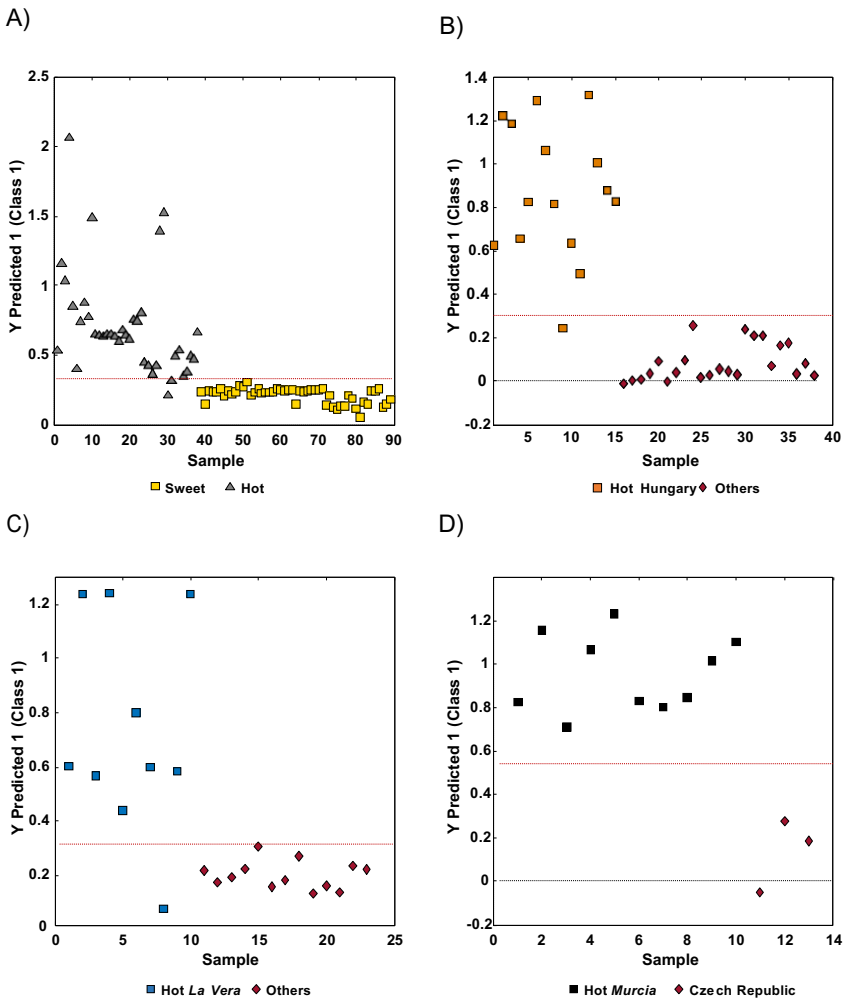


Figure S3: Classification plot depicting Samples *vs*. Y Predicted 1 Scores plot for the PLS-DA calibration models of (A) hot *vs.* sweet, (B) hot Hungary *vs.* others, (C) hot *La Vera vs.* others, and (D) hot *Murcia vs.* Czech Republic.

2.3. DISCUSSION

This chapter's results (Publication I and Publication II) present novel data on the metabolomic composition of paprika samples and the application of profiling LC−MS-based methods for their chemometric analysis. In this context, considering that the individual conclusions of each study are already detailed in the corresponding scientific article, this section discusses general findings and compares some of the obtained results.

In both studies, the instrumental quality parameters —including the instrumental limit of detection (iLODs), the instrumental limit of quantification (iLOQs), the linearity, the precision (both run-to-run and day-to-day), and the trueness— were evaluated. In general, each method obtained satisfactory results for most of these parameters. Nevertheless, focusing on the sensitivity, it is noteworthy to point out that different approaches were used to estimate the iLODs because of the mass analysers' characteristics. On the one hand, in Publication I, where a QqQ was used under the MRM mode, this parameter was calculated based on a signal-to-noise ratio of 3:1. Hence, the LC−MS/MS method developed for determining 36 phenolic compounds presented iLODs ranging from 0.01 $\mu g \cdot kg^{-1}$ to 1.4 $mg \cdot kg^{-1}$, with 28 of the studied compounds showing values below 0.3 $mg \cdot kg^{-1}$. In this line, the considerable differences within analytes' detectability could be attributed to their chemical structural differences since several phenolic classes were encompassed (*i.e.*, phenolic acids, flavonoids, stilbenes, other phenolic, and derivatives), leading to diverse ESI efficiencies.

On the other hand, in Publication II, the iLODs were estimated as the smallest analyte concentration providing a detectable well-resolved chromatographic peak. This fact was due to the lack of baseline noise in the LC−HRMS extracted ion chromatograms because of the Q-Orbitrap's high mass accuracy and resolving power. As a result, the LC−HRMS method for capsaicinoid and carotenoid profiling reached iLOD values from 0.001 to 0.25 $mg \cdot kg^{-1}$. Interestingly, although the applied APCI and MS

working conditions had been previously developed by Arrizabalaga-Larrañaga *et al*. for the carotenoid determination (alongside chlorophylls) [52], capsaicinoids were the ones reaching lower detection levels and, hence, presenting better APCI efficiencies.

Hence, after the method development and the quality parameters' evaluation, paprika samples were analysed by the optimised methodologies. In this context, the accurate quantification of secondary metabolites in food products is challenging as they are generally endogenous compounds and, therefore, no blank samples are available. In this line, the lack of blank samples prevents accurate quantification through matrix-matched calibration. Besides, other options, such as standard addition calibration or isotope dilution mass spectrometry (IDMS), cannot be effectively implemented for the proposed applications herein due to time and money requirements, respectively. In addition, in the case of IDMS, no internal labelled reference standards are available for all the targeted compounds. Under these circumstances, the semi-quantification through external calibration is currently the most extended practice for determining these compounds.

In both studies, in an attempt to evaluate ion suppression or ion enhancement for those targeted analytes co-eluting, the LC−MS peak area signals resulting from the injection of the individual reference standard solution or the mixture of the co-eluting analytes were compared. As a result, none of these effects were significantly reported (variations below 10%). Nevertheless, in Publication II, the matrix effect was also estimated since, aside from the targeted co-eluting compounds, other matrix-related metabolites could alter the ionisation step (either with ESI or APCI). In this case, values ranging from 10 to 50% were obtained depending on the analyte. Therefore, although it is evident that the matrix effect influences the accuracy of the quantification, it can also be an indirect source of variation between the sample classes under study.

Furthermore, as reported in Publication I, 20 of the 36 targeted phenolic compounds were detected and semi-quantified in the paprika samples under study through the proposed LC−MS/MS method. Among these compounds, 16 were detected in all the analysed samples (*i.e.*, *La Vera*, *Murcia*, and the Czech Republic). In contrast, betulinic acid was specific from *La Vera* samples, homogentisic acid from the Czech Republic samples, umbelliferone from the hot Czech Republic samples, and homovanillic acid from *La Vera* and the sweet-smoked Czech Republic samples. In this line, some of the detected phenolic compounds had been observed in previous studies. For instance, in the Mudrić *et al*. study [56], which contained the targeted and suspected determination of polyphenols by LC−HRMS in Serbian paprika, six phenolic compounds were also found in all the analysed samples (*i.e.*, vanillin, rutin, and caffeic, *p*-coumaric, ferulic, and sinapic acids), two in most of them (*i.e.*, gallic and chlorogenic acids), and two only in some (*i.e.*, quercetin and umbelliferone). Moreover, interestingly, while Mudrić *et al*. reported concentration values for *p*-hydroxybenzoic and syringic acids in Serbian paprika samples, they were not detected in Spanish nor Czech paprika samples.

Instead, capsaicinoid and carotenoid content in paprika is already compared with other studies in the literature in Publication II. However, Table 2.3 contains the pungency levels of the analysed paprika samples, expressed as Scoville heat units (SHU), obtained by multiplying the capsaicin and dihydrocapsaicin concentration ($\mu g \cdot kg^{-1}$) by 15. In this line, according to the Scoville scale [57], the obtained average SHU values classified the samples according to their pungency level. Thus, independently of the geographical origin of the sample, while sweet and bittersweet paprika were encompassed in the '*non-pungent*' level (0 – 700 SHU), hot samples were included in the '*moderately pungent*' one (3,000 – 25,000 SHU). Furthermore, considering that the SHU values depend on secondary metabolites naturally occurring in red pepper, the significant variations within sample types can be considered acceptable.

Table 2.3. Pungency of the analysed paprika samples expressed in SHU.

	La Vera			*Murcia*		The Czech Republic		Hungary	
	Hot	**Sweet**	**Bittersweet**	**Hot**	**Sweet**	**Hot**	**Sweet**	**Hot**	**Sweet**
Pungency (SHU)	13,477 ± 7,837	134 ± 95	442 ± 446	7,871 ± 773	145 ± 18	9,014 ± 402	161 ± 21	7,079 ± 6,493	111 ± 68

Regarding the chemometric analysis, given the results described in Publication I and II, the proposed phenolic compounds seemed better chemical descriptors for classifying paprika samples according to their geographical origin than the capsaicinoids and carotenoids profile. However, to compare both profiles under similar conditions, a PLS-DA model —encompassing *La Vera*, *Murcia*, and the Czech Republic samples— was built in each case (Hungary paprika samples were not considered for this comparison since they were not available when performing the analyses corresponding to Publication I). In this context, Figure 2.1 presents the results obtained after external validation of the PLS-DA models, using 70% of the samples as the calibration set and 30% as the prediction set. As reported in Publication I, the determined phenolic LC−MS/MS profile allowed excellent classification of paprika samples according to the three regions under study. Instead, the capsaicinoid and carotenoid profile showed worse discriminant ability through a unique multi-class PLS-DA model, even at the calibration set.

Therefore, considering that, as introduced in Section 1.2.6, the number of studied classes highly influences a chemometric discriminant analysis model, Publication II proposed a classification decision tree strategy to improve the observed classification results. As a result, breaking down the classification into successive binary PLS-DA models improved the final performance in terms of class sensitivity and specificity, as well as classification accuracy. Nevertheless, according to the results, the targeted phenolic and polyphenolic compounds have proven to be better chemical descriptors

to assess the geographical origin of paprika than the determined capsaicinoids and carotenoids.

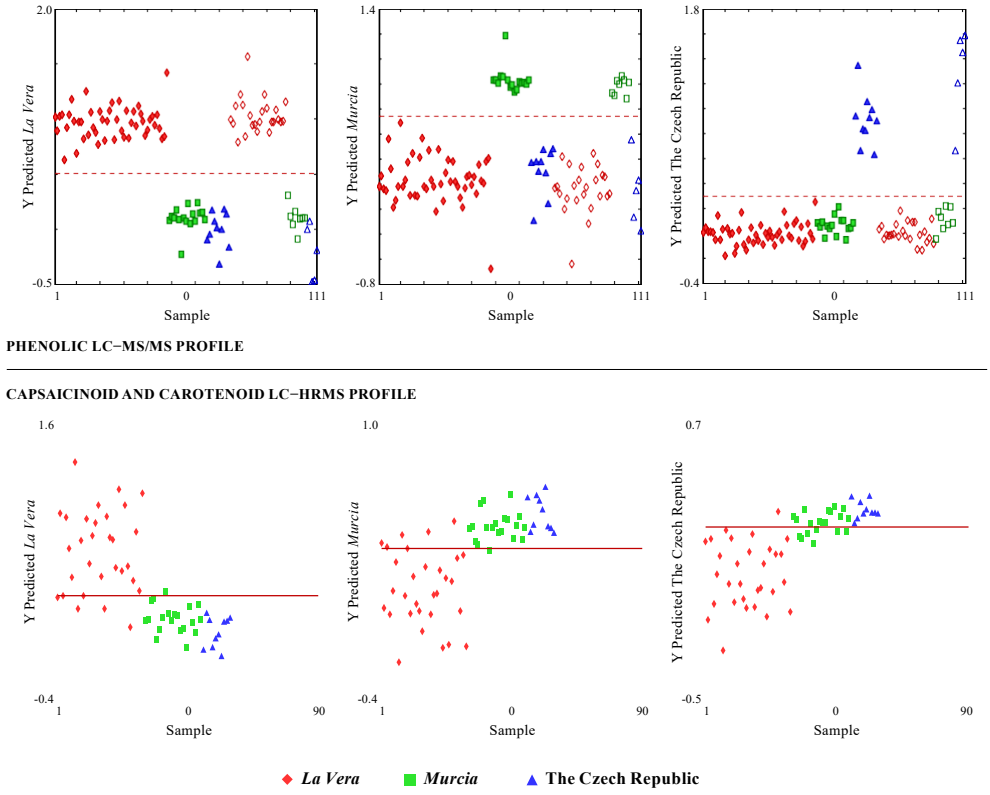


Figure 2.1. External validation results for the PLS-DA model built using the phenolic LC−MS/MS data

Finally, this chapter has stated and demonstrated both the advantages and the disadvantages of metabolomic profiling-based LC−MS approaches for food authentication, specifically for paprika classification according to its geographical origin. In summary, targeted metabolites can successfully act as chemical markers of a particular class of samples (*e.g.*, a geographical origin, production system, botanical origin, or cultivar). In this line, in contrast to non-targeted methods, targeted ones focusing on known and previously reported chemical markers can be easily transferred between laboratories and analytically validated following a harmonised

protocol [58]. However, guidelines for the validation of multivariate analysis still must be developed. Moreover, considering that many factors influence targeted markers' content (*e.g.*, cultivation, soil conditions, feeding, or climate), their specificity and universality are challenging. Therefore, the current knowledge of known and reported chemical markers for food authentication is scarce, and more research on that objective is required [6].

2.4. REFERENCES

[1] Z. Liu, M. Zhang, P. Chen, J.M. Harnly, J. Sun, Mass spectrometry-based nontargeted and targeted analytical approaches in fingerprinting and metabolomics of food and agricultural research, J. Agric. Food Chem. 70 (2022) 11138–11153. https://doi.org/10.1021/acs.jafc.2c01878.

[2] G. Campmajó, N. Núñez, O. Núñez, The role of liquid chromatography-mass spectrometry in food integrity and authenticity, in: G.S. Kamble (Ed.), Mass Spectrom. - Futur. Perceptions Appl., IntechOpen, 2019: pp. 1–22. https://doi.org/10.5772/intechopen.85087.

[3] S. Medina, R. Perestrelo, P. Silva, J.A.M. Pereira, J.S. Câmara, Current trends and recent advances on food authenticity technologies and chemometric approaches, Trends Food Sci. Technol. 85 (2019) 163–176. https://doi.org/10.1016/j.tifs.2019.01.017.

[4] V.G. Zaikin, R.S. Borisov, Mass spectrometry as a crucial analytical basis for omics sciences, J. Anal. Chem. 76 (2021) 1567–1587. https://doi.org/10.1134/S1061934821140094.

[5] K.J. Margolin Eren, O. Elkabets, A. Amirav, A comparison of electron ionization mass spectra obtained at 70 eV, low electron energies, and with cold EI and their NIST library identification probabilities, J. Mass Spectrom. 55 (2020) e4646. https://doi.org/10.1002/jms.4646.

[6] X. Dou, L. Zhang, R. Yang, X. Wang, L. Yu, X. Yue, F. Ma, J. Mao, X. Wang, W. Zhang, P. Li, Mass spectrometry in food authentication and origin traceability, Mass Spectrom. Rev. e21779 (2022) 1–36. https://doi.org/10.1002/mas.21779.

[7] L.M. Crawford, J.L. Janovick, N. Carrasquilla-Garcia, E. Hatzakis, S.C. Wang, Comparison of DNA analysis, targeted metabolite profiling, and non-targeted NMR fingerprinting for differentiating cultivars of processed olives, Food Control. 114 (2020) 107264. https://doi.org/10.1016/j.foodcont.2020.107264.

[8] L. Zhang, Z. Yuan, P. Li, X. Wang, J. Mao, Q. Zhang, C. Hu, Targeted multivariate adulteration detection based on fatty acid profiles and Monte Carlo one-class partial least squares, Chemom. Intell. Lab. Syst. 169 (2017) 94–99. https://doi.org/10.1016/j.chemolab.2017.09.002.

[9] D. Amorello, S. Orecchio, A. Pace, S. Barreca, Discrimination of almonds (Prunus dulcis) geographical origin by minerals and fatty acids profiling, Nat. Prod. Res. 30 (2016) 2107–2110. https://doi.org/10.1080/14786419.2015.1107559.

[10] B. Mehari, M. Redi-Abshiro, B.S. Chandravanshi, S. Combrinck, R. McCrindle, M. Atlabachew, GC-MS profiling of fatty acids in green coffee (Coffea arabica L.) beans and chemometric modeling for tracing geographical origins from Ethiopia, J. Sci. Food Agric. 99 (2019) 3811–3823. https://doi.org/10.1002/jsfa.9603.

[11] V. Giannetti, M.B. Mariani, P. Mannino, Characterization of the authenticity of Pasta di Gragnano protected geographical indication through flavor component analysis by

gas chromatography-mass spectrometry and chemometric tools, J. AOAC Int. 99 (2016) 1279–1286. https://doi.org/10.5740/jaoacint.16-0137.

[12] T.J. Kim, J.G. Park, S.K. Ahn, K.W. Kim, J. Choi, H.Y. Kim, S.H. Ha, W.D. Seo, J.K. Kim, Discrimination of adzuki bean (Vigna angularis) geographical origin by targeted and non-targeted metabolite profiling with gas chromatography time-of-flight mass spectrometry, Metabolites. 10 (2020) 112. https://doi.org/10.3390/metabo10030112.

[13] O. Núñez, H. Gallart-Ayala, Core-shell column technology in fast liquid chromatography, in: O. Núñez, H. Gallart-Ayala, C.P.B. Martins, P. Lucci (Eds.), Fast Liq. Chromatogr. Spectrom. Methods Food Environ. Anal., Imperial College Press, 2015: pp. 33–56. https://doi.org/10.1142/9781783264940_0002.

[14] L. Righetti, G. Paglia, G. Galaverna, C. Dall'Asta, Recent advances and future challenges in modified mycotoxin analysis: Why HRMS has become a key instrument in food contaminant research, Toxins (Basel). 8 (2016) 361. https://doi.org/10.3390/toxins8120361.

[15] M. Azir, S. Abbasiliasi, T.A.T. Ibrahim, Y.N.A. Manaf, A.Q. Sazili, S. Mustafa, Detection of lard in cocoa butter—its fatty acid composition, triacylglycerol profiles, and thermal characteristics, Foods. 6 (2017) 98. https://doi.org/10.3390/foods6110098.

[16] H.S. Green, X. Li, M. De Pra, K.S. Lovejoy, F. Steiner, I.N. Acworth, S.C. Wang, A rapid method for the detection of extra virgin olive oil adulteration using UHPLC-CAD profiling of triacylglycerols and PCA, Food Control. 107 (2020) 106773. https://doi.org/10.1016/j.foodcont.2019.106773.

[17] M. Sud, E. Fahy, D. Cotter, A. Brown, E.A. Dennis, C.K. Glass, A.H. Merrill, R.C. Murphy, C.R.H. Raetz, D.W. Russell, S. Subramaniam, LMSD: LIPID MAPS structure database, Nucleic Acids Res. 35 (2006) 527–532. https://doi.org/10.1093/nar/gkl838.

[18] S. Imperiale, E. Kaneppele, K. Morozova, F. Fava, D. Martini-Lösch, P. Robatscher, G. Peratoner, E. Venir, D. Eisenstecken, M. Scampicchio, Authenticity of hay milk vs. milk from maize or grass silage by lipid analysis, Foods. 10 (2021) 2926. https://doi.org/10.3390/foods10122926.

[19] I. Samarra, C. Masdevall, E. Foguet-Romero, M. Guirro, M. Riu, P. Herrero, N. Canela, A. Delpino-Rius, Analysis of oxylipins to differentiate between organic and conventional UHT milks, Food Chem. 343 (2021) 128477. https://doi.org/10.1016/j.foodchem.2020.128477.

[20] P. Lucci, J. Saurina, O. Núñez, Trends in LC-MS and LC-HRMS analysis and characterization of polyphenols in food, TrAC - Trends Anal. Chem. 88 (2017) 1–24. https://doi.org/10.1016/j.trac.2016.12.006.

[21] S. Piovesana, C. Cavaliere, A. Cerrato, C.M. Montone, A. Laganà, A.L. Capriotti, Developments and pitfalls in the characterization of phenolic compounds in food: From targeted analysis to metabolomics-based approaches, TrAC - Trends Anal. Chem. 133 (2020) 116083. https://doi.org/10.1016/j.trac.2020.116083.

[22] A. Rivera-Pérez, P. García-Pérez, R. Romero-González, A. Garrido Frenich, L. Lucini, An untargeted strategy based on UHPLC-QTOF-HRMS metabolomics to identify markers revealing the terroir and processing effect on thyme phenolic profiling, Food Res. Int. 162 (2022) 112081. https://doi.org/10.1016/j.foodres.2022.112081.

[23] J. Saurina, Characterization of wines using compositional profiles and chemometrics, TrAC - Trends Anal. Chem. 29 (2010) 234–245. https://doi.org/10.1016/j.trac.2009.11.008.

[24] R. González-Domínguez, A. Sayago, I. Akhatou, Á. Fernández-Recamales, Multi-chemical profiling of strawberry as a traceability tool to investigate the effect of cultivar and cultivation conditions, Foods. 9 (2020) 96. https://doi.org/10.3390/foods9010096.

[25] N.P. Kalogiouri, V.F. Samanidou, HPLC fingerprints for the characterization of walnuts and the detection of fraudulent incidents, Foods. 10 (2021) 2145. https://doi.org/10.3390/foods10092145.

[26] Z. Li, J. Zhang, Q. Meng, L. Yang, M. Qiu, Y. Li, S. Yao, W. Wei, C. Yao, Q. Bi, J. Li, D. Guo, The content and distribution of 18 phenolic compounds in 462 batches of edible flowers from 73 species commercially available in China, Food Res. Int. 166 (2023) 112590. https://doi.org/10.1016/j.foodres.2023.112590.

[27] L.R. Silva, A.C. Gonçalves, A.R. Nunes, G. Alves, Authentication of honeys from Caramulo region (Portugal): Pollen spectrum, physicochemical characteristics, mineral content, and phenolic profile, J. Food Sci. 85 (2020) 374–385. https://doi.org/10.1111/1750-3841.15023.

[28] I.K. Karabagias, M. V. Vavoura, C. Nikolaou, A. V. Badeka, S. Kontakos, M.G. Kontominas, Floral authentication of Greek unifloral honeys based on the combination of phenolic compounds, physicochemical parameters and chemometrics, Food Res. Int. 62 (2014) 753–760. https://doi.org/10.1016/j.foodres.2014.04.015.

[29] M. Jungen, P. Lotz, C.-D. Patz, C.B. Steingass, R. Schweiggert, Coumarins, psoralens, and quantitative 1H-NMR spectroscopy for authentication of lemon (Citrus limon [L.] Burm.f.) and Persian lime (Citrus × latifolia [Yu.Tanaka] Tanaka) juices, Food Chem. 359 (2021) 129804. https://doi.org/10.1016/j.foodchem.2021.129804.

[30] L. Puigventós, M. Navarro, É. Alechaga, O. Núñez, J. Saurina, S. Hernández-Cassou, L. Puignou, Determination of polyphenolic profiles by liquid chromatography-electrospray-tandem mass spectrometry for the authentication of fruit extracts, Anal. Bioanal. Chem. 407 (2015) 597–608. https://doi.org/10.1007/s00216-014-8298-2.

[31] F.I. Brigante, A. Lucini Mas, N.B. Pigni, D.A. Wunderlin, M. V. Baroni, Targeted metabolomics to assess the authenticity of bakery products containing chia, sesame and flax seeds, Food Chem. 312 (2020) 126059. https://doi.org/10.1016/j.foodchem.2019.126059.

[32] N. Núñez, O. Vidal-Casanella, S. Sentellas, J. Saurina, O. Núñez, Characterization, classification and authentication of turmeric and curry samples by targeted LC-HRMS polyphenolic and curcuminoid profiling and chemometrics, Molecules. 25 (2020)

2942. https://doi.org/10.3390/molecules25122942.

[33] S. Barbosa, N. Pardo-Mates, M. Hidalgo-Serrano, J. Saurina, L. Puignou, O. Núnez, Detection and quantitation of frauds in the authentication of cranberry-based extracts by UHPLC-HRMS (Orbitrap) polyphenolic profiling and multivariate calibration methods, J. Agric. Food Chem. 66 (2018) 9353–9365. https://doi.org/10.1021/acs.jafc.8b02855.

[34] J. Zhang, Q. Yu, H. Cheng, Y. Ge, H. Liu, X. Ye, Y. Chen, Metabolomic approach for the authentication of berry fruit juice by liquid chromatography quadrupole time-of-flight mass spectrometry coupled to chemometrics, J. Agric. Food Chem. 66 (2018) 8199–8208. https://doi.org/10.1021/acs.jafc.8b01682.

[35] S. Drakopoulou, E. Orfanakis, I. Karagiannaki, F. Gaitis, S. Skoulika, A. Papaioannou, G. Boukouvalas, G. Petropoulos, V. Katsoudas, R. Kontzedaki, A. Philippidis, A. Zoumi, M. Dasenaki, N.S. Thomaidis, M. Velegrakis, Comparative evaluation of different targeted and untargeted analytical approaches to assess Greek extra virgin olive oil quality and authentication, Molecules. 27 (2022) 1350. https://doi.org/10.3390/molecules27041350.

[36] Đ.D. Krstić, P.M. Ristivojević, U.M. Gašić, M. Lazović, M.M. Fotirić Akšić, J. Milivojević, G.E. Morlock, D.M. Milojković-Opsenica, J.Đ. Trifković, Authenticity assessment of cultivated berries via phenolic profiles of seeds, Food Chem. 402 (2023) 134184. https://doi.org/10.1016/j.foodchem.2022.134184.

[37] B. Mehari, M. Redi-Abshiro, B.S. Chandravanshi, S. Combrinck, M. Atlabachew, R. McCrindle, Profiling of phenolic compounds using UPLC-MS for determining the geographical origin of green coffee beans from Ethiopia, J. Food Compos. Anal. 45 (2016) 16–25. https://doi.org/10.1016/j.jfca.2015.09.006.

[38] M. Yang, X. Tian, M. Zhang, J. Wei, Y. Niu, J. Hou, Y. Jin, Y. Du, A holistic comparison of flavor signature and chemical profile in different harvesting periods of Chrysanthemum morifolium Ramat. based on metabolomics combined with bioinformatics and molecular docking strategy, RSC Adv. 12 (2022) 34971–34989. https://doi.org/10.1039/d2ra05698d.

[39] G. Xue, S. Su, P. Yan, J. Shang, J. Wang, C. Yan, J. Li, Q. Wang, Y. Du, L. Cao, H. Xu, Quality control of Zingiberis Rhizoma and its processed products by UHPLC-Q-TOF/MS-based non-targeted metabonomics combining with SIBDV method, Food Res. Int. 154 (2022) 111021. https://doi.org/10.1016/j.foodres.2022.111021.

[40] B. Senizza, G. Rocchetti, S. Ghisoni, M. Busconi, M. De Los Mozos Pascual, J.A. Fernandez, L. Lucini, M. Trevisan, Identification of phenolic markers for saffron authenticity and origin: An untargeted metabolomics approach, Food Res. Int. 126 (2019) 108584. https://doi.org/10.1016/j.foodres.2019.108584.

[41] R.L. Crizel, J.F. Hoffmann, G.P. Zandoná, P.M.S. Lobo, R.O. Jorge, F.C. Chaves, Characterization of extra virgin olive oil from Southern Brazil, Eur. J. Lipid Sci. Technol. 122 (2020) 1900347. https://doi.org/10.1002/ejlt.201900347.

[42] A.M. Mustafa, S. Angeloni, D. Abouelenein, L. Acquaticci, J. Xiao, G. Sagratini, F. Maggi, S. Vittori, G. Caprioli, A new HPLC-MS/MS method for the simultaneous

determination of 36 polyphenols in blueberry, strawberry and their commercial products and determination of antioxidant activity, Food Chem. 367 (2022) 130743. https://doi.org/10.1016/j.foodchem.2021.130743.

[43] S. Barbosa, J. Saurina, L. Puignous, O. Núñez, 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 (2020) 1623. https://doi.org/10.3390/molecules25071623.

[44] B. Senizza, P. Ganugi, M. Trevisan, L. Lucini, Combining untargeted profiling of phenolics and sterols, supervised multivariate class modelling and artificial neural networks for the origin and authenticity of extra-virgin olive oil: A case study on Taggiasca Ligure, Food Chem. 404 (2023) 134543. https://doi.org/10.1016/j.foodchem.2022.134543.

[45] J.A. Rothwell, J. Perez-Jimenez, V. Neveu, A. Medina-Remón, N. M'Hiri, P. García-Lobato, C. Manach, C. Knox, R. Eisner, D.S. Wishart, A. Scalbert, Phenol-Explorer 3.0: a major update of the Phenol-Explorer database to incorporate data on the effects of food processing on polyphenol content, Database. 2013 (2013) bat070. https://doi.org/10.1093/database/bat070.

[46] F. Ma, Q. Yang, B. Matthäus, P. Li, Q. Zhang, L. Zhang, Simultaneous determination of capsaicin and dihydrocapsaicin for vegetable oil adulteration by immunoaffinity chromatography cleanup coupled with LC-MS/MS, J. Chromatogr. B. 1021 (2016) 137–144. https://doi.org/10.1016/j.jchromb.2015.12.017.

[47] C. Zhou, D. Ma, W. Cao, H. Shi, Y. Jiang, Fast simultaneous determination of capsaicin, dihydrocapsaicin and nonivamide for detecting adulteration in edible and crude vegetable oils by UPLC-MS/MS, Food Addit. Contam. - Part A Chem. Anal. Control. Expo. Risk Assess. 35 (2018) 1447–1452. https://doi.org/10.1080/19440049.2018.1457801.

[48] M. Vázquez-Espinosa, A. V. González-de-Peredo, E. Espada-Bellido, M. Ferreiro-González, G.F. Barbero, M. Palma, Simultaneous determination by UHPLC-PDA of major capsaicinoids and capsinoids contents in peppers, Food Chem. 356 (2021) 129688. https://doi.org/10.1016/j.foodchem.2021.129688.

[49] K. Hurkova, J. Rubert, M. Stranska-Zachariasova, J. Hajslova, Strategies to document adulteration of food supplement based on sea buckthorn oil: a case study, Food Anal. Methods. 10 (2017) 1317–1327. https://doi.org/10.1007/s12161-016-0674-4.

[50] A. Cichelli, G. Pietro Pertesana, High-performance liquid chromatographic analysis of chlorophylls, pheophytins and carotenoids in virgin olive oils: Chemometric approach to variety classification, J. Chromatogr. A. 1046 (2004) 141–146. https://doi.org/10.1016/j.chroma.2004.06.093.

[51] A.A. D'Archivio, A. Giannitto, M.A. Maggi, F. Ruggieri, Geographical classification of Italian saffron (Crocus sativus L.) based on chemical constituents determined by high-performance liquid-chromatography and by using linear discriminant analysis, Food Chem. 212 (2016) 110–116. https://doi.org/10.1016/j.foodchem.2016.05.149.

[52] A. Arrizabalaga-Larrañaga, P. Rodríguez, M. Medina, F.J. Santos, E. Moyano,

Pigment profiles of Spanish extra virgin olive oils by ultra-high-performance liquid chromatography coupled to high-resolution mass spectrometry, Food Addit. Contam. - Part A Chem. Anal. Control. Expo. Risk Assess. 37 (2020) 1075–1086. https://doi.org/10.1080/19440049.2020.1753891.

[53] S. Van Ruth, M. Alewijn, K. Rogers, E. Newton-Smith, N. Tena, M. Bollen, A. Koot, Authentication of organic and conventional eggs by carotenoid profiling, Food Chem. 126 (2011) 1299–1305. https://doi.org/10.1016/j.foodchem.2010.11.081.

[54] O. Núñez, The role of capillary electrophoresis in food integrity and authenticity, in: O. Núñez, G. Campmajó (Eds.), Chromatogr. Relat. Sep. Tech. Food Integr. Authenticity. Vol. A Adv. Chromatogr. Tech., World Scientific Publishing Europe Ltd., 2021: pp. 71–104. https://doi.org/https://doi.org/10.1142/9781786349958_0004.

[55] M. Creydt, M. Fischer, Food metabolomics: Latest hardware—developments for nontargeted food authenticity and food safety testing, Electrophoresis. 43 (2022) 2334–2350. https://doi.org/10.1002/elps.202200126.

[56] S. Mudrić, U.M. Gašić, A.M. Dramićanin, I. Ćirić, D.M. Milojković-Opsenica, J.B. Popović-Đorđević, N.M. Momirović, Ž.L. Tešić, The polyphenolics and carbohydrates as indicators of botanical and geographical origin of Serbian autochthonous clones of red spice paprika, Food Chem. 217 (2017) 705–715. https://doi.org/10.1016/j.foodchem.2016.09.038.

[57] T. Kopta, A. Sekara, R. Pokluda, V. Ferby, G. Caruso, Screening of chilli pepper genotypes as a source of capsaicinoids and antioxidants under conditions of simulated drought stress, Plants. 9 (2020) 364. https://doi.org/10.3390/plants9030364.

[58] N. Mialon, B. Roig, E. Capodanno, A. Cadiere, Untargeted metabolomic approaches in food authenticity: A review that showcases biomarkers, Food Chem. 398 (2023) 133856. https://doi.org/10.1016/j.foodchem.2022.133856.

CHAPTER 3

METABOLOMIC FINGERPRINTING APPROACHES

3.1. INTRODUCTION

Due to its non-targeted nature, the metabolomic fingerprinting approach involves the comprehensive detection of instrumental responses related to the sample composition in a non-selective way, aiming to detect as many metabolites as possible. In this context, as in the metabolomic profiling approach (introduced in Chapter 2), GC and LC play a crucial role and have been widely applied in the food authentication field. As described by Cuadros-Rodríguez *et al*. [1], different types of analytical data can be obtained by chromatographic techniques. Thus, first-order signals (two dimensions) consist of a data vector formed by signal intensities, each characterised by a particular retention time value. For instance, GC-FID or LC-UV (at a fixed absorption wavelength) provide first-order signals. Instead, in the case of second-order signals (three dimensions), the signal intensities are defined by two variables: the retention time and either the absorption wavelength (in LC-UV using a DAD) or the mass-to-charge ratio (in LC−MS). Furthermore, in recent years, direct MS-based analysis techniques —*e.g.*, direct infusion mass spectrometry (DIMS), flow injection analysis coupled to mass spectrometry (FIA–MS), or ambient ionisation mass spectrometry (AIMS)— have gained significant relevance in this field and under this approach because of their short analysis time and high throughput [2]. These techniques generally produce first-order signals, although second-order signals can also be obtained when ion mobility spectrometry (IMS) is performed before MS.

Thus, on the one hand, several methods based on GC have been developed to assess diverse food fraud cases. For instance, GC-FID first-order signals have been subjected to chemometrics, following a fingerprinting approach, to address food issues such as the effect of processing and ageing in *Brandy de Jerez* PGI [3], the geographical origin of Arbequina EVOO [4], or the adulteration of saffron with turmeric or marigold [5]. Moreover, as mentioned in Section 2.1, GC−MS presents excellent properties for the non-targeted analysis of VOCs, known as '*volatilomics*'. Thus, the main advantages of this technique are its high reproducibility of retention time and mass spectra, reached by calculating the compounds' retention time indexes

(usually through a reference mix of *n*-alkenes) and by the well-established application of EI at 70 eV, respectively [6]. Thus, the possibility to compare reproducible mass spectra containing molecular fragmentation data with in-house or commercial libraries turns this technique into a unique tool for metabolite annotation. Moreover, low-resolution mass analysers are mostly used in metabolomics analysis. Therefore, for instance, a single quadrupole was used in the GC−MS analysis for the cultivar and geographical origin classification of Greek monovarietal EVOO, leading to the identification of 72 VOCs [7]. Similarly, GC−MS, with a QqQ and an IT as the mass spectrometers, was proposed to authenticate citrus monofloral honey [8] and the variety classification of Italian carrots [9], respectively. Nevertheless, some recent studies have also been developed using high-resolution mass spectrometers. For example, Yang *et al*. [10] proposed a non-targeted GC−HRMS method, using a Q-Orbitrap as the mass analyser, for discriminating five meat species.

On the other hand, LC has also demonstrated excellent capabilities for non-targeted approaches. Hence, this introduction includes the following book chapter (Publication III) that examines the role of thin-layer chromatography (TLC) and LC fingerprinting methodologies in guaranteeing food authenticity. Particularly, the chapter distinguishes between LC with conventional detection systems, such as UV or FLD, and LC−HRMS. Therefore, it describes the most common practices and trends —in terms of sample treatment, chromatographic separation, detection system and conditions, and data handling— by discussing some of the most recent applications.

Publication III: Book chapter

Chromatographic fingerprinting approaches in food authentication.

Campmajó, G.; Núñez, O.

In: Núñez, O., Campmajó, G. (Eds.), *Chromatographic and related separation techniques in food integrity and authenticity. Volume A: Advances in chromatographic techniques.* World Scientific Publishing Europe Ltd. 2021, 137 – 166.

3.1.1. PUBLICATION III

Chromatographic fingerprinting approaches in food authentication.

Campmajó, G.; Núñez, O.

In: Núñez, O., Campmajó, G. (Eds.), *Chromatographic and related separation techniques in food integrity and authenticity. Volume A: Advances in chromatographic techniques.* World Scientific Publishing Europe Ltd. 2021, 137 – 166.

https://doi.org/10.1142/9781786349958_0006

Chapter 6

Chromatographic Fingerprinting Approaches in Food Authentication

Guillem Campmajó[*,†,§] and Oscar Núñez[*,†,‡]

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

[†]*Research Institute in Food Nutrition and Food Safety (INSA), University of Barcelona, Santa Coloma de Gramenet, Barcelona, Spain*

[‡]*Serra Húnter Fellow, Generalitat de Catalunya, Barcelona, Spain*

[§]*gcampmajo@ub.edu*

Abstract

Food authentication has become an issue of great interest for consumers, the food industry, regulatory authorities, as well as the scientific community, because of the substantial grown of food fraud in the past years. Historically, the detection of fraudulent practices has been based on the use of targeted methods, focusing on the quantification of primary markers, directly linked to the authenticity problematic. Nevertheless, nowadays this strategy is in some cases insufficient because of the huge number of adulterant substances that can potentially be added in food as well as the increase of complex food authentication issues (e.g., deception on the geographical origin or production method). As an alternative, non-targeted methods, consisting of the analysis of instrumental responses without assuming any previous knowledge about the sample composition, have gained relevance when combined with chemometrics. The present chapter examines the role of thin-layer chromatography (TLC) and liquid chromatography (LC) fingerprinting methodologies, usually combined with chemometric techniques, to guarantee

food authenticity. In the case of LC, fingerprints obtained either by conventional detection systems or high-resolution mass spectrometry (HRMS) are discussed.

6.1. Introduction

Food authentication has emerged in the past decades as being of great interest for consumers, the food industry, regulatory authorities, and the scientific community, mainly due to the increase of food fraud practices. Food fraud, which can be defined as an intentional procedure that violates food laws and generally deceives the consumers, for an economic purpose, costs the global food industry approximately 30 billion euros a year. Usually, illegal practices such as substitution, adulteration, dilution, deliberate mislabeling or counterfeiting, among many others, involve the modification of the organoleptic and sensorial properties of goods to get consumer attention, reduce the total production costs by using cheaper products than the original one, or ensure a longer expiration time. Besides, the complexity of the food chain has increased the opportunities to conduct fraud throughout it. For that reason, several national and international governmental organizations — i.e., European Food Safety Authority (EFSA), in the European Union (EU), the China Food and Drug Administration (CFDA), in China, or the Food and Drug Administration (FDA), in the United States of America (USA) — have established strict directives and regulations to protect the essence of original products, which is strongly related to their quality, geographical origin, and production method as well as processing technologies [1]. Moreover, the manipulation of foodstuff can sometimes imply not only an economic deception but also a threat to consumer's health, especially when dealing with an allergen or toxic substance.

Till date, most of the strategies proposed for the detection of fraudulent practices are targeted methods, focusing on the determination and quantification of a specific analyte or group of analytes. Their concentrations or peak signals are then used as food features to address food authenticity. Depending on the relation between these compounds, which are so-called analytical markers, and the food authentication subject, they are defined as primary or secondary markers. In this line, the targeted analysis of primary markers directly links to solving the authenticity problem, since these markers are usually the known added illegal substances, complying with legal requirements regarding limits. For instance, Arrizabalaga-Larrañaga *et al*. [2] developed an ultra-high-performance liquid chromatography coupled to tandem mass spectrometry (UHPLC-MS/MS) method for the determination of the adulteration of olive oils with natural pigments (e.g., carotenoids and chlorophylls) and the food additive E-141i (green copper chlorophyll complex), which are illegally added to enhance low-quality olive oil color. In contrast, secondary marker profiling (multiple targeted analysis) indirectly allows the

certification of foodstuff, as a variation in their content is a consequence of the food authentication issue. As an example, Barbosa *et al*. [3] authenticated the geographical origin of paprika samples, profiling a total of 36 phenolic compounds by UHPLC-MS/MS.

Nevertheless, the vast amount and diversity of food and beverage products supplied by food manufacturers, the wide range of particularities that each foodstuff can present, the great variety of frauds that can be committed, and the existence of numerous potential adulterant substances (many of which are unknown) have made the detection of these illegal practices by traditional targeted methods more difficult. Therefore, in the past years, non-targeted methods, consisting of the analysis of instrumental responses without assuming any previous knowledge about the sample composition, have emerged as a reliable alternative [4]. It should be pointed out that the nomenclature used in this chapter for both targeted and non-targeted methodologies bases on Ballin *et al*. [5] proposition.

Cifuentes first defined *foodomics* as a *discipline that studies the food and nutrition domains through the application of omics technologies* [6]. Among others, omics technologies encompass *genomics*, *proteomics*, *isotopolomics*, and *metabolomics*, which focus on the analysis of the entirety of the DNA, proteins, element or isotopic profiles, and metabolites, respectively [7]. Hence, *metabolomics* studies a massive number of small molecules (<1500 Da), corresponding to intermediate or end products of metabolic pathways, and with either endogenous or exogenous origins. Therefore, since it is the closest *omics* discipline to the phenotype of biological systems, *metabolomics* fingerprinting approaches have been widely proposed to solve food authentication issues through chromatographic techniques. Nevertheless, because of the complexity and dynamic range of all these chemical compounds (e.g., different polarities and other chemical properties), a single analytical technique cannot be used to perform the complete metabolome analysis [8], and different analytical platforms are commonly proposed. In this context, the analytical technique, as well as the chosen working conditions, directly influence on the fingerprinting acquisition. Thus, in this chapter, the application of thin-layer chromatography (TLC) and liquid chromatography (LC) *metabolomics* fingerprinting is discussed.

As in any other analytical chemistry approach, data accuracy, reproducibility, and reliability, highly depend on correct sample preparation. While in a targeted analysis it centers on the optimization of some concrete compound signals with tedious, time-consuming, and expensive procedures, to achieve high sensitivity and hence low limits of detection (LODs); in fingerprinting approaches, it usually focuses on detecting as many components of the food matrix as possible by using unspecific sample treatments to avoid losing information [9]. Therefore, they commonly use fast and straightforward sample treatments. However, due to the wide diversity of existing compounds in food samples, sometimes specific extraction fractions (e.g., polar and non-polar fractions) are studied.

Moreover, non-targeted methodologies provide a multitude of data, making its interpretation a crucial and challenging step. In fact, in order to use these sample data sets to characterize, classify, and authenticate food products, as well as to extract chemical information from them, chemometrics is necessary. In this line, while non-supervised exploratory chemometric techniques — principal component analysis (PCA), cluster analysis (CA), and hierarchical cluster analysis (HCA) — provide trends and clusters between samples according to their similarities and dissimilarities, supervised classificatory techniques allow sample discrimination according to well-known classes. Even though supervised linear techniques — linear discriminant analysis (LDA), partial least squares regression-discriminant analysis (PLS-DA), orthogonal projections to latent structures-discriminant analysis (OPLS-DA), and soft independent modelling of class analogy (SIMCA) — are mainly proposed with classificatory purposes, in some cases supervised nonlinear techniques — support vector machine (SVM), artificial neuronal networks (ANN), k-nearest neighbor (kNN) — are also chosen. In some other applications, partial least squares (PLS) regression is employed to correlate food chromatographic fingerprints with a specific parameter, such as percentage of an adulterant.

In the next sections, several examples dealing with *metabolomics* fingerprinting chromatographic methodologies, in combination with chemometric techniques, to guarantee food authentication, are addressed.

6.2. High-Performance Thin-Layer Chromatography

High-performance thin-layer chromatography (HPTLC), also known as planar chromatography, has been an extensively used technique in food and natural product analysis for qualitative and quantitative results. For instance, despite the dynamic development of other chromatographic techniques, HPTLC is still employed in the phytochemical analysis [10]. Besides, it fits ideally with fingerprinting approaches since it provides short chromatograms, and excellent reproducibility, with minimal sample preparation, high-throughput screening, and an inexpensive and environmentally friendly analysis. In this line, HPTLC fingerprinting approaches have been applied not only in the natural product field but also in the food authentication one. As an example, Table 6.1 summarizes some of the applications of HPTLC in this field developed in the last five years.

As can be seen in Table 6.1, HPTLC has been mainly used for the analysis of liquid and viscous samples, such as honey, wine, oil, tea, and fruit extracts. Although in some cases no sample treatment or just one stage of sample dissolution was required, some authors performed liquid-liquid extraction (LLE), solid-liquid extraction (SLE), or even solid phase extraction (SPE) methods. The lack

Table 6.1. Compilation of some HPTLC fingerprinting methodologies developed in the last five years to address food authentication.

Sample	Extraction	HPTLC[a]	Data analysis	Ref.
Adulteration				
Chamomile tea	SLE with 96% ethanol	Silica gel 60 NH_2 plate F_{254}s (20 × 10 cm) Ethyl acetate:formic acid:acetic acid:water (30:1.5:1.5:3, *v/v/v/v*) NP/PEG 400 UV light (366 nm)	VI, densitometric data analysis, PCA, and HCA	[11]
Wine	—	Silica gel 60 plate F_{254} (20 × 10 cm) Dichloromethane: methanol:formic acid (73:20:7, *v/v/v*) and 0.14 M sodium dodecyl sulfate in *n*-butanol:water:heptane (13:4:83, *v/v/v*) 1% 2-aminoethyl diphenylborinate (*w/v*) UV light (366 nm)	Densitometric data analysis, PCA, HCA, and ANN	[12]
Cultivar/botanical origin				
Sandalwood oil	Dissolution in hexane	Silica gel 60 plate F_{254} (20 × 10 cm) Toluene:ethyl acetate (85:15, *v/v*) *p*-anisaldehyde sulfuric acid White and UV (254 and 366 nm) light	VI and densitometric data analysis	[13]
Honey	SPE with a glass column packed with amberlite XAD-2	Silica gel 60 plate F_{254} (20 × 10 cm) Chloroform:ethyl acetate:formic acid (50:40:10, *v/v/v*) 1% methanolic $AlCl_3$, Ce-P-Mo, and 1% methanolic 2-aminoethyl diphenylborate White and UV (254 and 366 nm) light	VI and densitometric data analysis	[14, 15]
Honey	UAE with dichloromethane SPE with styrene-divinylbenzene cartridges	Silica gel 60 plate F_{254} (20 × 10 cm) Toluene:ethyl acetate (80:20, *v/v*) Anisaldehyde White and UV (254 and 366 nm) light	VI, PCA, and HCA	[16, 17]

(*Continued*)

Table 6.1. (*Continued*)

Sample	Extraction	HPTLC[a]	Data analysis	Ref.
Honey	LLE with dichloromethane	Silica gel 60 plate F_{254} (20 × 10 cm) Toluene:ethyl acetate:formic acid (6:5:1, *v/v/v*) Vanillin White and UV (254 and 366 nm) light	VI and densitometric data analysis	[18, 19]
Wild fruits	SLE with 70% ethanol	Silica gel 60 plate F_{254} (20 × 10 cm) Ethyl acetate:toluene:formic acid:water (30:1.5:4:3, *v/v/v/v*) 0.5% ethanolic 2-aminoethyl diphenylborate UV light (365 nm)	PCA and CA	[20]
Cultivar/botanical origin, producer, and crop year				
White wine	—	Silica gel 60 plate F_{254} (20 × 10 cm) Ethyl acetate:formic acid:acetic acid:water (20:2:2:4, *v/v/v/v*) Silica gel 60 RP-18 plate F_{254}s (20 × 10 cm) Methanol:water:formic acid (5:5:0.1, *v/v/v*) PEG UV light (254 and 366 nm)	VI and CA	[21]

Notes: [a] From top to bottom: stationary phase, mobile phase, derivatization reagent, and detection system.
Natural products (NP), polyethylene glycol (PEG), ultrasound-assisted extraction (UAE), ultraviolet (UV), visual inspection (VI).

of a tedious pre-treatment step, as well as the possibility to simultaneously analyze several amounts of samples on one plate, is one of the advantages of this chromatographic technique when compared to other more complex ones.

Moreover, the selection of the stationary and mobile phases — which directly affect the separation of the sample compounds, the use of a derivatization reagent, and the detection system — play a crucial role in the visualization of rich and selective HPTLC fingerprints. Although, commonly, HPTLC fingerprinting employs a normal phase system as the stationary phase, using a polar adsorbent such as silica gel, and a mix of aprotic organic solvents (dichloromethane, chloroform, ethyl acetate, and toluene) as the mobile phase, other chromatographic systems have also been employed. For instance, Guzelmeric *et al*. [11], who studied the adulteration of chamomile tea with other species having similar flowers, used NH_2 modified silica gel plates, and ethyl acetate:formic acid:acetic acid:water (30:1.5:1.5:3, *v/v/v/v*) solution as the mobile phase, to obtain characteristic chemical fingerprints. Instead, Hosu *et al*. [21] developed two different HPTLC systems for the authentication of white wines according to their variety, vineyard, and crop year. Thus, the authors proposed both silica gel and C_{18}-modified silica gel stationary phases, with ethyl acetate:formic acid:acetic acid:water (20:2:2:4, *v/v/v/v*) and methanol:water:formic acid (5:5:0.1, *v/v/v*) as the mobile phases, respectively, for the separation of wine white compounds. Besides, the images of the plates were made under UV light at 254 and 366 nm. Several variations in terms of abundance and number of major compounds were observed by visual inspection (VI) between samples (except that the images of silica gel plates under 254 nm that presented similar fingerprints for all the studied samples). However, the individual evaluation of each chromatographic system (silica gel plates at 366 nm, and reversed-phase plates at 254 nm and 366 nm) by CA could not achieve the correct classification of white wines. Alternatively, when combining these fingerprint data, samples were successfully clustered in three main groups — S*auvignon Blanc*, *Feteasca Alba*, and *Riesling wines* — suggesting that wine variety has more influence on the HPTLC fingerprints than the vineyard and the crop year.

Regarding the detection of the HPTLC fingerprints, it is usually conducted using UV light, particularly at 254 and 366 nm, and in some cases white light. Moreover, as it can be observed in Table 6.1, the use of silica gel plates coated with green or blue inorganic fluorescent indicators (F_{254} and F_{254}s, respectively) allows the detection of colorless substances by fluorescence quenching. Besides, several derivatization reagents are usually sprayed to the plate in order to enhance the fluorescent signal.

The application of HPTLC fingerprinting to address the classification and authentication of foodstuff can aid in three strategies: VI, processing of densitometric data, and chemometrics [22]. VI is based on the subjective perception of the analyst and focuses on the visual detection of chemical patterns between samples (e.g., the detection of unique bands related to specific sample groups or

the presence of major and minor bands based on their intensity). Instead, quantitative comparison through the evaluation of densitometric measurements, consisting of the optical density obtained along the chromatogram and generally acquired with a device such as charge-coupled device (CCD), provides more reliable information as well as some advantages (i.e., fast and straightforward image processing), without implying more difficulty. Finally, multivariate analysis allows using all the information present at the acquired data for classification purposes. Generally, it requires pre-treatment steps involving baseline drift, peak alignment, normalization, and autoscaling.

As an example, Agatonovic-Kustrin *et al*. [12], who aimed to determine the major variety grape used in a given wine, processed the obtained data by studying the corresponding densitograms as well as employing chemometric techniques. The evaluation of the obtained densitograms provided characteristic peaks for some of the varieties studied (e.g., four and five peaks were associated with *Shiraz* and *Merlot* wines, respectively). Instead, PCA, HCA, and ANN were carried out using the whole HPTLC fingerprint. In this last case, the method provided good discrimination of samples according to grape varieties, with a classification rate of 96%, 100%, and 66%, for the training, testing, and external validation set.

6.3. Liquid Chromatography (LC)

The complexity and diversity of food matrices, as well as the significant variability of compounds present in them — differing in polarity, structure, and concentration ranges (from $g \cdot kg^{-1}$ to trace level) — makes their analysis difficult. In this context, the versatility of LC, allowing the selection of different separation and detection choices (either for polar or non-polar compounds), can provide an enormous amount of information, especially when employing a fingerprinting approach. Besides, the recent use of sub-2-μm fully porous or sub-3-μm core-shell particles, shifting from high-performance liquid chromatography (HPLC) to ultra-high-performance liquid chromatography (UHPLC) has opened new possibilities to achieve high throughput chromatographic analytical separations, increasing chromatographic resolution and peak capacity.

The LC *metabolomics* fingerprinting approach can be classified according to the detection system employed. On the one hand, LC in combination with conventional detection systems — such as UV, fluorescent detection (FLD), electrochemical detector (ECD), or charged aerosol detector (CAD) — provides 2D fingerprints (retention time and peak intensity) based on the simultaneous detection of instrumental data points. The potentiality of these fingerprints, when subjected to multivariant statistical analysis in the food authentication field, is shown in Section 6.3.1. On the other hand, as can be seen in Section 6.3.2, LC coupled to high-resolution mass spectrometry (HRMS) adds a third dimension (peak *m/z*)

and unspecified targets are simultaneously detected, obtaining characteristic fingerprints, which can be used not only for sample classificatory purposes through chemometrics but also as a further identification step of the most discriminant biomarkers.

6.3.1. *Conventional detection*

In this section, the applicability of LC combined with different conventional detection system fingerprinting techniques to characterize and classify food through chemometrics is evaluated. In this line, Table 6.2 summarizes some of the latest applications of this approach for different food authentication items, such as adulterations, identification of the cultivar or geographical origin, and confirmation of the processing method used.

Thus, Table 6.2 shows a wide variety of goods that has been analyzed by this kind of fingerprints. Most of the examined food matrices derive from the plant kingdom. For instance, Cetó *et al*. [23] used HPLC-UV fingerprints to quantify by PLS both type and geographical origin adulterations of paprika, which is a red seasoning powder obtained from different varieties of red pepper (*Capsicum annuum* L.). Instead, some authors have also analyzed foodstuff of animal origin, such as Campmajó *et al*. [29], who studied the processing method for mislabeling of hen eggs by HPLC-UV fingerprinting.

Similar to other fingerprinting methodologies, LLE and SLE are commonly used as the sample treatment choice, because of their simplicity, ease of use, and capacity to extract polar or non-polar chemical compounds directly from the food matrix. However, specific applications required a preceding step. For example, prior to an LLE with hexane, Jiménez-Carvelo *et al*. [26], who aimed to discriminate olive oil from other types of edible oils, carried out a methyl-transesterification reaction to characterize the HPLC-CAD compositional fingerprint of an organic phase extract. This process allowed the liberation of sterols and, thus, their subsequent detection. Therefore, the organic extract contained compounds such as fatty acid methyl esters (FAME), sterols, and mono- and diglycerides, among others.

Regarding the chromatographic separation, several stationary and mobile phases have been used depending on the polarity of the analyzed sample extract, although most of the works used a C_{18} column. Moreover, some authors have evaluated the authentication capability of fingerprints obtained with different chromatographic systems. For instance, Bikrani *et al*. [28] compared the results obtained with normal- and reversed-phase HPLC-UV fingerprinting — cyano and C_{30} columns were used, respectively — for the geographical origin certification of fat-spread products. Figure 6.1 shows the HPLC-UV fingerprints obtained for a Moroccan and a Spanish margarine/spread samples using normal- and

Table 6.2. Compilation of some LC fingerprinting methodologies, using spectroscopic, electrochemical or charged aerosol detection systems, developed in the last five years to address food authentication.

Sample	Extraction	LC and detection system[a]	Data analysis	Ref.
Adulteration				
Paprika	SLE with water:acetonitrile (20:80, *v/v*)	HPLC-UV (280 nm) Kinetex C_{18} column (100 × 4.6 mm, 2.6 μm) Gradient elution (1 mL·min^{-1}): (A) water with 0.1% formic acid (*v/v*) and (B) methanol	LDA and PLS	[23]
Cultivar/botanical origin				
White wine	—	HPLC-ECD (1000 mV) SeQuant ZIC *p*-HILIC column (150 × 4.6 mm, 5 μm) Isocratic elution (1 mL·min^{-1}): 0.05% (*v/v*) TFA:acetonitrile (19:81, *v/v*)	PLS-DA	[24]
Avocado	SLE with hexane:isopropanol (3:2, *v/v*)	HPLC-CAD Lichrospher 100 CN column (250 × 4 mm, 5 μm) Isocratic elution (1.2 mL·min^{-1}): *n*-hexane:isopropanol (96:4, *v/v*)	SIMCA and PLS-DA	[25]
Olive oil	Transesterification with sodium methoxide and LLE with hexane	HPLC-CAD Lichrospher 100 CN column (250 × 4 mm, 5 μm) Isocratic elution (1.2 mL·min^{-1}): *n*-hexane:isopropanol (96:4, *v/v*)	SIMCA, PLS-DA, kNN, and SVM-C	[26]
Olive oil	LLE with methanol:water (60:40, *v/v*) and hexane to defat	UHPLC-UV (280 nm) and UHPLC-FLD (λ_{exc}: 280 nm/λ_{em}: 339nm) Zorbax C_{18} column (150 × 4.6 mm, 1.8 μm) Gradient elution (0.8 mL·min^{-1}): (A) water with 0.5% acetic acid (*v/v*) and (B) acetonitrile	PCA, SIMCA, PLS-DA, and kNN	[27]

Geographical origin				
Fat-spread product	Dilution with *n*-hexane	HPLC-UV (210 and 254 nm) DevelosilTM C30-UG-5 column (250 × 4.6 mm, 5 μm) Isocratic elution (1.2 mL·min^{-1}): acetonitrile:isopropanol (40:60, *v/v*)	PCA, SIMCA, and PLS-DA	[28]
Processing				
Hen egg	SLE with water:acetonitrile (20:80, *v/v*)	HPLC-UV (250 nm) Kinetex C_{18} column (100 × 4.6 mm, 2.6 μm) Gradient elution (0.4 mL·min^{-1}): (A) water with 0.1% formic acid (*v/v*) and (B) methanol	PCA and PLS-DA	[29]
Adulteration and cultivar/botanical origin				
Nuts	SLE with acetone:water (70:30, *v/v*) and hexane to defat	HPLC-FLD (λ_{exc}: 280 nm/λ_{em}: 350 nm) Kinetex C_{18} column (100 × 4.6 mm, 2.6 μm) Gradient elution (0.4 mL·min^{-1}): (A) water with 0.1% formic acid (*v/v*) and (B) methanol	PCA, PLS-DA, and PLS	[30]
Fruit-based product	LLE with acetone:water:hydrochloric acid (70:29.9:0.1, *v/v/v*)	HPLC-UV (280 nm) Kinetex C_{18} column (100 × 4.6 mm, 2.6 μm) Gradient elution (1 mL·min^{-1}): (A) water with 0.1% formic acid (*v/v*) and (B) methanol	PCA and PLS	[31]
Cranberry-based product	LLE with acetone:water:hydrochloric acid (70:29.9:0.1, *v/v/v*)	HPLC-UV (370 nm) Kinetex C_{18} column (100 × 4.6 mm, 2.6 μm) Gradient elution (1 mL·min^{-1}): (A) water with 0.1% formic acid (*v/v*) and (B) methanol	PCA and PLS	[32]

Notes: [a]From top to bottom: chromatographic technique and detection system, column, and elution mode and mobile phase.
Hydrophilic interaction liquid chromatography (HILIC), support vector machine-classification (SVM-C).

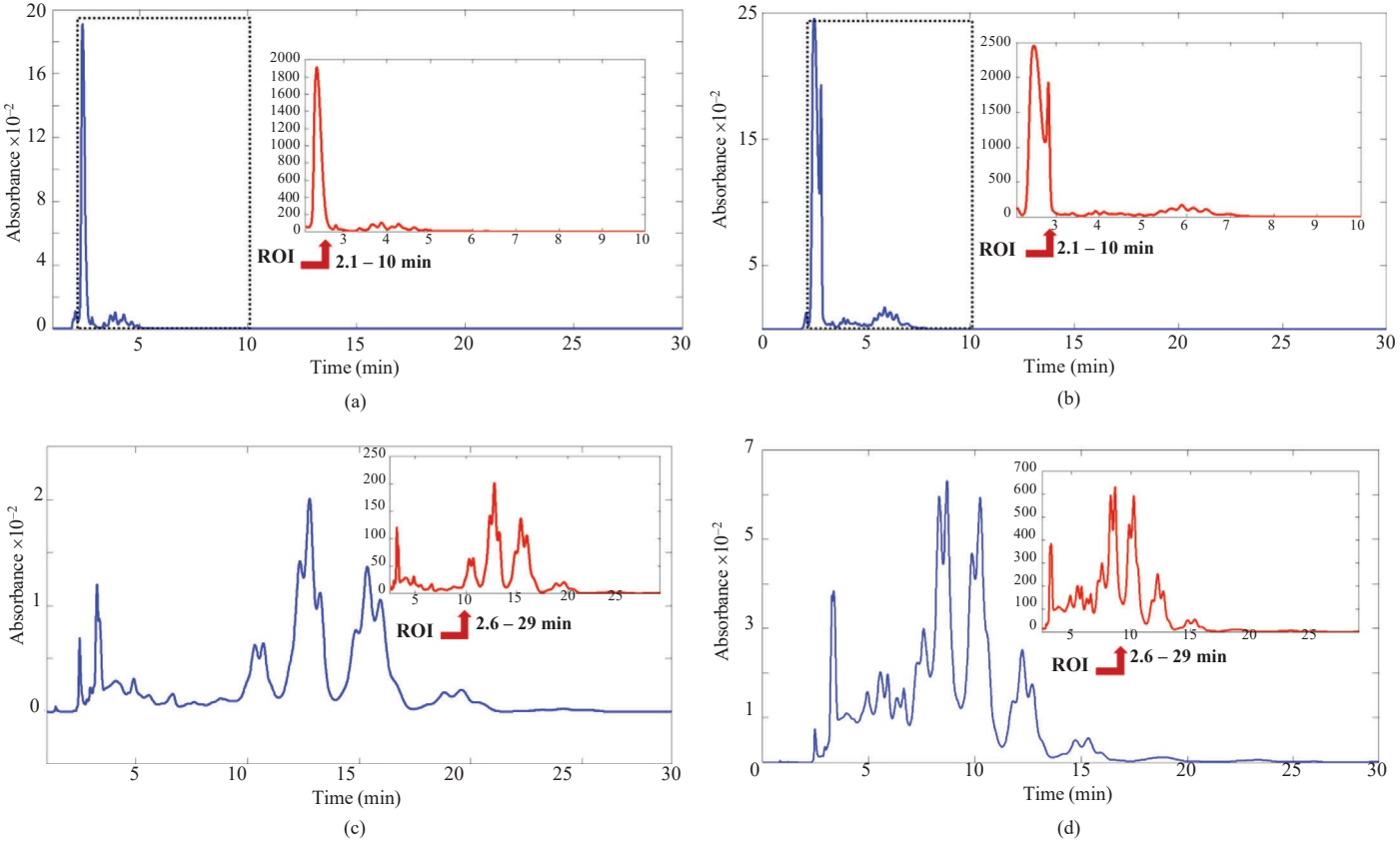


Fig. 6.1. HPLC-UV fingerprints obtained for a Moroccan and a Spanish margarine/spread samples using normal- (a, b) and reversed-phase (c, d) HPLC-UV, respectively.

Source: Reproduced from Open Access Ref. [28].

reversed-phase HPLC-UV. In this case, better PLS-DA classification was obtained with reversed-phase HPLC-UV fingerprints, achieving a rate of 100%.

Moreover, many and various conventional detectors are combined with HPLC, with UV being the most typically employed, in order to obtain distinctive fingerprints. Table 6.2 presents some other applications where FLD, ECD, or CAD have also been used. For instance, Campmaj$\acute{o}$ *et al*. [30] used HPLC-FLD fingerprints for the classification of nuts and the detection of almond-based products adulteration, as an alternative to the corresponding HPLC-UV ones. In comparison, HPLC-FLD fingerprints contained a fewer number of peaks, since the selection of excitation and emission acquisition wavelengths provides more selective data than HPLC-UV. However, in view of the better predictive figures obtained with HPLC-FLD, it seemed that this more selective detection was related to the acquisition of relevant chemical descriptors, such as flavanols (e.g., catechin, epicatechin, and related species). Alternatively, although they are not included in Table 6.2, refractive index (RI) and evaporative light scattering detector (ELSD) can also be employed for the detection of chromatographic fingerprints.

The use of the chromatographic fingerprints and the multivariate analysis strategy depend on the nature of the food authentication issue to be solved. In this line, Pardo-Mates *et al*. [31], who studied the authentication of fruit-based extracts — such as cranberry, raspberry, blueberry, and grape — by HPLC-UV fingerprints, proposed PCA for sample discrimination and PLS for the quantitation of adulterant levels in cranberry fruit extracts. Thus, subsequent to an LLE with acetone:water:hydrochloric acid (70:29.9:0.1, *v/v/v*) mix, HPLC-UV fingerprints were recorded by using a C_{18} column and a gradient elution with water, acidified with formic acid, and methanol. Then, the combination of different time segments of the chromatographic fingerprints (4.7–6.5, 8–14, 15–17, and 29–30 min) provided the best sample discrimination by PCA, although cranberry- and grape-based extracts appeared quite close; 3 to 23 min HPLC-UV segments were subjected to PLS to successfully quantitate grape, blueberry, or raspberry percentage of adulteration in the cranberry extracts, with prediction errors below 4.3%.

Otherwise, Bajoub *et al*. [27] evaluated the varietal origin of extra-virgin olive oil (EVOO) by using different data fusion approaches consisting of the combination of the obtained HPLC-UV and HPLC-FLD fingerprints. In order to obtain the sample extracts, the authors proposed an LLE with methanol:water (60:40, *v/v*), also adding hexane for defatting. Both HPLC-UV (with an excitation wavelength of 280 nm) and HPLC-FLD (with excitation and emission wavelengths of 280 and 339 nm, respectively) were acquired using a C_{18} column. Regarding the data treatment, besides the individual employment of HPLC-UV and HPLC-FLD fingerprints as chemical descriptors, Bajoub *et al*. constructed "low-level" — HPLC-UV and HPLC-FLD data were concatenated — and "mid-level" — optimal segments of both HPLC-UV and HPLC-FLD were concatenated — data fusion

data matrices. In this line, "mid-level" data fusion matrix provided the best sample prediction ability when using PLS-DA as well as SIMCA.

6.3.2. *High-resolution mass spectrometry (HRMS) detection*

In the last few years, there has been a trend towards the application of non-targeted UHPLC-HRMS fingerprinting methodologies — containing not only chromatographic time and peak signals but also *m/z* values — in the food authentication field. In this context, time-of-flight (TOF) and Orbitrap instruments are the most employed mass analyzers for this purpose because of their high resolving power, which is the ability to distinguish two adjacent ions of equal intensity. While TOF instruments present a resolving power in the range of 12,000–50,000 full-width at half maximum (FWHM), Orbitrap mass analyzers reach values up to 500,000 FWHM. Thus, both TOF and Orbitrap instruments provide accurate mass measurements with mass error values below 5 and 2 ppm, respectively [33]. Molecular *formulae* of specific ions can be determined, taking advantage of this high accuracy, as well as isotopic abundance ratios; while structural information can also be obtained dissociating the compound by fragmentation. Hybrid configurations such as quadrupole-Orbitrap (Q-Orbitrap) and quadrupole-TOF (Q-TOF) allow the fragmentation of specific ions by tandem mass spectrometry (MS/MS), bringing more information for the identification of unknown compounds.

Furthermore, given the excellent sensitivity and selectivity, high-throughput nature, and depth coverage of HRMS, UHPLC-HRMS perfectly fits with *metabolomics* fingerprinting approaches. Table 6.3 summarizes some of the latest applications of UHPLC-HRMS in the food authenticity field.

In accordance with the HPTLC and LC fingerprint applications seen in previous sections (Section 6.2 and Section 6.3.1, respectively), most of the analyzed samples arise from the plant kingdom products. However, some authors also propose this methodology for animal-derived matrices such as hen eggs [35, 36]. Moreover, although fingerprinting strategies do not require prior knowledge of the sample chemical composition to detect different compound classes, some authors focus the whole analytical strategy (i.e., sample treatment, chromatographic separation, and mass spectrometric acquisition) to obtain a fingerprint strongly related to a specific family. In this line, a UHPLC-HRMS fingerprinting method favoring the detection of phenolic and polyphenolic compounds, which are a family of secondary aromatic metabolites ubiquitously spread through the plant kingdom, was developed by Barbosa *et al*. for the authentication of paprika [46]. Thus, an SLE with water:acetonitrile (20:80, *v/v*) and a chromatographic separation based on a C_{18} column and water and acetonitrile (both acidified with 0.1% of formic acid), as the mobile phase components, were performed. Besides, heated-electrospray ionization (H-ESI) in negative mode, which is by far the most generalized ionization source employed in the determination of polyphenols [48], was used.

Table 6.3. Compilation of some UHPLC-HRMS fingerprinting methodologies developed in the last five years to address food authentication.

Sample	Extraction	LC and HRMS conditions[a]	Data analysis	Ref.
Adulteration				
Cocoa product	Polar extract: SLE with methanol:water (90:10, *v/v*) with 5 mM ammonium acetate Non-polar extract: SLE with isopropanol:chloroform (80:20, *v/v*) with 20 mM ammonium acetate	Accucore RP-MS column (150 × 2.1 mm, 2.6 μm) Gradient elution (0.35 mL·min^{-1}): (A) water with 10 mM ammonium formate and (B) isopropanol:acetonitrile (60:40, *v/v*) with 10 mM ammonium formate Cogent Diamond Hydride column (150 × 2.1 mm, 2.2 μm) Gradient elution (0.6 mL·min^{-1}): (A) water with 0.1% acetic acid (*v/v*) and (B) acetonitrile with 0.1% acetic acid (*v/v*) H-ESI (±) Q-TOF (full-scan mode 60-1200 *m/z*)	PCA and SPLS	[34]
Aging period				
Hen egg	LLE with dichloromethane:methanol (75:25, *v/v*)	Accucore RP-MS column (100 × 2.1 mm, 2.6 μm) Gradient elution (0.3 mL·min^{-1}): (A) water with 0.1% formic acid (*v/v*) and 5 mM ammonium acetate and (B) methanol with 0.1% formic acid (*v/v*) and 5 mM ammonium acetate H-ESI (+) Q-TOF (full-scan mode 100-1000 *m/z* and MS/MS)	PCA	[35]
Hen egg	LLE with acetonitrile:water (80:20, *v/v*)	Luna Omega C_{18} column (150 × 2.1 mm, 1.6 μm) Gradient elution (0.3 mL·min^{-1}): (A) water with 0.1% formic acid (*v/v*) and 5 mM ammonium formate and (B) methanol with 0.1% formic acid (*v/v*) and 5 mM ammonium formate H-ESI (±) Q-Orbitrap (full-scan mode 75-1000 *m/z* and DDA)	PCA and OPLS-DA	[36]

(*Continued*)

Table 6.3. (*Continued*)

Sample	Extraction	LC and HRMS conditions[a]	Data analysis	Ref.
Cultivar/botanical origin				
Vaccinum berries	SLE with methanol	HSS T3 column (100 × 2.1 mm, 1.8 μm) Gradient elution (0.4 mL·min^{-1}): (A) water with 0.1% formic acid (*v/v*) and 5 mM ammonium formate and (B) methanol:water (95:5, *v/v*) with 0.1% formic acid (*v/v*) and 5 mM ammonium formate H-ESI (±) Q-TOF (full-scan mode 100–1200 *m/z* and IDA)	PLS-DA	[37]
Geographical origin				
Green tea	SLE with dichloromethane:methanol (50:50, *v/v*)	BEH C_{18} column (100 × 2.1 mm, 1.7 μm) Gradient elution (0.4 mL·min^{-1}): (A) water:methanol (95:5, *v/v*) with 0.1% formic acid (*v/v*) and 5 mM ammonium formate and (B) isopropanol:methanol:water (65:30:5, *v/v/v*) with 0.1% formic acid (*v/v*) and 5 mM ammonium formate H-ESI (±) Q-TOF (full-scan mode 50–1200 *m/z*, MS^E, and MS/MS)	PCA, PLS-DA, and OPLS-DA	[38]
Coffee	Polar extract: SLE with acetonitrile:water: formic acid (79:20:1, *v/v/v*) Semi-polar extract: SLE with acetone	Cortecs C_{18} column (100 × 2.1 mm, 2.7 μm) Gradient elution (0.3 mL·min^{-1}): (A) water with 0.01% formic acid (*v/v*) and (B) methanol with 0.01% formic acid (*v/v*) BEH HILIC column (100 × 2.1 mm, 1.7 μm) Gradient elution (0.3 mL·min^{-1}): (A) acetonitrile with 0.01% formic acid (*v/v*) and 10 mM ammonium acetate and (B) water with 0.01% formic acid (*v/v*) and 10 mM ammonium acetate BEH C_{18} column (100 × 2.1 mm, 2.7 μm)	PCA and PLS-DA	[39]

		Gradient elution (0.3 mL·min^{-1}): (A) acetonitrile with 0.1% formic acid (*v/v*) and 0.5 mM ammonium acetate and (B) butanol with 0.1% formic acid (*v/v*) and 0.5 mM ammonium acetate H-ESI (±) Q-TOF (full-scan mode 50–1200 *m/z*)		
Wine	Dilution with water	Hypersil GOLD aQ column (100 × 2.1 mm, 1.9 μm) Gradient elution (0.6 mL·min^{-1}): (A) water with 0.1% formic acid (*v/v*) and (B) acetonitrile BEH C_{18} column (100 × 2.1 mm, 1.7 μm) Gradient elution (0.3 mL·min^{-1}): (A) water with 0.01% formic acid (*v/v*) and (B) methanol with 0.01% formic acid (*v/v*) H-ESI (-) Orbitrap (full-scan mode 50–1000 *m/z* and AIF) Q-TOF (full-scan mode 50–1000 *m/z* and MS^E)	PCA, PLS-DA and OPLS-DA	[40]
Processing				
Tomato	LLE with methanol	Luna C_8 column (100 × 2 mm, 3 μm) Gradient elution (0.35 mL·min^{-1}): (A) water:methanol (98:2, *v/v*) with 0.1% formic acid (*v/v*) and 5 mM ammonium formate and (B) methanol:water (98:2, *v/v*) with 0.1% formic acid (*v/v*) and 5 mM ammonium formate H-ESI (+) Q-Orbitrap (full-scan mode 74–1100 *m/z* and AIF)	PCA	[41]

(*Continued*)

Table 6.3. (*Continued*)

Sample	Extraction	LC and HRMS conditions[a]	Data analysis	Ref.
Carrot	Polar extract: SLE with water:methanol (60:40, *v/v*) Non-polar extract: SLE with chloroform	BEH C_{18} column (100 × 2.1 mm, 1.7 μm) Gradient elution (0.27 mL·min^{-1}): (A) water:methanol (95:5, *v/v*) with 0.2% formic acid (*v/v*) and (B) methanol with 0.2% formic acid (*v/v*) H-ESI (±) TOF (full-scan mode 50–1000 *m/z*)	PCA and OPLS-DA	[42]
Adulteration and cultivar/botanical origin				
Berry fruit juice	Dilution with water	C_{18} column (100 × 2.1 mm, 2.6 μm) Gradient elution (0.3 mL·min^{-1}): (A) water with 0.1% formic acid (*v/v*) and (B) methanol with 0.1% formic acid (*v/v*) H-ESI (±) Q-TOF (full-scan mode 50–1000 m/z and IDA)	PCA-DA and OPLS-DA	[43]
Aging period and processing				
Wine	Dilution with acetonitrile	BEH amide column (150 × 2.1 mm, 1.7 μm) Gradient elution (0.25-0.4 mL·min^{-1}): (A) acetonitrile:water (95:5, *v/v*) with 20 mM ammonium formate and (B) water: acetonitrile (98:2, *v/v*) with 20 mM ammonium formate H-ESI (+) Q-TOF (full-scan mode 30–1000 *m/z*)	PCA and OPLS-DA	[44]

Cultivar/botanical origin and geographical origin				
Honey	LLE with water	Hypersil GOLD C_{18} column (100 × 2.1 mm, 1.9 μm) Gradient elution (0.3 mL·min^{-1}): (A) water with 0.1% formic acid (*v/v*) and (B) acetonitrile with 0.1% formic acid (*v/v*) H-ESI (±) Q-Orbitrap (full-scan mode 80–1200 *m/z* and DDA)	PCA and PLS-DA	[45]
Geographical origin and variety				
Paprika	SLE with water:acetonitrile (20:80, *v/v*)	Ascentis Express C_{18} column (100 × 2.1 mm, 2.7 μm) Gradient elution (0.3 mL·min^{-1}): (A) water with 0.1% formic acid (*v/v*) and (B) acetonitrile with 0.1% formic acid (*v/v*) H-ESI (-) Q-Orbitrap (full-scan mode 100–1500 *m/z*)	PCA and PLS-DA	[46]
Aging period, geographical origin, and processing				
Golden rum	Dilution with water	Hypersil ODS C_{18} column (250 × 4.6 mm, 5 μm) Gradient elution (0.2 mL·min^{-1}): (A) water with 30 mM ammonium acetate (pH 5) and (B) methanol H-ESI (±) Orbitrap (full-scan mode 50–1000 *m/z* and AIF)	PCA, HCA, and PLS-DA	[47]

Notes: [a]From top to bottom: column, elution mode and mobile phase, and mass spectrometry conditions.

All-ion fragmentation (AIF), data-dependent acquisition (DDA), information dependent acquisition (IDA), principal component analysis-discriminant analysis(PCA-DA), sparse partial least squares (SPLS) regression.

As can be seen in Table 6.3, sample dilution avoiding its treatment is extensively used in liquid matrices, such as alcoholic beverages or fruit juices, as it ensures minimal sample handling, and hence, it maintains almost the whole components of the original matrix. Instead, solid or viscous matrices generally require fast and straightforward sample treatment, which usually consists of a SLE or LLE, respectively. Therefore, since the sample preparation strategy for fingerprinting analysis needs to be capable of extracting a broad range of compounds, minimizing potential interferents, and providing a good reproducibility for compounds with different chemical properties (i.e., size, charge, acidity, and polarity), the chosen extracting solvent must be optimized. In this line, different optimization strategies have been proposed. On the one hand, Cain *et al.* [34], who aimed to determine the cocoa shell content (*Theobroma cacao L.*) in cocoa products, evaluated the extraction process by studying its effect on the detection and signal intensity of 30 targeted polar and non-polar potential key metabolites, covering the entire retention and mass range characteristics. As a result, the most efficient extraction was obtained with methanol:water (90:10, *v/v*) with 5 mM ammonium acetate for polar compounds, and with isopropanol:chloroform (80:20, *v/v*) with 20 mM ammonium acetate for non-polar substances. Similarly, Cubero-Leon *et al*. [42], aiming to assess the performance of an SLE extraction method in a study for the identification of the agronomic production system of carrots, used different standards encompassing compounds already reported to be present in carrots, such as vitamins, flavonoids, and phenolic acids. Thus, the authors provided values of repeatability of the extraction method, expressed as the relative standard deviation (RSD, %) of the peak intensity of the spiked samples with the standard compounds from six independent extractions.

Alternatively, in order to obtain the most comprehensive information for sample authentication, Navratilova *et al.* [38] evaluated the extraction efficiency of several solvents employed in an SLE method by comparing the number of detected molecular features. Since the authors aimed to authenticate green tea according to their geographical origin by UHPLC-HRMS fingerprinting, a pool of samples representing three geographical regions was subjected to SLE using water, methanol, and dichloromethane:methanol (50:50, *v/v*). Lately, the extracts were analyzed by UHPLC-HRMS using H-ESI in both positive and negative modes. For peak detection, the use of appropriate software, as well as the establishment of specific parameters (i.e., *m/z* range, intensity threshold, mass and retention time window, and data deisotoping), was required. As a result, dichloromethane:methanol (50:50, *v/v*) was chosen as the extracting solvent since it significantly increased the number of detected molecular features, especially in the H-ESI-positive ionization mode.

Regarding the chromatographic separation, as can be observed in Table 6.3, reversed-phase columns are by far the most selected ones, especially those based on C_{18} chemistry. This fact is mainly due to the great performance of these

columns to separate semi-polar metabolites (e.g., polyphenols, carotenoids, and capsaicinoids), which encompass a wide number of compounds that have been demonstrated to be excellent secondary markers for food authentication issues. Moreover, instead of traditional trimethyl silane (TMS) end capping, some authors proposed polar end-capped C_{18} columns, such as T3, since they enhance the retention of polar metabolites, allowing a broader compound coverage [37, 40]. Instead, *metabolomics* fingerprints based on polar compounds have been obtained by several authors using normal-phase [34], HILIC [39], or amide [44] columns.

Another step that influences in the selectivity of the non-targeted UHPLC-HRMS methods is the ionization source used. Almost all the methodologies available in literature employ H-ESI as the selected interface between the liquid and the gas phase, because of its more remarkable universality in terms of the number of ionizing compounds when compared to other atmospheric pressure ionization (API) sources, such as atmospheric pressure chemical ionization (APCI) or atmospheric pressure photoionization (APPI). In this line, both H-ESI-positive and -negative modes are run (in two different chromatographic runs) to obtain the widest range of detectable compounds, although in some cases only one of them is used. Nevertheless, APCI, APPI, or even multimode ionization source — combining ESI-APCI, ESI-APPI, or APCI-APPI — could reach a broader analysis of compounds with different hydrophobicity, polarity, and volatility, and therefore their application in *metabolomics* fingerprinting could be of high interest.

Non-targeted UHPLC-HRMS analysis generate massive datasets, requiring dedicated software programs to handle the obtained data, with this being a critical step since inappropriate data-processing may compromise the whole analytical method. These software programs aim to convert raw data into a matrix consisting of the retention time, *m/z* values, and the area or signal of each peak detected. Thus, noise filtering, peak picking, peak deconvolution, retention time alignment, and removal of isotopic peaks are carried out by establishing certain parameters such as the peak width, signal-to-noise (S/N) ratio, or mass tolerance. At this point, although the generated matrix can already be subjected to statistical analysis, reduction of data dimensionality is recommended. For instance, Hurkova *et al*. [37], who aimed to find differences between the phytochemical content of lingonberries and cranberries, reduced a matrix containing 7046 molecular features in positive and 1833 in negative ionization mode to only 935 and 307, respectively, by removing isotopic and background peaks as well as using univariate data analysis. Instead, in a UHPLC-HRMS fingerprinting study to find chemical composition differences in chicken eggs stored for different lengths of time, Johnson *et al*. [35] removed all those molecular features with a RSD bigger than 30% in the quality control (QC) samples (sample constructed by pooling equal aliquots of each sample), as well as those with a significance of $P > 0.01$ in a one-way analysis of variance (ANOVA).

As can be observed in Table 6.3, both unsupervised and supervised chemometric techniques are widely used with UHPLC-HRMS fingerprinting data. In the case of unsupervised chemometric techniques, PCA is by far the most used one, allowing an initial interrogation of data. However, since most of the methods reported in this section aim to discover significant differences in terms of molecular features (or unspecified targets), supervised chemometric techniques such as PLS-DA or OPLS-DA perfectly fit with them, not only for classificatory purposes but also to identify the most discriminant biomarkers when required. For instance, on the one hand, Barbosa *et al*. [46] proposed PLS-DA models in order to authenticate paprika samples according to geographical origin through UHPLC-HRMS fingerprinting, reaching a classification rate of 100% for each of the origins studied as shown in Fig. 6.2. In this case, as the fingerprint was enough to discriminate the agricultural origin of paprika, the authors did not carry out further steps of identification.

On the other hand, Hoyos Ossa *et al*. [39], aiming the assessment of protected designation of origin (PDO) for Colombian coffees, used PLS-DA models not only to achieve the best sample classification but also to identify the most discriminant molecular features. In this case, the authors tested three different chromatographic separation modes to select the best one to perform faster analysis. Therefore, for a polar fraction extracted with a mix consisting of acetonitrile:water:formic acid (79:20:1, *v/v/v*), reversed-phase liquid chromatography (RPLC) and HILIC were proposed for the separation of polar and highly polar compounds, respectively; whereas for a semi-polar extract obtained using acetone, RPLC was employed. Then, since the RPLC analysis of the semi-polar extract showed the worst sample classification performance, it was no longer used. Instead, both RPLC and HILIC analysis of the polar extract provided similar classificatory results, but RPLC was chosen because of its advantages over HILIC (e.g., less stabilization time needed, fewer additives for the injection, and more stability against variations).

Moreover, in order to select the most discriminant variables of the built PLS-DA models — each model corresponding to positive or negative ionization mode — the variable importance in projection (VIP) score, which summarizes the overall contribution of the variables in the supervised chemometric models, was chosen. In this line, the variables were listed according to their VIP value, and reduced PLS-DA models were built using the minimum number of variables with higher VIP score that could allow a classification rate above 90%. As a result, only 13 molecular features were required to obtain a satisfactory PLS-DA model for the negative ionization mode, while 30 molecular features could not achieve a correct classification for the positive ionization mode. Therefore, only the markers obtained by the RPLC analysis of the polar extract and in the negative ionization mode were subjected to further identification steps, consisting of the study of the exact mass and the isotopic pattern to propose a candidate molecular formula followed by MS/MS spectra comparison with databases or in silico fragmentation

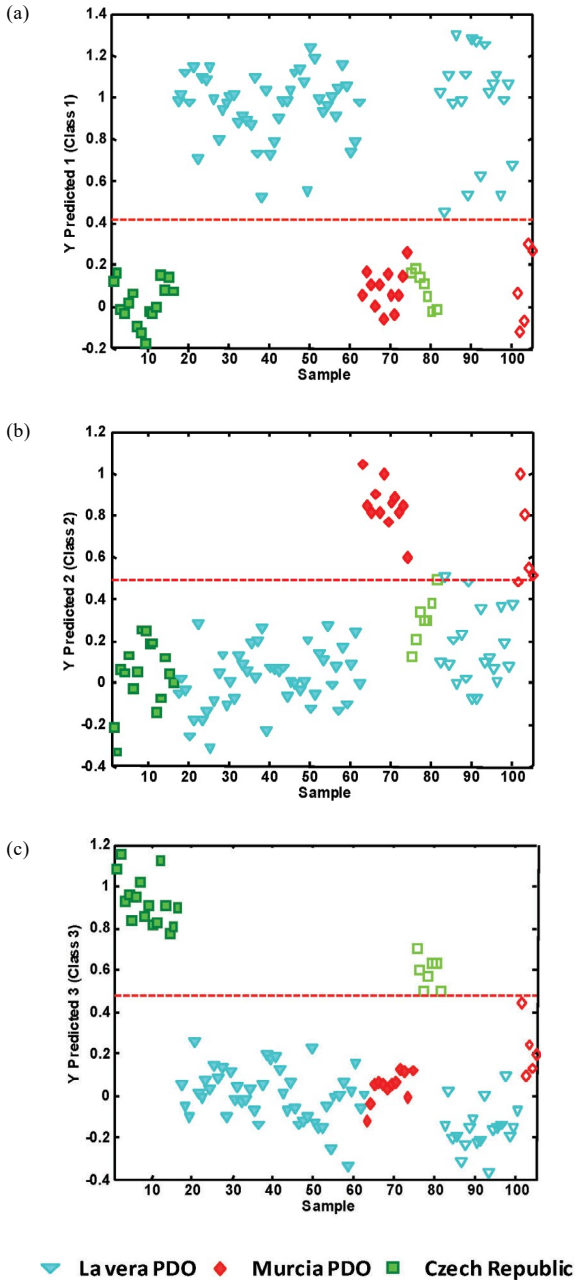


Fig. 6.2. PLS-DA classification plots, obtained by processing UHPLC-HRMS fingerprints, according to sample geographical origin: (a) *La Vera* vs. *Murcia* and the Czech Republic, (b) *Murcia* vs. *La Vera* and the Czech Republic, and (c) the Czech Republic vs. *La Vera* and *Murcia*.

Source: Reproduced from Open Access Ref. [46].

software. For instance, from the 13 biomarkers proposed, eight could be tentatively identified following the explained elucidation process.

As previously mentioned, *metabolomics* fingerprinting approaches are strongly influenced by the global experimental strategy proposed, comprising sample treatment, separation, detection system, data processing, and instrumentation. This fact was evidenced by the study carried out by Díaz *et al*. [40], where two different platforms were compared in the annotation of discriminant metabolites in a fingerprinting approach for the classification of wine samples according to three Spanish PDOs. In this line, although both platforms could correctly distinguish the samples employing the whole fingerprint, a strong divergence among the annotated discriminant metabolites was observed (a total of eight and nine molecular features were identified for each platform, although none of them was common), showing the difficulties of obtaining robust results in terms of identified biomarkers at the end of the workflow.

6.4. Summary and Conclusions

This chapter has presented the role of chromatographic fingerprinting to address the authentication of food through some selected applications published in the last years. In this line, different TLC and LC fingerprinting approaches, mostly in combination with chemometrics, have been discussed.

Food fraud, consisting of an intentional procedure that violates food laws, deceives the consumers, and in some cases poses a risk to human health (e.g., use of toxic or allergenic substances), has grown substantially in the last years. Thus, food authentication has become an issue of great interest for consumers, the food industry, regulatory authorities, as well as the scientific community. Historically, the detection of fraudulent practices has been based on the use of targeted methods, focusing on the quantification of primary markers, directly linked to the authenticity problematic. Despite the great utility of this strategy, when a specific adulterant is suspected, in some cases is insufficient because of the huge number of adulterant substances that can potentially be added in food as well as the increase of complex food authentication issues, such as the identification of the geographical origin or the production method employed. For that reason, profiling strategies based on the qualitative or quantitative determination of secondary markers (indirectly provide information about the food authentication issue) have been widely exploited for authenticity purposes in combination with chemometrics. Nevertheless, when neither primary nor secondary markers are available, non-targeted methods, consisting of the analysis of instrumental responses without assuming any previous knowledge about the sample composition, have emerged as a reliable alternative when combined with chemometrics, especially in the case of *metabolomics* fingerprinting approaches, since *metabolomics* is the closest *omics* discipline to the phenotype of biological systems.

This chapter has focused on the application of HPTLC and LC (with conventional or MS-based detection) fingerprinting in order to address food authentication issues. On the one hand, HPTLC fits ideally with fingerprinting approaches since it provides short chromatograms and excellent reproducibility, with minimal sample preparation, high-throughput screening, and an inexpensive and environmentally friendly analysis. Its use centers on the analysis of liquid or viscous samples coming from the plant kingdom (e.g., honey, wine, and tea). Besides, the selection of stationary and mobile phases, derivatization reagent, and the detection system directly influences on the HPTLC fingerprints selectivity. Regarding data processing, three different strategies can be employed: VI, processing of densitometric data, and chemometrics. However, lower sensitivity and separation performance are the main drawbacks of HPTLC when compared with other chromatographic techniques.

On the other hand, LC fingerprinting has been demonstrated to be a reliable alternative for the authentication of more complex food matrices. For instance, in addition to plant-derived matrices, some authors have analyzed animal-derived products, such as hen eggs. Moreover, as happens in any other chromatographic technique, the obtained LC fingerprints are strongly influenced by the global experimental strategy proposed, comprising sample treatment, separation, detection system, data processing, and instrumentation.

In this chapter, the LC *metabolomics* fingerprinting approach has been classified according to the detection system employed. On the one hand, LC in combination with conventional detection systems — such as UV, FLD, ECD, or CAD — is based on the simultaneous detection of instrumental data points. On the other hand, LC coupled to HRMS adds a third dimension (fingerprints build not only with the retention time and peak signal but also with peak *m/z*) and aids in the detection of unspecified targets. In both cases, chemometrics has been shown to play a crucial role for classificatory purposes, for instance using PLS-DA, for the detection and quantitation of adulteration levels, employing PLS, or in the case of LC-HRMS fingerprinting, for the determination of the most discriminant variables through VIP score for further identification.

In conclusion, both HPTLC and LC *metabolomics* fingerprinting techniques, especially when combined with chemometrics, are powerful tools to address food authentication in terms of quality, geographical indication, and production system. Moreover, the application of a *metabolomics* fingerprinting approach is already being used with other chromatographic or related techniques such as gas chromatography (GC), either multidimensional liquid or gas chromatography (MDLC, MDGC), or even ion mobility (IM), proving its potentiality.

Acknowledgments

This work was supported by the Agency for Administration of University and Research Grants (Generalitat de Catalunya, Spain) under the project 2017 SGR-310.

References

[1] Medina, S., Perestrelo, R., Silva, P., Pereira, J.A.M., and Câmara, J.S. Current trends and recent advances on food authenticity technologies and chemometric approaches. *Trends Food Sci. Technol*, 2019, 85(January):163–176. Available from: https://doi.org/10.1016/j.tifs.2019.01.017.

[2] Arrizabalaga-Larrañaga, A., Rodríguez, P., Medina, M., Santos, F.J., and Moyano, E. Simultaneous analysis of natural pigments and E-141i in olive oils by liquid chromatography–tandem mass spectrometry. *Anal. Bioanal. Chem*., 2019, 411(21): 5577–5591.

[3] Barbosa, S., Campmaj◌̊G., Saurina, J., Puignou, L., and Núñez, O. Determination of phenolic compounds in paprika by ultrahigh performance liquid chromatography-tandem mass spectrometry: Application to product designation of origin authentication by chemometrics. *J. Agric. Food Chem.*, 2020, 68(2):591–602.

[4] Cuadros-Rodríguez, L., Ruiz-Samblás, C., Valverde-Som, L., Pérez-Castaño, E., and González-Casado, A. Chromatographic fingerprinting: An innovative approach for food "identitation" and food authentication — A tutorial. *Anal. Chim. Acta*., 2016, 909:9–23.

[5] Ballin, N.Z. and Laursen, K.H. To target or not to target? Definitions and nomenclature for targeted versus non-targeted analytical food authentication. *Trends Food Sci. Technol*., 2019, 86(2018):537–543. Available from: https://doi.org/10.1016/j.tifs.2018.09.025.

[6] Cifuentes, A. Food analysis and foodomics. *J. Chromatogr. A*., 2009, 1216(43):7109.

[7] Creydt, M. and Fischer, M. Omics approaches for food authentication. *Electrophoresis*, 2018, 39(13):1569–1581.

[8] Böhme, K., Calo-Mata, P., Barros-Velázquez, J., and Ortea, I. Recent applications of omics-based technologies to main topics in food authentication. *TrAC — Trends. Anal. Chem*., 2019, 110:221–232.

[9] Esslinger, S., Riedl, J., and Fauhl-Hassek, C. Potential and limitations of non-targeted fingerprinting for authentication of food in official control. *Food Res. Int.*, 2014, 60:189–204. Available from: http://dx.doi.org/10.1016/j.foodres.2013.10.015.

[1O] Juszczak, A.M., Zovko-Koncié, M., and Tomczyk, M. Recent trends in the application of chromatographic techniques in the analysis of luteolin and its derivatives. *Biomolecules*, 9, 2019.

[11] Guzelmeric, E., Ristivojevié, P., Vovk, T., Milojkovié-Opsenica, D., and Yesilada, E. Quality assessment of marketed chamomile tea products by a validated HPTLC method combined with multivariate analysis. *J. Pharm. Biomed. Anal*., 2017, 132:35–45.

[12] Agatonovic-Kustrin, S., Milojkovié-Opsenica, D., Morton, D.W., and Ristivojevié, P. Chemometric characterization of wines according to their HPTLC fingerprints. *Eur. Food Res. Technol*., 2017, 243(4):659–667.

[13] Ofori, H., Hettiarachchi, D., Sostaric, T., Busetti, F., and Boyce, M.C. High-performance thin-layer chromatographic fingerprinting of sandalwood essential oils. *J. Planar. Chromatogr. — Mod TLC*., 2019, 32(3):205–210.

[14] Stanek, N., Teper, D., Kafarski, P., and Jasicka-Misiak, I. Authentication of phacelia honeys (Phacelia tanacetifolia) based on a combination of HPLC and HPTLC analyses as well as spectrophotometric measurements. *Lwt*., 2019, 107(March):199–207.

[15] Stanek, N. and Jasicka-Misiak, I. HPTLC Phenolic Profiles as Useful Tools for the Authentication of Honey. *Food Anal. Methods*., 2018, 11(11):2979–2989.

[16] Makowicz, E., Jasicka-Misiak, I., Teper, D., and Kafarski, P. Botanical origin authentication of polish phacelia honey using the combination of volatile fraction profiling by HS-SPME and Lipophilic Fraction Profiling by HPTLC. *Chromatographia*, 2019, 82(10):1541–1553. Available from: https://doi.org/10.1007/s10337-019-03778-x.

[17] Makowicz, E., Jasicka-Misiak, I., Teper, and D., Kafarski, P. HPTLC fingerprinting — Rapid method for the differentiation of honeys of different botanical origin based on the composition of the lipophilic fractions. *Molecules*, 2018, 23(7): 1811–1822.

[18] Locher, C., Neumann, J., and Sostaric, T. Authentication of honeys of different floral origins via high-performance thin-layer chromatographic fingerprinting. *J. Planar. Chromatogr. — Mod. TLC*, 2017, 30(1):57–62.

[19] Locher, C., Tang, E., Neumann, J., and Sostaric, T. High-performance thin-layer chromatography profiling of Jarrah and Manuka honeys. *J. Planar. Chromatogr — Mod. TLC*, 2018, 31(3):181–189.

[20] Sima, I.A. and Sârbu, C. Evaluation of polyphenolic fingerprints and antioxidant profiles of wild fruits. *Int. J. Food Sci. Technol*., 2016, 51(6):1433–1440.

[21] Hosu, A. and Cimpoiu, C. HPTLC fingerprinting: A useful tool for white wines authentication. *J. Liq. Chromatogr. Relat. Technol*., 2016, 39(5–6):303–7. Available from: http://dx.doi.org/10.1080/10826076.2016.1163470.

[22] Ristivojević, P., Trifković, J., Andrić, F., and Milojković-Opsenica, D. Recent trends in image evaluation of HPTLC chromatograms. *J. Liq. Chromatogr. Relat. Technol*., 2020, 0(0):1–9. Available from: https://doi.org/10.1080/10826076.2020.1725555.

[23] Cetó X., Sánchez, C., Serrano, N., Díaz-Cruz, J.M., and Núñez, O. Authentication of paprika using HPLC-UV fingerprints. *Lwt*, 2020, 124(February).

[24] Pérez-Ràfols, C., Subirats, X., Serrano, N., and Díaz-Cruz, J.M. New discrimination tools for harvest year and varieties of white wines based on hydrophilic interaction liquid chromatography with amperometric detection. *Talanta,* 2019, 201(January): 104–110. Available from: https://doi.org/10.1016/j.talanta.2019.03.099.

[25] Martín-Torres, S., Jiménez-Carvelo, A.M., González-Casado, A., and Cuadros-Rodríguez, L. Differentiation of avocados according to their botanical variety using liquid chromatographic fingerprinting and multivariate classification tree. *J. Sci. Food Agric*., 2019, 99(11):4932–4941.

[26] Jiménez-Carvelo, A.M., Pérez-Castaño, E., González-Casado, A., and Cuadros-Rodríguez, L. One input-class and two input-class classifications for differentiating olive oil from other edible vegetable oils by use of the normal-phase liquid chromatography fingerprint of the methyl-transesterified fraction. *Food Chem*., 2017, 221:1784–1791.

[27] Bajoub, A., Medina-Rodríguez, S., Gómez-Romero, M., Ajal, E.A., Bagur-González, M.G., Fernández-Gutiérrez, A,, *et al*. Assessing the varietal origin of extra-virgin

olive oil using liquid chromatography fingerprints of phenolic compound, data fusion and chemometrics. *Food Chem*., 2017, 215:245–255. Available from: http://dx.doi.org/10.1016/j.foodchem.2016.07.140.

[28] Bikrani, S., Jiménez-Carvelo, A.M., Nechar, M., Bagur-González, M.G., Souhail, B., and Cuadros-Rodríguez, L. Authentication of the geographical origin of margarines and fat-spread products from liquid chromatographic UV-absorption fingerprints and chemometrics, 2019, 8:588.

[29] Campmaj$\acute{o}$G., Cayero, L., Saurina, J., and Núñez, O. Classification of hen eggs by HPLC-UV fingerprinting and chemometric methods. *Foods*, 2019, 8:310.

[30] Campmaj$\acute{o}$G., Saez-Vigo, R., Saurina, J., and Núñez, O. 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*., 2020, 114:107265. Available from: https://doi.org/10.1016/j.foodcont.2020.107265.

[31] Pardo-Mates, N., Vera, A., Barbosa, S., Hidalgo-Serrano, M., Núñez, O., Saurina, J., *et al*. Characterization, classification and authentication of fruit-based extracts by means of HPLC-UV chromatographic fingerprints, polyphenolic profiles and chemometric methods. *Food Chem*., 2017, 221:29–38. Available from: http://dx.doi.org/10.1016/j.foodchem.2016.10.033.

[32] Puigventos, L., Nunez, O., and Saurina, J. HPLC Fingerprints for the authentication of cranberry-based products based on multivariate calibration approaches. *Curr. Anal. Chem*., 2016, 13(3):256–261.

[33] Rubert, J., Zachariasova, M., and Hajslova, J. Advances in high-resolution mass spectrometry based on metabolomics studies for food — a review. Food Addit Contam — *Part A Chem. Anal. Control. Expo. Risk. Assess*., 2015, 32(10): 1685–1708. Available from: http://dx.doi.org/10.1080/19440049.2015.1084539.

[34] Cain, N., Alka, O., Segelke, T., von Wuthenau, K., Kohlbacher, O., and Fischer, M. Food fingerprinting: Mass spectrometric determination of the cocoa shell content (*Theobroma cacao L.*) in cocoa products by HPLC-QTOF-MS. *Food Chem*., 2019, 298(April):125013. Available from: https://doi.org/10.1016/j.foodchem.2019.125013.

[35] Johnson, A.E., Sidwick, K.L., Pirgozliev, V.R., Edge, A., and Thompson, D.F. Metabonomic profiling of chicken eggs during storage using high-performance liquid chromatography-quadrupole time-of-flight mass spectrometry. *Anal. Chem*., 2018, 90(12):7489–7494.

[36] Cavanna, D., Catellani, D., Dall'asta, C., and Suman, M. Egg product freshness evaluation: A metabolomic approach. *J. Mass. Spectrom*., 2018, 53(9):849–861.

[37] Hurkova, K., Uttl, L., Rubert, J., Navratilova, K., Kocourek, V., Stranska-Zachariasova, M., *et al*. Cranberries versus lingonberries: A challenging authentication of similar Vaccinium fruit. *Food Chem*., 2019, 284(2018):162–170. Available from: https://doi.org/10.1016/j.foodchem.2019.01.014.

[38] Navratilova, K., Hrbek, V., Kratky, F., Hurkova, K., Tomaniova, M., Pulkrabova, J., *et al*. Green tea: Authentication of geographic origin based on UHPLC-HRMS fingerprints. *J. Food Compos. Anal*., 2019, 78(2018):121–128.

[39] Hoyos Ossa, D.E., Gil-Solsona, R., Peñuela, G.A., Sancho, J.V., and Hernández, F.J. Assessment of protected designation of origin for Colombian coffees based on HRMS-based metabolomics. *Food Chem*., 2018, 250(2017):89–97. Available from: https://doi.org/10.1016/j.foodchem.2018.01.038.

[40] Díaz, R., Gallart-Ayala, H., Sancho, J.V., Nuñez, O., Zamora, T., Martins, C.P.B., *et al*. Told through the wine: A liquid chromatography-mass spectrometry interplatform comparison reveals the influence of the global approach on the final annotated metabolites in non-targeted metabolomics. *J. Chromatogr. A*., 2016, 1433:90–97.

[41] Martinez Bueno, M.J., Diaz-Galiano, F.J., Rajski, L., Cutillas, V., and Fernández-Alba, A.R. A non-targeted metabolomic approach to identify food markers to support discrimination between organic and conventional tomato crops. *J. Chromatogr. A*., 2018, 1546:66–76. Available from: https://doi.org/10.1016/j.chroma.2018.03.002.

[42] Cubero-Leon, E., De Rudder, O., and Maquet, A. Metabolomics for organic food authentication: Results from a long-term field study in carrots. *Food Chem*., 2018, 239(555):760–770. Available from: http://dx.doi.org/10.1016/j.foodchem.2017.06.161.

[43] Zhang, J., Yu, Q., Cheng, H., Ge, Y., Liu, H., Ye, X., *et al*. Metabolomic approach for the authentication of berry fruit juice by liquid chromatography quadrupole time-of-flight mass spectrometry coupled to chemometrics. *J. Agric. Food Chem*., 2018, 66(30):8199–8208.

[44] Arapitsas, P., Corte, A. Della, Gika, H., Narduzzi, L., Mattivi, F., and Theodoridis, G. Studying the effect of storage conditions on the metabolite content of red wine using HILIC LC-MS based metabolomics. *Food Chem*., 2016, 197:1331–1340.

[45] Li, Y., Jin, Y., Yang, S., Zhang, W., Zhang, J., Zhao, W., *et al*. Strategy for comparative untargeted metabolomics reveals honey markers of different floral and geographic origins using ultrahigh-performance liquid chromatography-hybrid quadrupole-orbitrap mass spectrometry. *J. Chromatogr. A*., 2017, 1499:78–89.

[46] Barbosa, S., Saurina, J., Puignou, L.,and Núñez, O. Classification and authentication of paprika by UHPLC-HRMS fingerprinting and multivariate calibration methods (PCA and PLS-DA). *Foods*, 2020, 9(4):486–495.

[47] Belmonte-Sánchez, J.R., Romero-González, R., Arrebola, F.J., Vidal, J.L.M., and Garrido Frenich, A. An innovative metabolomic approach for golden rum classification combining ultrahigh-performance liquid chromatography-orbitrap mass spectrometry and chemometric strategies. *J. Agric. Food Chem*., 2019, 67(4): 1302–1311.

[48] Lucci, P., Saurina, J., and Núñez, O. Trends in LC-MS and LC-HRMS analysis and characterization of polyphenols in food. *TrAC — Trends. Anal. Chem*., 2017, 88:1–24.

Furthermore, as previously mentioned, aside from chromatographic-based non-targeted methods, direct MS-based techniques have gained relevance in food authentication through a fingerprinting approach. DIMS, FIA–MS, and AIMS are the most employed among these techniques. However, only DIMS and FIA–MS are herein introduced since the direct MS-based methods developed in this thesis are based on them.

In this context, both are the main analytical techniques used to directly introduce liquid samples (*i.e.*, raw samples or sample extracts) into the MS system. On the one hand, DIMS directly introduces the sample through a syringe pump, leading to its continuous ionisation. On the other hand, in FIA–MS, a small volume of sample is injected into a continuous stream that carries the sample bolus up to the mass spectrometer ion source [2]. Thus, the major advantages of these techniques over conventional chromatographic ones are the greater method simplicity and the lower analysis time, increasing the analytical throughput and allowing the analysis of many more samples. Moreover, the lack of chromatographic separation reduces solvent consumption and improves the method's repeatability since no retention time drifts can affect the analysis. Instead, their main limitations are the impossibility of distinguishing isomeric compounds and the ion suppression that can lead to a loss of detection sensitivity [11,12]. Nevertheless, it is noteworthy that in the case of fingerprinting approaches, which do not focus on specific markers, ion suppression due to competition between matrix ions may not be a drawback but another source of discrimination between samples.

Table 3.1 compiles recent DIMS and FIA–MS fingerprinting methods to address different food fraud issues. In this line, different methods based on LRMS have been developed. For instance, Gamboa-Becerra *et al*. [13] used the mass spectra acquired with a single quadrupole by DIMS to classify coffee products by Random Forest models according to their cultivar origin, geographical origin, and processing. Similarly, FIA–MS methods, employing IT technologies, were developed to assess

the cultivar and botanical origin authentication of pumpkin seed oil, cinnamon, and tea samples [14–16].

Instead, some other authors have chosen HRMS instruments, which because of their higher resolution can efficiently resolve near-isobaric compounds, enhancing the selectivity and increasing the total amount of molecular features detected. For example, the geographical origin of garlic samples was addressed through FIA–HRMS (Q-Orbitrap) and OPLS-DA [17]. Besides, Nikou *et al*. [18] proposed a magnetic resonance mass spectrometer (MRMS) to detect the direct infusion high-resolution mass spectrometry (DIHRMS) fingerprints of olive oil, which were subsequently subjected to chemometrics for the geographical origin and production system classification.

Moreover, IMS has become another alternative to increase the selectivity of direct MS-based methodologies. This technique separates the ions in a neutral gas phase based on their mobility, which is strongly related to their charge, size, and shape. In this context, IMS technologies are commonly classified into three categories according to their ion separation mechanism: time-dispersive, space-dispersive, and ion-trapping with selective release techniques [19]. Thus, IMS provides an additional separation dimension to direct MS-based methods, enhancing their selectivity. Besides, it usually improves the method sensitivity due to reducing background noise. Finally, in some cases, it can provide collision cross-section (CCS) values, a complementary parameter for compound identification [20].

Only Masike *et al*. [21] have developed a fingerprinting method combining IMS with a direct MS-based technique. Thus, in this study, non-targeted flow injection analysis–ion mobility spectrometry coupled to high-resolution mass spectrometry (FIA–HRMS), using travelling-wave ion mobility spectrometry (TWIMS), was employed to observe honeybush and rooibos tea sample trends, regarding their botanical origin.

Table 3.1. Compilation of some DIMS and FIA–MS fingerprinting methodologies to address different food authentication issues.

Sample	Extraction	Injection approach	Ion mobility	Mass spectrometry	Data analysis	Ref.
Cultivar/botanical origin						
Pumkin seed oil	Dilution with *n*-propanol:methanol (85:15, *v/v*) with 40 mM ammonia	DI	-	LRMS (IT) Full-scan (*m/z* 100 – 800) H-ESI (-)	LDA	[14]
Cinnamon	SLE with methanol:water (60:40, *v/v*)	FIA Water:acetonitrile 60:40 (*v/v*) with 0.1% formic acid (*v/v*)	-	LRMS (IT) Full-scan (*m/z* 100 – 2000) H-ESI (+)	PCA	[15]
Tea	SLE with water	FIA Water:methanol 50:50 (*v/v*) with 0.1% formic acid (*v/v*)	-	LRMS (LIT) Full-scan (*m/z* 100 – 550) H-ESI (±)	PCA, PLS-DA, and PLS	[16]
Honeybush and rooibos tea	SLE with methanol:water (51:49, *v/v*) with 0.2% formic acid (*v/v*)	FIA Water:acetonitrile 50:50 (*v/v*) with 0.1% formic acid (*v/v*)	TWIMS	HRMS (Q-TOF) Full-scan (*m/z* 100 – 1500) H-ESI (-)	PCA	[21]
Geographical origin						
Garlic	SLE with methanol:water (80:20, *v/v*)	FIA Water and methanol	-	HRMS (TOF) Full-scan (*m/z* 50 – 1200) H-ESI (±)	PCA and OPLS-DA	[17]

Processing						
Lettuce	SLE with methanol:water (60:40, *v/v*)	FIA Water:acetonitrile 60:40 (*v/v*) with 0.1% formic acid (*v/v*)	-	HRMS (LTQ-Orbitrap) Full-scan (*m/z* 100 – 1000) H-ESI (-)	PCA and ANOVA-PCA	[22]
Daylily flowers	SLE with methanol:water (70:30, *v/v*)	FIA Water:acetonitrile 50:50 (*v/v*) with 0.1% formic acid (*v/v*)	-	HRMS (LTQ-Orbitrap) Full-scan (*m/z* 100 – 1500) ESI (-)	PCA	[23]
Production system						
Tomato	SLE with methanol:water (50:50, *v/v*) with 0.1% formic acid (*v/v*)	DI	-	HRMS (Q-Orbitrap) Full-scan (*m/z* 150 – 1700) ESI (+)	Algorithm adaptive boosting	[24]
Cultivar/botanical origin and geographical origin						
Maca (*Lepidium meyenii*)	SLE with methanol:water (70:30, *v/v*)	FIA Water:acetonitrile 20:80 (*v/v*) with 0.1% formic acid (*v/v*)	-	HRMS (LTQ-Orbitrap) Full-scan (*m/z* 100 – 1000) H-ESI (±)	PCA and MANOVA-PCA	[25]
Wolfberry	SLE with ethanol:water (80:20, *v/v*)	FIA Water:acetonitrile 50:50 (*v/v*) with 0.1% formic acid (*v/v*)	-	HRMS (Q-TOF) Full-scan (*m/z* 100 – 1000) H-ESI (-)	PLS-DA	[26]

Geographical origin and production system					
Olive oil	[a]Dilution in dichloromethane:methanol (50:50, *v/v*) with 10 mM ammonium acetate [b]LLE with methanol:water (80:20, *v/v*), evaporation, and dissolution in methanol:water (50:50, *v/v*) with 10 mM ammonium acetate	DI	-	HRMS (MRMS) Full-scan ([a]*m/z* 147 – 3000 and [b]*m/z* 107 – 3000) H-ESI (-)	PCA and OPLS-DA	[18]
Cultivar/botanical origin, geographical origin, and processing					
Coffee	SLE with methanol	DI	-	LRMS (quadrupole) Full-scan (*m/z* 15 – 2000) H-ESI (+)	HCA and Random Forest	[13]

Analysis of variance-principal component analysis (ANOVA-PCA), linear ion trap (LIT), magnetic resonance mass spectrometer (MRMS), multivariate analysis of variance-principal component analysis (MANOVA-PCA), travelling-wave ion mobility spectrometry (TWIMS)

3.2. RESULTS

This section comprises seven scientific publications in which metabolomic fingerprinting approaches and chemometrics were applied to assess different food fraud cases.

First, several methods based on LC with spectroscopic detection systems (*i.e.*, LC-UV and LC-FLD) are described. In this line, Publication IV and Publication V describe the application of LC-UV fingerprinting to classify nut samples —according to their type— and hen egg samples —concerning their production system—, respectively. Then, in Publication VI, LC-FLD fingerprints were used as chemical markers to assess the nut type classification and the adulteration of almond-based products with hazelnut or peanut. Instead, Publication VII deals with the geographical origin classification of paprika samples by LC-FLD fingerprinting, as well as the quantitation of their adulterant blends.

Then, in contrast, Publication VIII contains the development of an exhaustive non-targeted LC−HRMS method, aiming for nut-type classification and the tentative annotation/identification of the most discriminant metabolites. Moreover, as a case study, the adulteration of almond-based products with hazelnut or peanut was addressed through a targeted LC−HRMS method based on the previously identified chemical markers.

Finally, direct MS-based fingerprinting approaches were also tested for food authentication purposes. For instance, Publication IX reports the application of FIA–HRMS fingerprinting to several food authentication issues: the geographical origin of Spanish red wines, the geographical origin of three European paprikas, and the botanical origin authentication and quality assessment of olive oils. Instead, in Publication X, fingerprints obtained through differential mobility spectrometry coupled to mass spectrometry (DMS−MS), which is a space-dispersive IMS technology, were evaluated as markers for classifying Spanish paprika.

In summary, this chapter's results are presented in the subsequent publications.

Publication IV: Scientific article

Non-targeted HPLC-UV fingerprinting as chemical descriptors for the classification and authentication of nuts by multivariate chemometric methods.

Campmajó, G.; Navarro, G. J.; Núñez, N.; Puignou, L.; Saurina, J.; Núñez, O.

Sensors. 2019, 19, 1388.

Publication V: Scientific article

Classification of hen eggs by HPLC-UV fingerprinting and chemometric methods.

Campmajó, G.; Cayero, L.; Saurina, J.; Núñez, O.

Foods. 2019, 8, 310.

Publication VI: Scientific article

High-performance liquid chromatography with fluorescence detection fingerprinting combined with chemometrics for nut classification and the detection and quantitation of almond-based product adulterations.

Campmajó, G.; Saez-Vigo, R.; Saurina, J.; Núñez, O.

Food Control. 2020, 114, 107265.

Publication VII: Scientific article

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

Campmajó, G.; Rodríguez-Javier, L. R.; Saurina, J.; Núñez, O.

Food Chemistry. 2021, 352, 129397.

Publication VIII: Scientific article

Liquid chromatography coupled to high-resolution mass spectrometry for nut classification and marker identification.

Campmajó, G.; Saurina, J.; Núñez, O.

Food Control. 2023, 152, 109834.

Publication IX: Scientific article

FIA–HRMS fingerprinting subjected to chemometrics as a valuable tool to address food classification and authentication: Application to red wine, paprika, and vegetable oil samples.

Campmajó, G.; Saurina, J.; Núñez, O.

Food Chemistry. 2022, 373, 131491.

Publication X: Scientific article

Differential mobility spectrometry coupled to mass spectrometry (DMS–MS) for the classification of Spanish PDO paprika.

Campmajó, G.; Saurina, J.; Núñez, O.; Sentellas, S.

Food Chemistry. 2022, 390, 133141.

3.2.1. PUBLICATION IV

Non-targeted HPLC-UV fingerprinting as chemical descriptors for the classification and authentication of nuts by multivariate chemometric methods.

Campmajó, G.; Navarro, G. J.; Núñez, N.; Puignou, L.; Saurina, J.; Núñez, O.

Sensors. 2019, 19, 1388.

sensors

MDPI

Article

Non-Targeted HPLC-UV Fingerprinting as Chemical Descriptors for the Classification and Authentication of Nuts by Multivariate Chemometric Methods

Guillem Campmajó [1], Gemma J. Navarro [1], Nerea Núñez [1], Lluís Puignou [1,2], Javier Saurina [1,2] and Oscar Núñez [1,2,3,*]

[1] Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, Martí i Franquès, 1-11, E08028 Barcelona, Spain; campma03@gmail.com (G.C.); g.navarro.albiol@gmail.com (G.J.N.); nereant7@gmail.com (N.N.); lluis.puignou@ub.edu (L.P.); xavi.saurina@ub.edu (J.S.)

[2] Research Institute in Food Nutrition and Food Safety, University of Barcelona, Recinte Torribera, Av. Prat de la Riba 171, Edifici de Recerca (Gaudí), Santa Coloma de Gramenet, E08921 Barcelona, Spain

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

* Correspondence: oscar.nunez@ub.edu; Tel.: +34-93-403-3706

Received: 21 February 2019; Accepted: 19 March 2019; Published: 21 March 2019

check far updates

Abstract: Recently, the authenticity of food products has become a great social concern. Considering the complexity of the food chain and that many players are involved between production and consumption; food adulteration practices are rising as it is easy to conduct fraud without being detected. This is the case for nut fruit processed products, such as almond flours, that can be adulterated with cheaper nuts (hazelnuts or peanuts), giving rise to not only economic fraud but also important effects on human health. Non-targeted HPLC-UV chromatographic fingerprints were evaluated as chemical descriptors to achieve nut sample characterization and classification using multivariate chemometric methods. Nut samples were extracted by sonication and centrifugation, and defatted with hexane; extracting procedure and conditions were optimized to maximize the generation of enough discriminant features. The obtained HPLC-UV chromatographic fingerprints were then analyzed by means of principal component analysis (PCA) and partial least squares-discriminant analysis (PLS-DA) to carry out the classification of nut samples. The proposed methodology allowed the classification of samples not only according to the type of nut but also based on the nut thermal treatment employed (natural, fried or toasted products).

Keywords: HPLC-UV; fingerprinting; food authentication; nuts; principal component analysis; partial least squares-discriminant analysis

1. Introduction

Nowadays, food manufacturers, researchers, and society, in general, have become very interested in the quality of food products, not only from the nutritional point of view but also in relation to food safety issues or regarding the presence of bioactive substances with beneficial properties for consumers (functional foods, nutraceuticals, etc.). Within this context and considering the complexity of the food chain in a globalized world where many players are involved between production and consumption, food manipulation and adulteration practices are rising because of the ease of perpetrating fraud that may remain undetected. In general, food adulteration is carried out to increase volume, to mask the presence of inferior quality components, and to replace the authentic substances for the seller's economic gain. For instance, a common fraud is the employment of a cheaper similar ingredient, which the consumer has difficulty recognizing and which is difficult to detect by current analytical methodologies. However, it must be considered that the deliberate adulteration of food and its

misrepresentation to deceive final consumers is illegal worldwide [1]. In addition, depending on the nature of the adulterants, the fraudulent food product can also represent a health risk for the consumer when prohibited substances are added to deceive its organoleptic properties or when the adulterant can produce allergy episodes. Sixty-eight percent of food fraud violations are perpetrated in animal and vegetable products with high-fat content becoming a crucial issue for food processing industries [2]. Therefore, the ability to guarantee food integrity and authenticity is a major concern in the food industry for both economic and safety reasons, requiring new analytical methodologies.

Nuts are food products with important health benefits to humans. Even though approximately 70% of their weight is attributed to fat, the amount of saturated fatty acids is very low. Thus, their high unsaturated fatty acid content induces the reduction of both total and low-density lipoprotein cholesterols [3], correlating their consumption to the decrease of ischemic heart disease [4]. They have also been shown to be important sources of antioxidant compounds, such as polyphenols [5–8], which are secondary metabolites and the largest group of phytochemicals in plants. In fact, the main dietary sources of total polyphenols are nuts, followed by tea and coffee, rich in flavanols and hydroxycinnamic acids, respectively [9]. Walnuts, almonds, and hazelnuts are the most commonly consumed nuts in the European countries (either raw, fried or toasted) where tree nuts are more consumed than peanuts or seeds. Furthermore, walnuts and almonds contain high levels of total polyphenols in comparison to other polyphenol-rich foods, such as apple juice or red wine [8].

Nuts should be considered as highly exposed to fraud practices since they can be relabeled with old or expired stock samples or replaced with cheaper ones, representing serious problems to consumers with allergies and intolerances [10]. Of the 177 cases of fraudulent practices reported in the European Union in 2016, 4% were related to nuts and seeds [2]. Several analytical methodologies in combination with chemometrics have been described in the literature to address nut authentication and to detect its fraud. For example, the determination of fatty acids by gas chromatography-flame ionization detector (GC-FID) in combination with principal component analysis (PCA) was proposed for the authentication of almond cultivars [11], for the classification and authentication of Iranian walnuts according to their geographical origin [12], and to authenticate several almond genotypes grown in Serbia [13]. In this last work, fatty acid content was combined with the determination of some phenolic compounds by ultra-high performance liquid chromatography-tandem mass spectrometry (UHPLC-MS/MS). The fatty acid profile obtained by GC-FID and analyzed by PCA, linear discriminant analysis (LDA), and partial least squares (PLS) regression was also proposed to detect and quantify the fraudulent addition of apricot kernel in almonds [14]. However, time-consuming sample treatments are typically needed for fatty acid determination by GC techniques, which require derivatization steps to obtain volatile fatty acid methyl esters. Thus, spectroscopic techniques, such as near-infrared (NIR), have been employed for the classification of hazelnuts according to protected designations of origin (PDO) [15] or to address the discrimination of peanuts from bulk cereals and nuts [16]. Multi-elemental analysis fingerprinting based on inductively coupled plasma optical emission measurements (ICP-OES) by determining 10 metallic elements and their analysis by PCA, LDA, and PLS were also described in adulteration studies of almond powder samples with peanut [10].

Most of the previously described methods are based on targeted approaches in which a specific group of known selected chemicals or belonging to the same family is determined. Generally, as the concentration (or peak signal) of these targeted compounds is used as a food feature (marker) to address food integrity and authenticity, this approach requires a previous quantification step using standards for each one. However, when dealing with food products, which are very complex matrices, the quantification of some of these chemicals may be a difficult task, especially due to the possibility of unknown interfering compounds. Hence, nowadays the use of non-targeted fingerprinting approaches, in which analytical signals related to the composition of foodstuffs are employed in a non-selective way (i.e., spectrum or chromatogram), are gaining popularity in food authentication [17–21]. Mathematical processing of the information in such fingerprints may allow the characterization and/or authentication of foodstuffs.

In this work, a non-targeted high-performance liquid chromatography with ultraviolet detection (HPLC-UV) fingerprinting method has been evaluated for the classification and authentication of different types of nuts. For that purpose, a total of 149 nut samples belonging to different classes (almonds, cashew nuts, hazelnuts, macadamia nuts, peanuts, pinions, pistachios, pumpkin seeds, sunflower seeds, and walnuts), some of them in different presentation formats according to the thermal treatment applied (natural, fried or toasted), were analyzed. Samples were extracted by a simple solid–liquid extraction method and the extracting solvent composition was optimized to maximize the total amount of non-targeted components extracted. Data corresponding to the non-targeted HPLC-UV fingerprints recorded at 280 nm were considered as a source of potential descriptors to be exploited for the characterization and classification of the analyzed nut samples by exploratory PCA and supervised partial least squares-discriminant analysis (PLS-DA).

2. Materials and Methods

2.1. Chemicals and Standard Solutions

All the reagents employed, unless otherwise stated, were of analytical grade. Methanol and acetonitrile (both UHPLC-gradient grade) were purchased from Panreac (Barcelona, Spain). Acetone, hexane, and formic acid (96%) were obtained from Sigma-Aldrich (St. Louis, MO, USA) and absolute ethanol from VWR International Eurolab S.L. (Barcelona, Spain). Water was purified using an Elix 3 coupled to a Milli-Q system (Millipore, Bedford, MA, USA) and filtered through a 0.22 μm nylon membrane integrated into the Milli-Q system.

2.2. Instrumentation

An Agilent 1100 Series HPLC instrument was used to obtain the HPLC-UV chromatograms employed as the data in the chemometric methods. The instrument was equipped with a quaternary pump (G1311A), a degasser (G1322A), an autosampler (G1329A), a diode-array detector (G1315B) and a computer with the Agilent Chemstation software, all of them from Agilent Technologies (Waldbronn, Germany). HPLC-UV runs were obtained in reversed-phase mode by employing a porous-shell Kinetex C18 column (1000 mm × 4.6 mm I.D., 2.6 μm particle size) from Phenomenex (Torrance, CA, USA) at room temperature. Gradient elution mode using 0.1% (v/v) formic acid aqueous solution (solvent A) and methanol (solvent B) as mobile phase components was applied as follows: 0–30 min, linear gradient from 5 to 75% B; 30–32.5 min, linear gradient from 75 to 95% B; 32.5–35 min, isocratic step at 95% B; 35–35.1 min, back to initial conditions at 5% B; and from 35.1–40 min, at 5% B for column re-equilibration. A mobile phase flow rate of 0.4 mL/min and an injection volume of 5 μL were applied. HPLC-UV chromatographic fingerprints registered at 280 nm were employed.

2.3. Samples and Sample Treatment

A total of 149 nut samples, obtained from Barcelona markets, were analyzed. Table 1 shows the description, including the abbreviation used in this manuscript, the number of samples regarding the sample treatment applied (natural, fried or toasted products) and the sample origin.

Sample extraction was performed as follows: 0.125 g of crushed and homogenized nut sample were weighed into a 15 mL polypropylene tube and extracted with 3 mL of acetone:water (70:30 *v*/*v*) solution by stirring in a Vortex (Stuart, Stone, United Kingdom) for 1 min and by sonication for 15 min (5510 Branson ultrasonic bath, Hampton, NH, USA). The solutions were then centrifuged for 30 min at 3400 rpm (Rotanta 460 RS Centrifuge, Hettich, Germany). The supernatant extract was transferred to a new 15 mL polypropylene tube, defatted with 3 mL hexane by stirring in a Vortex for 1 min and centrifuged again for 15 min at 3400 rpm. Finally, the sample extract was filtered through 0.22 μm nylon filter (Scharlab, Sentmenat, Barcelona, Spain) and stored at −18 °C in 2 mL glass injection vial until HPLC-UV analysis.

Table 1. Description of the nut samples analyzed.

Nut Type	Abbreviation	Number of Samples			Origin
		Natural	Fried	Toasted	
Almonds	AL	10	10	10	Spain-USA
Cashew nuts	CN	-	10	-	Brazil
Hazelnuts	HN	10	-	10	Spain-Turkey
Macadamia nuts	MN	10	-	-	South Africa
Peanuts	PN	-	10	10	Spain-Brazil-China-USA
Pinions	PI	10	-	-	Spain-China
Pistachios	PT	-	-	9	Spain-Germany-Iran
Pumpkin seeds	PS	-	10	10	Austria-China
Sunflower seeds	SS	-	-	9	Spain
Walnuts	WN	10	-	-	Chile-USA

In addition, a quality control (QC) was prepared by mixing 50 µL of each nut sample extract. This QC was employed to evaluate the repeatability of the method and the robustness of the chemometric results.

Samples were randomly analyzed with the proposed HPLC-UV method. A QC and a blank of water were injected at the beginning of the sequence and every 10 sample injections.

2.4. Data Analysis

SOLO chemometric software from Eigenvector Research was used for calculations with PCA and PLS-DA [22]. A detailed description of the theoretical background of these methods is given elsewhere [23].

The X-data matrices to be treated by PCA and PLS-DA consisted of the HPLC-UV chromatographic fingerprints obtained at 280 nm (absorbance intensities at different retention times). HPLC-UV chromatograms were pretreated to improve the data quality while minimizing solvent and matrix interferences, peak shifting, and baseline drifts. For additional details see reference [24]. The Y-data matrix in PLS-DA consisted of the sample class. Scatter plots of scores and loadings of the principal components (PCs), in PCA, and latent variables (LVs), in PLS-DA, were used to investigate the structure of maps of samples and variables, respectively. The plots of scores show the distribution of the samples revealing patterns that can be correlated to sample characteristics. The plots of loadings provide information on the most descriptive features contributing to sample discrimination.

3. Results and Discussion

3.1. Sample Treatment: Optimization of the Extracting Solvent Composition

In the present work, a simple solid–liquid extraction procedure by stirring and sonication, followed by a defatting step with hexane, was proposed for the extraction of non-targeted phytochemicals from nut samples. An optimization of the extracting solvent composition was performed to maximize the total amount of extracted compounds. Thus, seventeen sample extraction conditions were evaluated for three nut samples (almond, hazelnut, and walnut), as they were extracted with different solvent mixtures including pure water, ethanol, methanol, acetonitrile, and acetone, as well as with organic solvent:water mixtures in 50:50, 70:30, and 80:20 (*v*/*v*) ratios. Considering the high polyphenolic content of nuts, a reversed-phase HPLC-UV method, previously developed for the determination of polyphenols in wine samples [25], was employed with some modifications by using a porous-shell C18 column and a gradient elution with 0.1% formic acid aqueous solution and methanol as mobile phase components (conditions described in Section 2.2). As an example, Figure 1 shows the non-targeted HPLC-UV chromatographic fingerprint obtained for a walnut sample extracted with acetone:water 70:30 (*v*/*v*).

As seen in the figure, a high signal caused by the absorption of the acetone present in the extracting solvent was detected close to the dead volume (4–5 min). Moreover, a significant number of peak signals related to the non-targeted extracted compounds were obtained. In fact, to select the best extracting solvent, the total peak area (sum of each detected compound peak area without considering the one

related to the organic solvent) was employed as a chemical parameter and the obtained non-targeted HPLC-UV chromatographic fingerprints were evaluated. Thus, Figure 2 shows the resulting bar plot representing the total peak area of the extracted components for the three nut samples under the different extraction conditions. Consistent with the fact that walnut is the nut with higher total polyphenolic content (3733 ± 1190 mg/100 g dry mass expressed as gallic acid equivalent, GAE) [26], it presents a larger quantity of extracted chemicals than almond and hazelnut. Regarding the extracting solvent composition, it was observed that the extraction capacity of pure organic solvents was not as effective as that achieved by pure water, except for the walnut sample. In contrast, mixtures of water with different amounts of organic solvents were quite effective, although a high percentage of it seemed to keep the extraction capacity constant or decrease it. Among the different extracting conditions evaluated, a great variety in extraction capacities was observed depending on both the extracting solvent employed and the nut sample matrix studied. For instance, acetone:water (50:50 and 70:30 *v*/*v*) provided the best extraction efficiencies for the walnut matrix; pure water and with mixtures of acetonitrile:water and acetone:water (both at 50:50 and 70:30 *v*/*v*) for hazelnut matrix; whereas for the almond matrix, the best extraction results were achieved when employing pure water and acetonitrile:water (80:20 *v*/*v*). Therefore, taking into consideration that the aim of the present work is to study the viability of a non-targeted HPLC-UV fingerprinting strategy to address nut sample classification, as a compromise, acetone:water (70:30 *v*/*v*) was chosen as the optimal extraction solvent for future experiments because of its satisfactory extraction efficiencies for the hazelnut and walnut samples.

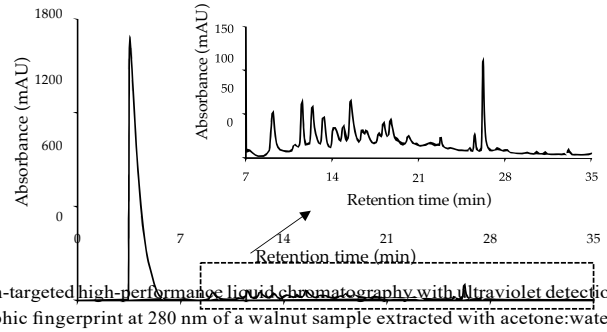


Figure 1. Non-targeted high-performance liquid chromatography with ultraviolet detection (HPLC-UV) chromatographic fingerprint at 280 nm of a walnut sample extracted with acetone:water 70:30 (*v*/*v*).

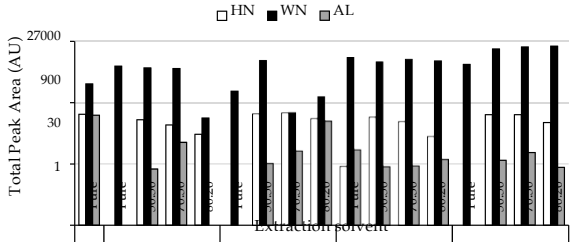


Figure 2. Total area of extracted phytochemical components (represented in logarithmic scale) obtained from HPLC-UV chromatographic fingerprints of a hazelnut (HN), walnut (WN) and almond (AL) sample extracted using different extraction solvent compositions. Ratios indicated correspond to organic solvent:water (*v*/*v*).

3.2. Non-Targeted HPLC-UV Chromatographic Fingerprints

Once the extracting solvent was selected, the studied samples described in Table 1 were extracted, and their respective extracts were randomly analyzed, along with a QC and a water blank every ten sample injections, with the proposed reversed-phase HPLC-UV method.

As an example, Figure 3 shows the chromatograms acquired at 280 nm for a selected sample within the ten types of nut samples under study. As can be seen, remarkable differences were obtained depending on the nut sample matrix analyzed, such as the number and distribution of extracted compounds or their signal intensity, with walnut and sunflower seed samples showing the highest ones. Therefore, due to these differences, the obtained fingerprints were proposed as possible chemical descriptors to achieve sample classification through the employment of chemometric methods.

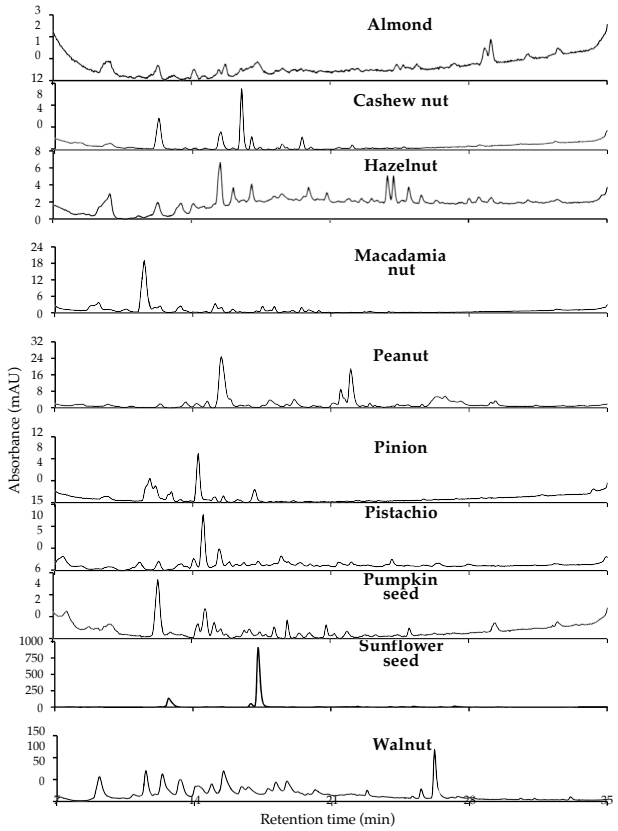


Figure 3. Non-targeted HPLC-UV chromatographic fingerprint acquired at 280 nm for a selected sample within each nut sample type. Chromatograms displayed only from minute 7 to 35 to remove absorption of the extracting solvent and column pre-conditioning step.

3.3. Chemometric Data Analysis

3.3.1. Characterization of Samples according to Nut Type: Non-Supervised PCA Study

The capability of the obtained non-targeted HPLC-UV chromatographic fingerprints to be used as discriminant chemical descriptors for sample classification and authentication depending on the nut fruit involved, independently of the presentation format (natural, fried or toasted), was evaluated by PCA. For that purpose, a 163 × 6001 dimension data matrix, which consisted of the absorbance signals recorded as a function of retention time for the 149 analyzed samples, as well as the QCs, was built. Moreover, autoscaling pretreatment was chosen to provide similar weight to all the variables (overall bioactive compound signals) as it suppresses differences in the magnitude and amplitude scale.

As a first result, Figure S1 (supplementary material) shows the obtained scores plot of PC1 vs. PC2, which does not present a compact group of the QCs as should be expected. In fact, when labelling them, a distribution based on their injection order is observed from the top-right section of the scores plot to the bottom-left section, which reveals a trend associated to the HPLC-UV sample sequence employed. Thus, this chromatographic behavior along the sequence affects systematically the fingerprinting signal registered for each sample, and therefore, the chemometric results displayed in this figure cannot be used, and a correction is required.

As QCs are injections of the same extract, they not only allow the detection of any possible instrumental or chromatographic issue but also to correct them. In this case, the fingerprints obtained for each sample were normalized by dividing their absorbance signal variables by those of the closest QC in the sequence, whereas each QC was divided by itself. After performing this correction, QC samples appear in the same position in the PCA scores plot (see Figure 4A), and chemometric results can be discussed. As can be seen, walnut samples are clearly separated at the left part of the plot from the other nut samples (displaying negative PC1 score values). This difference could be related to the fact that they are among the matrices with higher polyphenolic content, as previously mentioned.

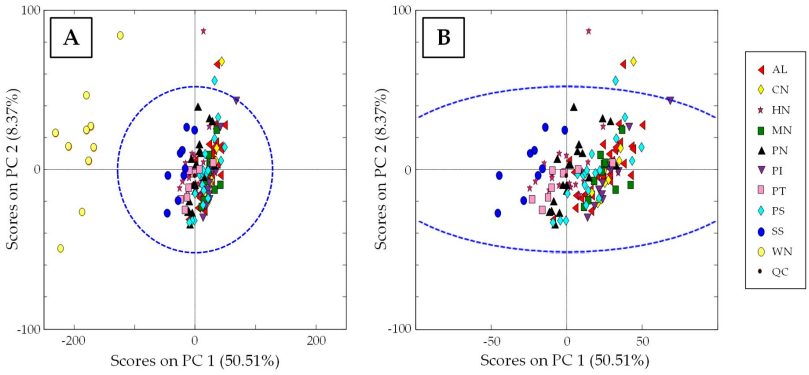


Figure 4. (**A**) PCA Score plot of PC1 vs. PC2 when using non-targeted HPLC-UV chromatographic fingerprints registered at 280 nm as chemical descriptors for all the analyzed samples. (**B**) Same PCA plot without including walnut samples.

To better observe the distribution of the other samples, Figure 4B shows an extension of the PCA scores plot without including walnut samples. A trend along the PC1 can be seen that slightly groups the samples according to their nut type. Moreover, PC1 could be related to the total polyphenolic content as it seems to be a decrease in the level of extracted compounds in samples from the left to right part of the scores plot. For instance, sunflower seeds, showing the highest polyphenol content in dry mass (1400 ± 90 mg GAE/100 g) after walnut samples [26], are the group of samples located more to the left of the plot. Then, being distributed consecutively to the right of the plot following PC1, hazelnuts (550 ± 130 mg GAE/100 g dry mass) and pistachios (642 ± 5 0 mg/100 g dry mass), peanuts

(460 ± 90 mg GAE/100 g dry mass), pumpkin seeds (140 ± 20 mg/100 g dry mass) [26], cashew nuts (133 mg GAE/100 g peeled dry mass) [27], macadamia nuts, almonds (58–159 mg GAE/100 g dry mass) [28,29], and pinions, can be found. Thus, even though there is not a clear separation between groups, those of higher polyphenolic content are distinguished from lower ones.

3.3.2. Classification of Samples According to Nut Type: Supervised PLS-DA Study

HPLC-UV chromatographic fingerprints were also used as chemical descriptors to address nut classification by using a supervised PLS-DA method, obtaining the plot of scores of LV1 vs. LV2 depicted in Figure 5A. Similar to the reported results by PCA, the observed discrimination between samples could be associated to the total amount of extracted compounds, even though in this case both LV1 and LV2 are contributing to it. In fact, PLS-DA maximized the classification of the two nuts with the highest polyphenolic content (walnuts and sunflower seeds), whereas the other samples are more concentrated in the center of the graph. To focus on this distribution, in Figure 5B, walnut and sunflower seed samples are excluded. Again, some samples, such as hazelnuts, pistachios, and peanuts, in which a higher number of extracted phytochemicals is expected according to the literature, can be distinguished from the other samples mainly due to LV1.

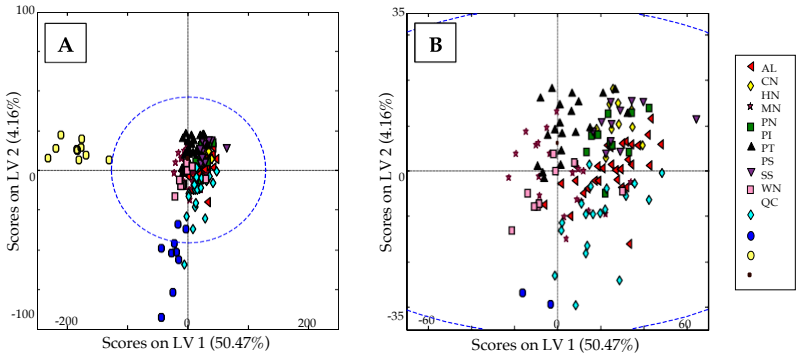


Figure 5. (**A**) Partial least squares-discriminant analysis (PLS-DA) Score plot of LV1 vs. LV2 when using non-targeted HPLC-UV chromatographic fingerprints registered at 280 nm as chemical descriptors for all the analyzed samples. (**B**) Same PLS-DA score plot without including walnuts and sunflower seeds samples.

The fact that the analyzed nut samples are not distributed randomly but more or less grouped according to the type of nut fruit, independently of the format of presentation, shows that non-targeted HPLC-UV fingerprints can be employed as adequate chemical descriptors to address nut sample classification and authentication.

As previously commented in the introduction section, the adulteration of high quality or expensive nut products by substituting them with a lower quality or cheaper nut is a common practice nowadays. For this reason, in the present work, PLS-DA models were built to study some nuts in pairs, i.e., almonds vs. hazelnuts, almonds vs. peanuts or pumpkin seeds vs. sunflower seeds. As can be seen in Figure S2 (supplementary material), the number of latent variables employed to generate each PLS-DA model was selected depending on the cross-validation classification error average, being approximately the first minimum point the most appropriate one. As a good classification was obtained for the studied pairs, the models were validated by using a 70% of each group of samples as the calibration set, while the remaining 30% of samples constituted the validation set. Figure 6 shows the obtained PLS-DA score plots projected on LV1 vs. LV2 as well as Samples vs. Y predicted 1 for (A) almonds vs. hazelnuts, (B) almonds vs. peanuts, and (C) pumpkin seeds vs. sunflower seeds, obtaining a perfect classification and discrimination between these nut samples and reaching a

Sensors **2019**, *19*, 1388

prediction rate of 100% in each case. Therefore, the proposed strategy based on the use of non-targeted HPLC-UV chromatographic fingerprints registered at 280 nm is a very promising method to achieve the characterization, classification, and authentication of nut samples, as well as to address the future identification of some nut frauds by means of adulteration with cheaper nut products.

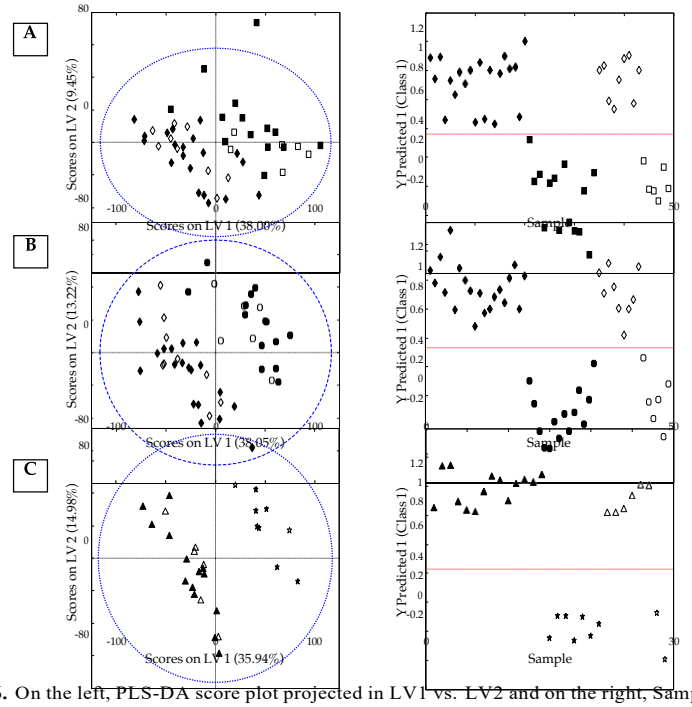


Figure 6. On the left, PLS-DA score plot projected in LV1 vs. LV2 and on the right, Sample vs. Y predicted 1 for (**A**) almonds ◆ vs. hazelnuts ■, (**B**) almonds ◆ vs. peanuts • and (**C**) pumpkin seeds ‗ vs. sunflower seeds ✶. Filled and empty symbols correspond to calibration and validation sets, respectively.

3.3.3. Classification of Samples According to Their Processing Thermal Treatment: Supervised PLS-DA Study

The applicability of non-targeted HPLC-UV chromatographic fingerprints as chemical descriptors to achieve nut sample classification regarding other nut food features, such as the nut format presentation according to the thermal processing treatment (natural and thermally processed, fried or toasted), was also evaluated. For that purpose, the chromatographic fingerprints of those nuts with different types of presentation (see Table 1) were used to create the data matrices, which were later subjected to supervised PLS-DA study. Thus, PLS-DA models were built for almonds, hazelnuts, peanuts, and pumpkin seeds following the same criterion for the number of latent variables selection

Sensors **2019**, *19*, 1388

as established in previous PLS-DA models, as can be seen in Figure S3 (supplementary material). A very acceptable discrimination of the analyzed samples according to the thermal processing method was achieved. Hence, 30% of the samples for each group were removed from the model and used as a validation set. As can be seen in Figure 7, where the PLS-DA score plots projected on LV1 vs. LV2, as well as Samples vs. Y predicted 1 for (A) almonds, (B) hazelnuts, (C) peanuts, and (D) pumpkin seeds, are represented, while for almonds the model showed a 78% classification rate between natural and thermally processed samples (toasted and fried), the other studied nut samples presented a value equal to 100%.

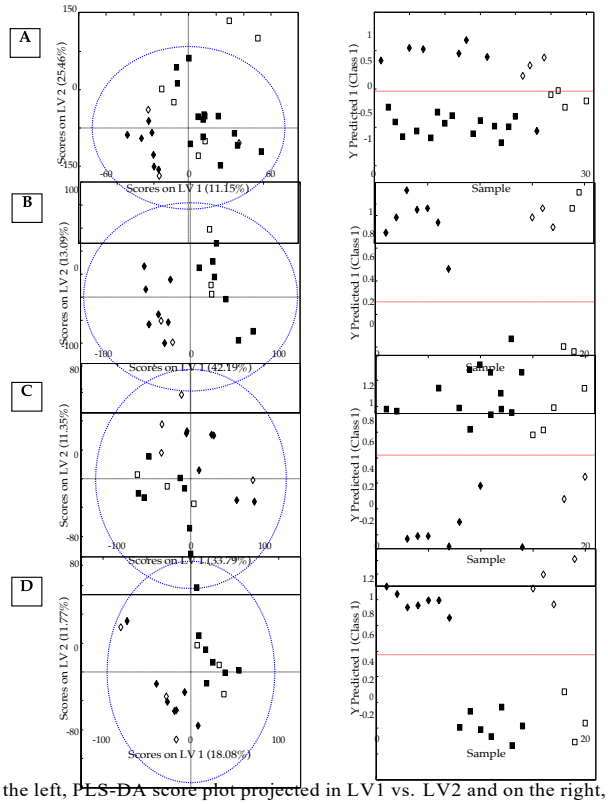


Figure 7. On the left, PLS-DA score plot projected in LV1 vs. LV2 and on the right, Sample vs. Y predicted 1 for (**A**) natural ◆ vs. thermally treated (fried or toasted) almonds ■, (**B**) natural ◆ vs. toasted hazelnuts ■, (**C**) fried ♦ vs. toasted peanuts ■, and (**D**) toasted ■ vs. natural pumpkin seeds ♦. Filled and empty symbols correspond to calibration and validation sets, respectively.

4. Conclusions

In this work, non-targeted HPLC-UV chromatographic fingerprints recorded at 280 nm have proved to be a useful and dependable tool for the classification and authentication of nuts, according to their nut type as well as their thermal treatment, when combined with chemometrics. In fact, the built PLS-DA models for the distinction of a determinate type of nut in front of another have reached a classification rate equal to 100%, independently of their thermal treatment. Moreover, supervised models have also allowed a discrimination capacity over 78% regarding the thermal processing treatment in each nut type. Therefore, this strategy could be proposed to detect frauds involving any of the nut samples studied.

Supplementary Materials: The following are available online at http://www.mdpi.com/1424-8220/19/6/1388/s1, Figure S1: PCA Score plot of PC1 vs. PC2 employing non-targeted HPLC-UV chromatographic fingerprints registered at 280 nm, without QC correction, for all the nut samples and QCs analyzed, Figure S2: Latent variable number vs. CV classification error average plots for the built PLS-DA models of: (A) almonds vs. hazelnuts, (B) almonds vs. peanuts, and (C) pumpkin seeds vs. sunflower seeds, Figure S3: Latent variable number vs. CV classification error average plots for the built PLS-DA models of (A) almonds, (B) hazelnuts, (C) peanuts, and (D) pumpkin seeds.

Author Contributions: G.C., J.S. and O.N. conceived and designed the experiments. G.C., G.J.N. and N.N. performed the experiments and processed the data. G.C., J.S., L.P. and O.N. discussed the first results and suggested additional experiments. The first draft of the manuscript was prepared by G.C. and O.N. and it was revised and substantially improved by L.P. and J.S.

Funding: This research was supported by the Agency for Administration of University and Research Grants (Generalitat de Catalunya, Spain) under the projects 2017SGR-171 and 2017SGR-310.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. European Union Regulation (EC) No 178/2002 of the European Parliament and of the Council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food saf. *Off. J. Eur. Communities* **2004**, *L269*, 1–15.
2. Valdés, A.; Beltrán, A.; Mellinas, C.; Jiménez, A.; Garrigós, M.C. Analytical methods combined with multivariate analysis for authentication of animal and vegetable food products with high fat content. *Trends Food Sci. Technol.* **2018**, *77*, 120–130. [CrossRef]
3. Feldman, E.B. The Scientific Evidence for a Beneficial Health Relationship Between Walnuts and Coronary Heart Disease. *J. Nutr.* **2002**, *132*, 1062S–1101S. [CrossRef]
4. Hu, F.B.; Stampfer, M.J.; Manson, J.E.; Rimm, E.B.; Colditz, G.A.; Rosner, B.A.; Speizer, F.E.; Hennekens, C.H.; Willett, W.C. Disease in Women: Prospective Cohort Study. *Br. Med. J.* **1998**, *317*, 1341–1345. [CrossRef]
5. Herbello-Hermelo, P.; Lamas, J.P.; Lores, M.; Domínguez-González, R.; Bermejo-Barrera, P.; Moreda-Piñeiro, A. Polyphenol bioavailability in nuts and seeds by an in vitro dialyzability approach. *Food Chem.* **2018**, *254*, 20–25. [CrossRef]
6. Li, A.N.; Li, S.; Zhang, Y.J.; Xu, X.R.; Chen, Y.M.; Li, H. Bin Resources and biological activities of natural polyphenols. *Nutrients* **2014**, *6*, 6020–6047. [CrossRef] [PubMed]
7. Bullo, M.; Lamuela-Raventós, R.M.; Salas-Salvado, J. Mediterranean Diet and Oxidation: Nuts and Olive Oil as Important Sources of Fat and Antioxidants. *Curr. Top. Med. Chem.* **2011**, *11*, 1–14. [CrossRef]
8. Torabian, S.; Haddad, E.; Rajaram, S.; Banta, J.; Sabaté, J. Acute effect of nut consumption on plasma total polyphenols, antioxidant capacity and lipid peroxidation. *J. Hum. Nutr. Diet.* **2009**, *22*, 64–71. [CrossRef]
9. Godos, J.; Marventano, S.; Mistretta, A.; Galvano, F.; Grosso, G. Dietary sources of polyphenols in the Mediterranean healthy Eating, Aging and Lifestyle (MEAL) study cohort. *Int. J. Food Sci. Nutr.* **2017**, *68*, 750–756. [CrossRef]
10. Esteki, M.; Vander Heyden, Y.; Farajmand, B.; Kolahderazi, Y. Qualitative and quantitative analysis of peanut adulteration in almond powder samples using multi-elemental fingerprinting combined with multivariate data analysis methods. *Food Control* **2017**, *82*, 31–41. [CrossRef]

11. Yildirim, A.N.; Akinci-Yildirim, F.; Şan, B.; Sesli, Y. Total Oil Content and Fatty Acid Profile of some Almond (*Amygdalus Communis* L.) Cultivars. *Polish J. Food Nutr. Sci.* **2016**, *66*, 173–178. [CrossRef]
12. Esteki, M.; Farajmand, B.; Amanifar, S.; Barkhordari, R.; Ahadiyan, Z.; Dashtaki, E.; Mohammadlou, M.; Vander Heyden, Y. Classification and authentication of Iranian walnuts according to their geographical origin based on gas chromatographic fatty acid fingerprint analysis using pattern recognition methods. *Chemom. Intell. Lab. Syst.* **2017**, *171*, 251–258. [CrossRef]
13. Čolić, S.D.; Fotirić Akšić, M.M.; Lazarević, K.B.; Zec, G.N.; Gašić, U.M.; Dabić Zagorac, D.; Natić, M.M. Fatty acid and phenolic profiles of almond grown in Serbia. *Food Chem.* **2017**, *234*, 455–463. [CrossRef] [PubMed]
14. Esteki, M.; Farajmand, B.; Kolahderazi, Y.; Simal-Gandara, J. Chromatographic Fingerprinting with Multivariate Data Analysis for Detection and Quantification of Apricot Kernel in Almond Powder. *Food Anal. Methods* **2017**, *10*, 3312–3320. [CrossRef]
15. Moscetti, R.; Radicetti, E.; Monarca, D.; Cecchini, M.; Massantini, R. Near infrared spectroscopy is suitable for the classification of hazelnuts according to Protected Designation of Origin. *J. Sci. Food Agric.* **2015**, *95*, 2619–2625. [CrossRef] [PubMed]
16. Ghosh, S.; Mishra, P.; Mohamad, S.N.H.; de Santos, R.M.; Iglesias, B.D.; Elorza, P.B. Discrimination of peanuts from bulk cereals and nuts by near infrared reflectance spectroscopy. *Biosyst. Eng.* **2016**, *151*, 178–186. [CrossRef]
17. Cuadros-Rodríguez, L.; Ruiz-Samblás, C.; Valverde-Som, L.; Pérez-Castaño, E.; González-Casado, A. Chromatographic fingerprinting: An innovative approach for food "identitation" and food authentication - A tutorial. *Anal. Chim. Acta* **2016**, *909*, 9–23. [CrossRef] [PubMed]
18. Sobolev, A.P.; Circi, S.; Capitani, D.; Ingallina, C.; Mannina, L. Molecular fingerprinting of food authenticity. *Curr. Opin. Food Sci.* **2017**, *16*, 59–66. [CrossRef]
19. Zhang, J.; Zhang, X.; Dediu, L.; Victor, C. Review of the current application of fingerprinting allowing detection of food adulteration and fraud in China. *Food Control* **2011**, *22*, 1126–1135. [CrossRef]
20. Esteki, M.; Shahsavari, Z.; Simal-Gandara, J. Use of spectroscopic methods in combination with linear discriminant analysis for authentication of food products. *Food Control* **2018**, *91*, 100–112. [CrossRef]
21. Riedl, J.; Esslinger, S.; Fauhl-Hassek, C. Review of validation and reporting of non-targeted fingerprinting approaches for food authentication. *Anal. Chim. Acta* **2015**, *885*, 17–32. [CrossRef] [PubMed]
22. Eigenvector Research Incorporated. Powerful Resources for Intelligent Data Analysis. Available online: http://www.eigenvector.com/software/solo.htm (accessed on 15 January 2019).
23. Massart, D.L.; Vandeginste, B.G.M.; Buydens, L.M.C.; de Jong, S.; Lewi, P.J.; Smeyers-Verbeke, J. *Handbook of Chemometrics and Qualimetrics*; Elsevier: Amsterdam, The Netherlands, 1997.
24. Pérez-Ráfols, C.; Saurina, J. Liquid chromatographic fingerprints and profiles of polyphenolic compounds applied to the chemometric characterization and classification of beers. *Anal. Methods* **2015**, *7*, 733–8739. [CrossRef]
25. Larrauri, A.; Núñez, O.; Hernández-Cassou, S.; Saurina, J. Determination of Polyphenols in White Wines by Liquid Chromatography: Application to the Characterization of Alella (Catalonia, Spain) Wines Using Chemometric Methods. *J. AOAC Int.* **2017**, *100*, 323–329. [CrossRef] [PubMed]
26. Zujko, M.E.; Witkowska, A.M. Antioxidant potential and polyphenol content of beverages, chocolates, nuts, and seeds. *Int. J. Food Prop.* **2014**, *17*, 86–92. [CrossRef]
27. Soares, D.J.; Silva, J.; Carmo, D.O.; Da, J.; Severino, S.; Maia, G.A.; Henrique, P.; Sousa, M.D.E. Polyphenols and Antioxidant Activity of Cashew Nuts from Conventional and Organic Cultivation in Different Stages of Processing. *Boletim Centro de Pesquisa de Processamento de Alimentos* **2013**, *31*, 173–181. [CrossRef]
28. Bolling, B.W. Almond Polyphenols: Methods of Analysis, Contribution to Food Quality, and Health Promotion. *Compr. Rev. Food Sci. Food Saf.* **2017**, *16*, 346–368. [CrossRef]
29. Bolling, B.W.; Dolnikowski, G.; Blumberg, J.B.; Chen, C.Y.O. Polyphenol content and antioxidant activity of California almonds depend on cultivar and harvest year. *Food Chem.* **2010**, *122*, 819–825. [CrossRef]

Supplementary Material

Non-targeted HPLC-UV Fingerprinting as Chemical Descriptors for the Classification and Authentication of Nuts by Multivariate Chemometric Methods

Guillem Campmajó [1], Gemma J. Navarro [1], Nerea Núñez [1], Lluís Puignou [1,2], Javier Saurina [1,2] and Oscar Núñez [1,2,3,*]

[1] Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, Martí i Franqués, 1-11, E08028 Barcelona, Spain; campma03@gmail.com (G.C.); g.navarro.albiol@gmail.com (G.J.N.); nereant7@gmail.com (N.N.), lluis.puignou@ub.edu (L.P.), xavi.saurina@ub.edu (J.S.)

[2] Research Institute in Food Nutrition and Food Safety, University of Barcelona, Recinte Torribera, Av. Prat de la Riba 171, Edifici de Recerca (Gaudí), Santa Coloma de Gramenet, E08921 Barcelona, Spain

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

* Correspondence: oscar.nunez@ub.edu; Tel.: +34-93-403-3706

Fig. S1. PCA Score plot of PC1 vs PC2 employing non-targeted HPLC-UV chromatographic fingerprints registered at 280 nm, without QC correction, for all the nut samples and QCs analyzed.

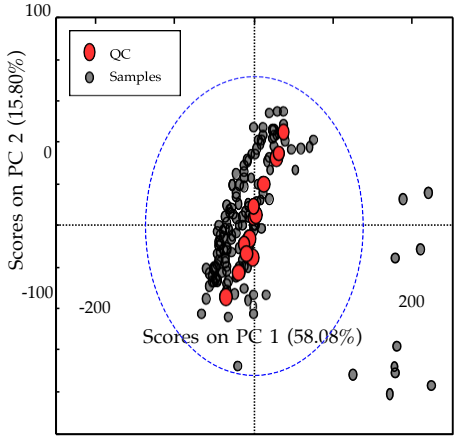


Fig. S2. Latent variable number *vs* CV classification error average plots for the built PLS-DA models of: (a) almonds *vs* hazelnuts, (b) almonds *vs* peanuts and (c) pumpkin seeds *vs* sunflower seeds.

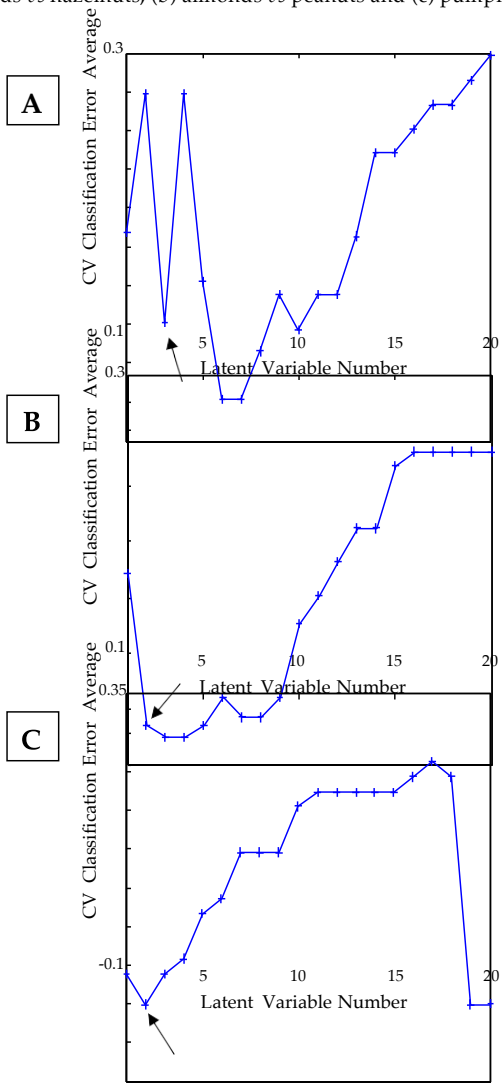
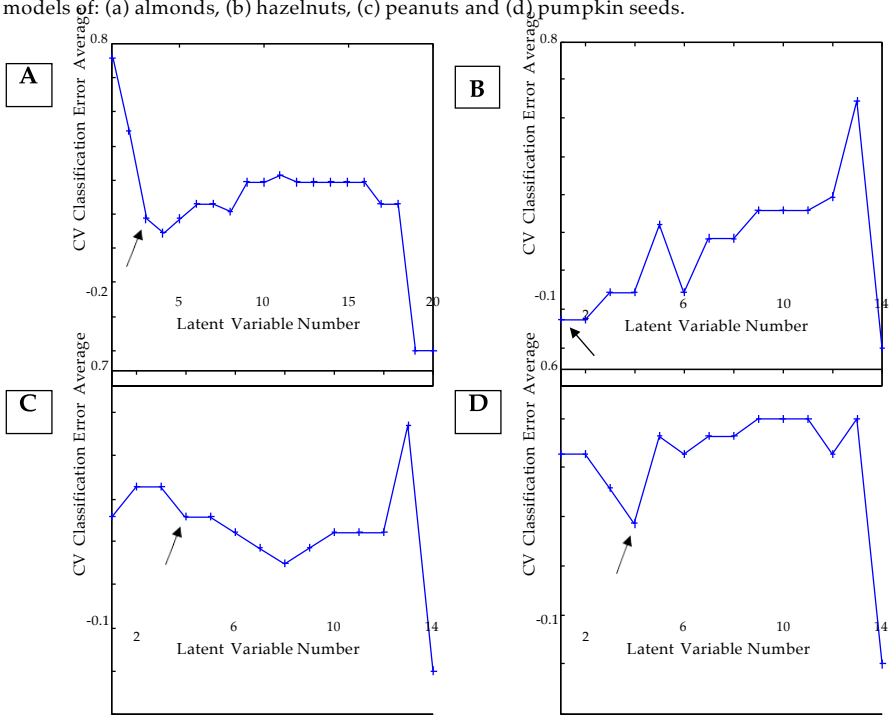


Fig. S3. Latent variable number *vs* CV classification error average plots for the built PLS-DA models of: (a) almonds, (b) hazelnuts, (c) peanuts and (d) pumpkin seeds.



3.2.2. PUBLICATION V

Classification of hen eggs by HPLC-UV fingerprinting and chemometric methods.

Campmajó, G.; Cayero, L.; Saurina, J.; Núñez, O.

Foods. 2019, 8, 310.

foods

MDPI

Article

Classification of Hen Eggs by HPLC-UV Fingerprinting and Chemometric Methods

Guillem Campmajó [1,*], Laura Cayero [1], Javier Saurina [1,2] and Oscar Núñez [1,2,3]

[1] Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, Martí i Franquès 1-11, E08028 Barcelona, Spain
[2] Research Institute in Food Nutrition and Food Safety, University of Barcelona, Recinte Torribera, Av. Prat de la Riba 171, Edifici de Recerca (Gaudí), Santa Coloma de Gramenet, E08921 Barcelona, Spain
[3] Serra Húnter Fellow, Generalitat de Catalunya, Rambla de Catalunya 19-21, E08007 Barcelona, Spain
* Correspondence: campma03@gmail.com; Tel.: +34-93-403-3706

Received: 21 June 2019; Accepted: 29 July 2019; Published: 1 August 2019

Abstract: Hen eggs are classified into four groups according to their production method: Organic, free-range, barn, or caged. It is known that a fraudulent practice is the misrepresentation of a high-quality egg with a lower one. In this work, high-performance liquid chromatography with ultraviolet detection (HPLC-UV) fingerprints were proposed as a source of potential chemical descriptors to achieve the classification of hen eggs according to their labelled type. A reversed-phase separation was optimized to obtain discriminant enough chromatographic fingerprints, which were subsequently processed by means of principal component analysis (PCA) and partial least squares-discriminant analysis (PLS-DA). Particular trends were observed for organic and caged hen eggs by PCA and, as expected, these groupings were improved by PLS-DA. The applicability of the method to distinguish egg manufacturer and size was also studied by PLS-DA, observing variations in the HPLC-UV fingerprints in both cases. Moreover, the classification of higher class eggs, in front of any other with one lower, and hence cheaper, was studied by building paired PLS-DA models, reaching a classification rate of at least 82.6% (100% for organic vs. non-organic hen eggs) and demonstrating the suitability of the proposed method.

Keywords: HPLC-UV; fingerprinting; food classification; hen eggs; principal component analysis; partial least square-discriminant analysis

1. Introduction

In the last years, the interest of society in the food they purchase and consume has been raised. In this line, products with value-added due to specific particularities such as organic production, protected designation of origin (PDO), protected geographical indication (PGI), or those with fair-trade certification, are now receiving special attention. These labels not only ensure and guarantee food quality and traceability, but also mean an increment in its price in comparison with conventional products.

Hen eggs are among the most commonly eaten foods worldwide, as they have a high nutritional value, cheap costs, and are widely employed in international cuisines. They consist of two parts: The egg white, which mainly consists of 85% water and 10% proteins (ovalbumin being the most abundant one) approximately, and the egg yolk, which is composed of almost 22% lipids [1,2]. Moreover, their intake provides all the essential amino acids, many vitamins (vitamin A, riboflavin, choline, vitamin B_{12}, and vitamin B_9), and minerals (phosphorus, potassium, iron, and zinc).

In Europe, where almost 8 million tons of hen eggs were produced in 2017 [3], rules on their trade regarding production, hygiene, labelling, and marketing are laid down by the European Union (EU) [4–6]. Thereby, according to the European labelling eggs rule, each A quality category egg,

which are those destined for human consumption, has to contain an identifier number code in its shell. Among other information that can be found on it, such as the country of origin (two-letter ISO - International Organization for Standardization- abbreviation code), the province, the municipality, and the producer establishment, the kind of hens and the breeding method employed are indicated by the first number digit:

Digit 0 is related to organic eggs (O), which means that they come from authorized and certified organic production farms. Thus, hens are fed with grown pasture and organic farming products, without employing transgenic substances nor antibiotics. The animals have a minimum space of 4 and 6 hens/m^2 outdoors and indoors, respectively. These are the most expensive eggs.

Digit 1 corresponds to free-range hen eggs (FR). In this case, their diet is mainly based on prepared cereal pellets, although grass can also be eaten. Antibiotics are mixed with food if needed. Moreover, similar space conditions to organic eggs are established.

Digit 2 indicates barn hen eggs (B). Hens do not have outdoor access, as they live in densely populated vessels and therefore, their diet consists of the prepared pellets and there is no entrance of natural light. Further, antibiotics are systematically provided with feed.

Digit 3 for eggs from caged hens (C), which are the cheapest ones. In these cages, hens can barely move (the minimum space allowed is of 12 hens/m^2) and there is no access to natural light either. Medical additives are provided with feed.

Due to the huge amount of produced eggs, two different frauds can be practiced. On one hand, in accordance with the European legislation, hen eggs have to reach the consumers within the 21 days of being laid [7], and their expiration date has to be fixed not more than 28 days after laying [6]. As there is no way to confirm whether those that are for sale are within the stipulated periods, some producers label them with erroneous dates, therefore giving a longer time before reaching their expiration date [8]. On the other hand, it is also difficult to distinguish hen eggs regarding their type. Although organic bodies may ensure the compliance of the established regulations, due to the high cost of the evaluation systems, some producers and distributors regulate themselves without adopting any national certification standard, leaving then the opportunity for food fraud [9].

The egg price increase from category 3 to 0 makes them susceptible to fraud, since a low category egg could be labelled as a superior one. Several methodologies have been previously developed in order to address egg authentication. For instance, profiling fatty acid composition by gas chromatography (GC), fitted with flame ionization detector (FID), in combination with chemometric techniques was proposed for the verification of organic against conventional eggs [1,10]. However, relatively time-consuming methodologies are usually required in order to determine the total lipid and fatty acid composition from the samples, also involving derivatization steps before GC separation. In another study, the carotenoid profile acquired by high-performance liquid chromatography and ultraviolet detection (HPLC-UV) was performed to classify both organic and conventional eggs [11]. Besides, in some cases, the authenticity of organic eggs and the assurance on their origin, was also approached by evaluating the level of several elements, including rare earth elements [12–14].

As can be seen, most of the methods described in the literature for egg authentication are based on targeted profiling approaches, which are focused on the specific determination of a given group of known selected chemicals. However, up to now, no specific biomarkers have been found in order to address hen eggs classification regarding their labelled class. Since many factors will affect the chemical composition of these products, non-targeted fingerprinting strategies that involve the determination of non-selective signals related to a range of potential discriminating compounds (i.e., spectrum or chromatogram), are promising approaches to address food authenticity issues [15–19]. As an example, a spectroscopic technique such as near infrared (NIR), in combination with principal component analysis (PCA), was proposed to achieve the classification of different type of eggs found in Chinese markets [20]. In the present work, HPLC-UV fingerprints recorded at 250 nm were proposed as a source of discriminant signals for hen eggs classification according to their production method by PCA and partial least squares-discriminant analysis (PLS-DA).

2. Materials and Methods

2.1. Chemicals and Standard Solutions

All the employed chemicals were of analytical grade. In the sample treatment, the acetonitrile and water (LC-MS Chromasolv® quality) used were purchased from Sigma-Aldrich (St. Louis, MO, USA). Methanol (UHPLC-gradient grade) was obtained from PanReac AppliChem (Barcelona, Spain) and formic acid (≥98%) from Sigma-Aldrich. For the mobile phase, water was purified using an Elix 3 coupled to a Milli-Q system from Millipore Corporation (Burlington, MA, USA) and filtered through a 0.22 µm nylon membrane integrated into the Milli-Q system.

2.2. Instrumentation

An Agilent 1100 Series HPLC instrument equipped with a quaternary pump (G1311A), a degasser (G1322A), an autosampler (G1329A), a diode array detector (G1315B), and a PC with the Agilent Chemstation software, all of them from Agilent Technologies (Waldbronn, Germany), was employed. HPLC-UV fingerprints were obtained by reversed-phase mode using a Kinetex C18 porous-shell column (100 mm × 4.6 mm I.D., 2.6 µm particle size) from Phenomenex (Torrance, CA, USA) at room temperature. Chromatographic separation was performed under gradient elution mode, using 0.1% (*v*/*v*) formic acid aqueous solution (solvent A) and methanol (solvent B) as mobile phase components, following the next elution program: 0–20 min, linear gradient from 15% to 95% solvent B; 20–30 min, isocratic elution at 95% solvent B; 30–30.1 min, back to initial conditions; and from 30.1–35 min, at 15% solvent B for column re-equilibration. The mobile phase flow rate was 0.4 mL/min, and the injection volume was 5 µL. The HPLC-UV fingerprints were registered at 250 nm.

2.3. Samples and Sample Treatment

Characterization and classification studies were carried out by analyzing 173 hen egg samples purchased from local markets (Barcelona, Spain). Table 1 classifies them according to their typology and manufacturer and defines their specified size as well as the number of samples.

Table 1. Description of the egg samples analyzed.

Egg Type	Manufacturer	Egg Size	Number of Samples
Organic hen eggs (O)	ViuBi	M/L	23
Free-range hen eggs (FR)	Vall de Mestral	-	23
	Ous Roig (Ebre)	-	23
	Ous Roig	L/XL	22
Barn hen eggs (B)	Liderou	M	24
	Eroski	L	24
	Ous Roig	L	11
Caged hen eggs (C)	Eroski	M	12
	Eroski	L	11

Sample extraction was performed following a previously described method [21], with some modifications. Briefly, 0.3 g of homogenized egg sample were weighed in an Eppendorf tube (Deltalab, Rubí, Spain), mixed with 1 mL of an acetonitrile:water 80:20 (*v*/*v*) solution by stirring in a Vortex (Stuart, Stone, United Kingdom) for 30 s, and then, centrifuged (Allegra™ 64R Centrifugue, Beckman Coulter, L'Hospitalet de Llobregat, Spain) for 10 min at 14,000 rpm and 4 °C. The supernatant extract was then filtered through 0.22 µm filter (Scharlab, Sentmenat, Spain) and stored at −18 °C in 2 mL glass injection vials until HPLC-UV analysis.

Moreover, a quality control (QC), which aimed to evaluate the repeatability of the method and the robustness of the chemometric results, was prepared by mixing 50 µL of each sample extract. A QC and a blank of acetonitrile were injected every 10 randomly injected samples.

2.4. Data Analysis

PCA and PLS-DA calculations were made by using SOLO 8.6 chemometric software (Eigenvector Research [22], Manson, WA, USA). Theoretical background of these methods in a detailed way is addressed elsewhere [23].

X-data matrices in both PCA and PLS-DA analysis consisted of the HPLC-UV chromatographic fingerprints obtained at 250 nm (absorbance intensities vs. retention time), whereas the Y-data matrix in PLS-DA defined each sample class. In order to improve the data quality, HPLC-UV chromatograms were smoothed, baseline-corrected, aligned, and autoscaled. Scatter plots of scores from principal components (PCs), in PCA, and latent variables (LVs), in PLS-DA, were used to study the distribution of samples, revealing patterns that could be correlated to their characteristics. In order to build both PCA and PLS-DA models, the first significant minimum point of the cross validation (CV) error from a Venetian blind approach was considered to be the most appropriate number of PCs or LVs, respectively.

3. Results and Discussion

3.1. HPLC-UV Chromatographic Separation

This work aimed to develop a HPLC-UV fingerprinting approach for the classification and discrimination of hen eggs according to their labelled typology. Thus, in order to obtain the richest chromatographic fingerprints, after a slightly modified simple liquid–liquid extraction procedure [19], the obtained extract of a B egg sample was employed for the optimization of the chromatographic separation by reversed-phase mode using 0.1% aqueous formic acid and methanol as mobile phase components. In a first separation consisting of a universal gradient, where methanol increased from 10% to 90% in 30 min, several compounds with different peak intensities were detected, although most of them elute close to the column dead volume. Thereby, different initial methanol percentages as well as the combination of gradient and isocratic steps were tested. As a compromise between the number of detected peaks, resolution, and analysis time, a final gradient consisting of an increase of methanol from 15% to 95% in 20 min followed by an isocratic step at 95% methanol for 10 min was selected. Figure 1 shows the obtained HPLC-UV chromatographic fingerprint registered at 250 nm for a B egg sample with the proposed gradient program.

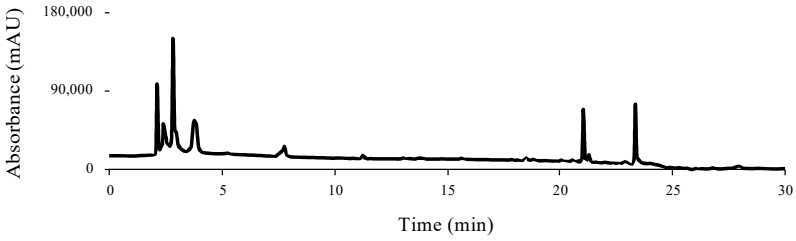


Figure 1. High performance liquid chromatography with ultraviolet detection (HPLC-UV) chromatogram at 250 nm obtained for a selected barn hen egg sample under the proposed gradient elution program (Section 2.2).

3.2. HPLC-UV Fingerprints

A total of 173 egg samples were analyzed by the proposed HPLC-UV method for classification purposes. For instance, Figure 2 shows the chromatograms at 250 nm for each one of the egg sample

groups (O, FR, B, and C) analyzed. At a first glance, similar chromatographic fingerprints were obtained independently of the egg type. In fact, according to the retention times, the detected compounds seemed to be the same in each of them. However, variations associated to peak intensities, as well as their abundance within the different peak signals detected in a same sample, can be easily remarked. Therefore, these chromatographic fingerprints were evaluated and proposed as chemical descriptors to achieve sample classification.

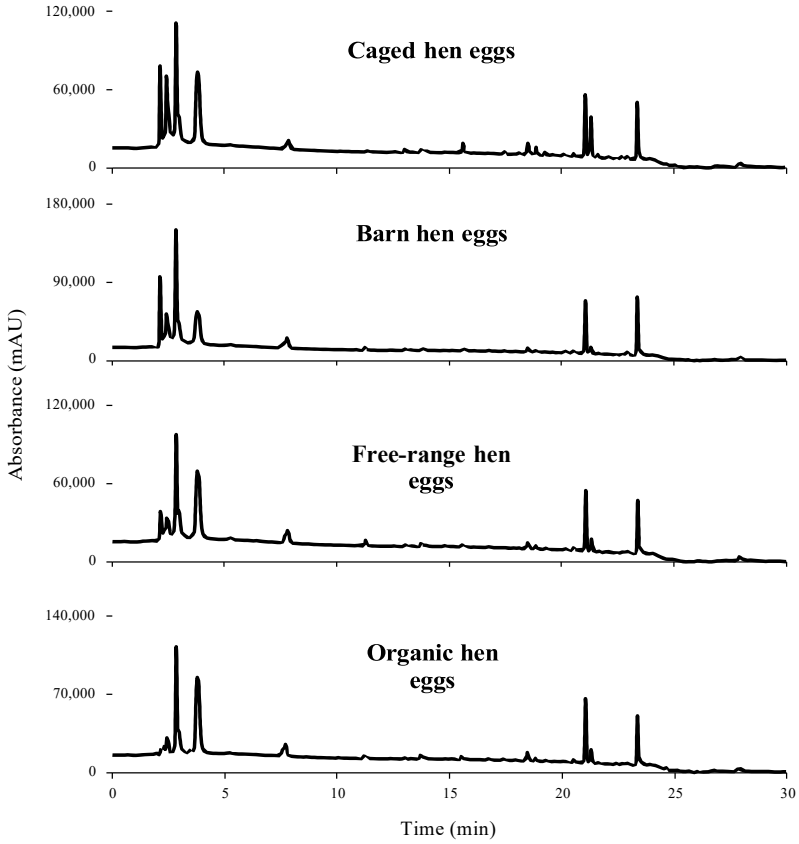


Figure 2. High performance liquid chromatography with ultraviolet detection (HPLC-UV) chromatographic fingerprints acquired at 250 nm for a selected sample within each egg type.

3.3. Classification of Samples According to Egg Type: PCA Study

As a first approach, a non-supervised PCA study was performed to evaluate the usefulness of HPLC-UV fingerprints for eggs classification according to their type. For that purpose, a data matrix (189 × 4506, samples × variables), which consisted of the recorded absorbance signals at 250 nm as a function of time for the analyzed egg samples and the QCs, was built. Moreover, data were pretreated as mentioned in Section 2.4, not only to reduce noise interferences, peak shifting, and baseline drifts, but also to provide the same weight to each variable by suppressing differences in their magnitude and amplitude scales. As a first result, the plot of scores of PC1 vs. PC2 (seven PCs were chosen for the PCA analysis), which is displayed in Figure 3A, shows that QC samples form a clear group (without any trend associated to a systematic error) in the upper side of the diagram, allowing the

consideration of the obtained chemometric results. As can be seen in the plot of scores PC1 vs. PC3 shown in Figure 3B (seven PCs were also chosen for the supervised analysis), when excluding QC samples, even though there is not an evident discrimination among the samples, both the highest (O eggs) and lowest (C eggs) quality eggs predominate above and on the left of the plot, respectively. Up to this point, the proposed HPLC-UV fingerprints appeared to be adequate chemical descriptors at least for the distinction of O and C eggs, although PCA is only a non-supervised exploratory chemometric method. Therefore, in order to better exploit the obtained data and to improve the results on sample distribution, a supervised PLS-DA chemometric classification method was evaluated.

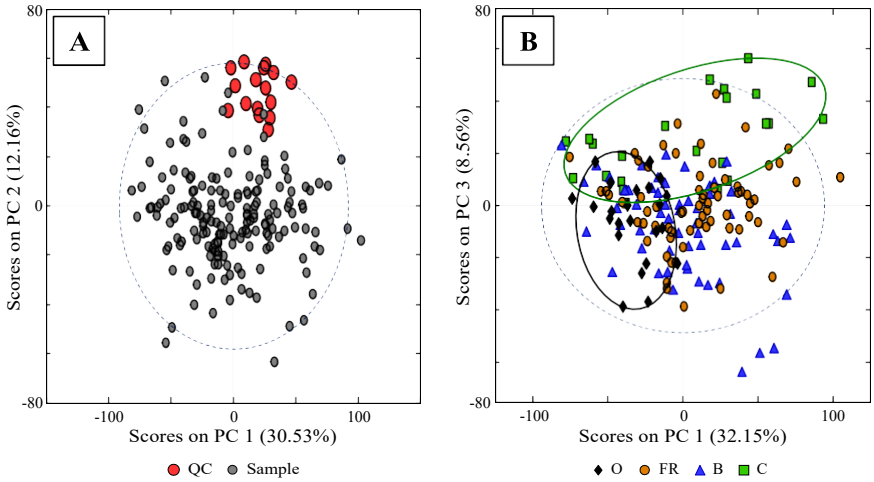


Figure 3. (**A**) Scores plot of PC1 vs. PC2 when using HPLC-UV fingerprints registered at 250 nm as chemical descriptors, showing a correct behavior of quality control (QC) samples. (**B**) Scores plot of PC1 vs. PC3, without including QC samples, showing a slight trend of organic hen (O) and caged hen (C) eggs.

3.4. Classification of Samples According to Egg Type: PLS-DA Study

The supervised chemometric study of all the analyzed egg samples was carried out by PLS-DA. For this reason, in addition to the X-data matrix previously described in the PCA study, a Y-matrix indicating the membership of each sample (O, FR, B, and C eggs) was used. The obtained scores plot of LV1 vs. LV2 (six LVs were chosen as optimum for the PLS-DA model, as detailed in Section 2.4), which is shown in Figure 4, improves non-supervised chemometric results as expected, and the obtained distribution seem to be directly related to the hens breeding method employed. In fact, eggs of hens with organic diet (O eggs) follow a particular trend mainly due to LV1, whereas LV2 affects those obtained from hens fed with a cereal-based diet and reared in cages (C eggs). On the other hand, in between these two group samples, both FR and B eggs, which as C eggs, are also collected from hens with a cereal-based diet but with better breeding conditions, apparently appeared randomly distributed.

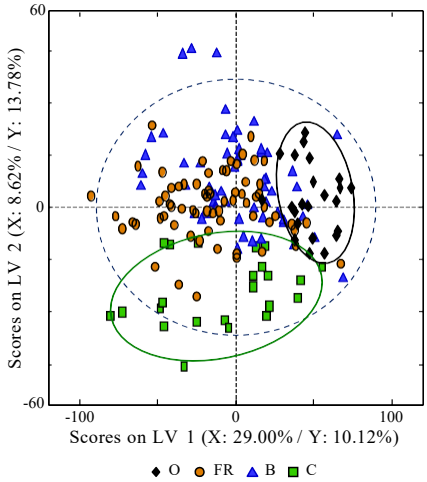


Figure 4. Partial least square-discriminant analysis (PLS-DA) scores plots of LV1 vs. LV2 when using HPLC-UV fingerprints registered at 250 nm as chemical descriptors.

A fact that should be taken into consideration is the number of manufacturers involved within the employed samples for each typology. While for O and C eggs all samples belonged only to one manufacturer, FR and B groups came from three. Thus, although according to the EU, same rules on the breeding process are established for a given quality egg class, additional sources of variance, such as the cereals employed in hens diet or the available grass and plants, could suppose a differential factor. Therefore, the applicability of HPLC-UV fingerprints as chemical descriptors to distinguish between the egg manufacturers was also evaluated by means of PLS-DA. For instance, Figure 5 shows the obtained scores plot of LV1 vs. LV2 when a 4 LVs PLS-DA model was built only for B egg samples. As can be observed, B eggs are clearly grouped according to their manufacturer, and thus, the proposed chromatographic fingerprints seem to be capable to remark these differences between different origins of production.

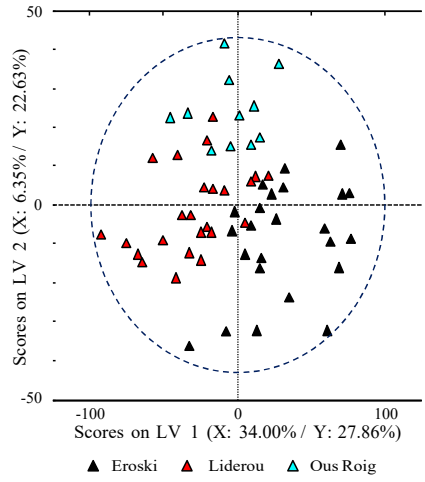


Figure 5. PLS-DA scores plots of LV1 vs. LV2 for B hen egg when using HPLC-UV chromatographic fingerprints registered at 250 nm as chemical descriptors.

Moreover, the size of the studied eggs was also evaluated by this methodology, as reported in Figure S1 (Supplementary Material). For that purpose, a matrix containing B and C egg samples, which were the only available classes labelled by size, was constructed. A clear distinction between M and L size eggs was achieved, independently of their class (B or C), denoting changes in the phytochemical fingerprint related to this morphological characteristic.

3.5. Supervised PLS-DA Method Validation

As the main goal of the present work was the discrimination of hen eggs according to their labelled class, and in order to demonstrate the applicability of the proposed method, the classification of higher class eggs in front of any other with one lower, and hence cheaper, was studied by building paired PLS-DA models (i.e., O vs. FR, B and C eggs; FR vs. B and C eggs; and B vs. C eggs). As can be observed in Figure S2 (Supplementary Material), the number of LVs employed to generate each classificatory model was selected considering the first significant minimum point of the CV error average as the most appropriate one.

For predicting the egg classes, the chemometric model was established using 70% of samples of each group as calibration set, while the remaining 30% was employed as "unknown" set for validation purposes. As can be seen in Figure 6A, O eggs, which are the most expensive ones, were clearly discriminated from those with lower prizes, reaching a classification rate of 100%. Further, while for the PLS-DA model of FR in front of B and C eggs (Figure 6B) a discrimination capacity of 82.6% was accomplished, B in front of C eggs (Figure 6C) resulted to be of 88%.

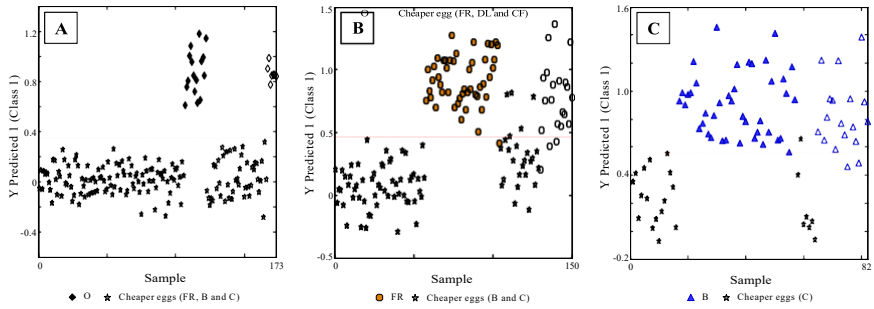


Figure 6. Sample vs. Y predicted 1 Scores plot for (**A**) O vs. FR, B and C eggs, (**B**) FR vs. B and C, and (**C**) B vs. C.

4. Conclusions

In this work, HPLC-UV chromatograms acquired at 250 nm have proved to be useful discriminant fingerprints for the classification of hen eggs according to their labelled typology. The PLS-DA models built for each egg category in front of those with lower one have reached at least a classification rate of 82.6%, showing satisfactory results of prediction. The distinction among organic and non-organic eggs has been especially satisfactory, in which 100% of sensitivity and selectivity has been reached. Moreover, the chromatographic fingerprints have also shown differences in egg phytochemical content among samples with different size independently of their type, as well as different manufacturers between samples from the same class.

Even though HPLC-UV fingerprints provided satisfactory results, the perfect classification of the four labelled hen egg groups was not achieved. At this point and in order to improve them, the evaluation of a new matrix, such as the egg yolk, rather than using the whole egg, could be an improvement to solve this problematic. Besides, fluorescence detection, which is characterized to be more selective than UV detection, could be proposed as an alternative detection technique for better descriptive models.

Finally, compared with biomarker-based strategies, the principal advantage of fingerprinting approaches is that the identification and quantification of selective species of each class are not essential for a successful sample classification. Here, in this regard, despite that specific markers have not been found, subtle differences in the content of components up- or down-expressed among classes have been exploited as the basis of the classification models. However, future work should also be directed towards biomarkers identification in order to address hen eggs authentication.

Supplementary Materials: The following are available online at http://www.mdpi.com/2304-8158/8/8/310/s1, Figure S1: PLS-DA scores plots of LV1 vs. LV2 for M and L size eggs when using HPLC-UV chromatographic fingerprints registered at 250 nm as chemical descriptors. Figure S2: Latent variable number vs. CV classification error average plots for the built PLS-DA models of: (A) O vs. FR, B and C eggs, (B) FR vs. B and C, and (C) B vs. C.

Author Contributions: G.C., J.S. and O.N. conceived and designed the experiments. G.C. and L.C. performed the experiments and processed the data. G.C. carried out the validation studies of all the chemometric methods proposed. G.C., J.S. and O.N. discussed the first results and suggested additional experiments. The first draft of the manuscript was prepared by G.C. and O.N. and it was revised and substantially improved by J.S.

Funding: This research was supported by the Agency for Administration of University and Research Grants (Generalitat de Catalunya, Spain) under the projects 2017SGR-171 and 2017SGR-310. G. Campmajó also thanks the University of Barcelona for the APIF fellowship.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cherian, G.; Holsonbake, T.B.; Goeger, M.P. Fatty acid composition and egg components of specialty eggs. *Poult. Sci.* **2002**, *81*, 30–33. [CrossRef] [PubMed]
2. Fredriksson, S.; Elwinger, K.; Pickova, J. Fatty acid and carotenoid composition of egg yolk as an effect of microalgae addition to feed formula for laying hens. *Food Chem.* **2006**, *99*, 530–537. [CrossRef]
3. European Comission. Agriculture and Rural Development. Available online: https://www.ec.europa.eu/agriculture/eggs_en (accessed on 18 June 2019).
4. European Union. Council Directive 1999/74/EC of 19 July 1999 laying down minimum standards for the protection of laying hens. *Off. J. Eur. Communities* **1999**, *L203*, 53–57.
5. European Union. Commission Directive 2002/4/EC of 30 January 2002 on the registration of establishments keeping laying hens, covered by Council Directive 1999/74/EC. *Off. J. Eur. Communities* **2002**, *L30*, 44–46.
6. European Union. Commission Regulation (EC) No 589/2008 of 23 June 2008 laying down detailed rules for implementing Council Regulation (EC) No 1234/2007 as regards marketing standards for eggs. *Off. J. Eur. Union* **2008**, *L163*, 6–23.
7. The European Parliament and the Council of the European Union. Regulation (EC) No 853/2004 of the European Parliament and of the Council of 29 April 2004 laying down specific hygiene rules for food business operators on the hygiene of foodstuffs. *Off. J. Eur. Union* **2004**, *L139*, 151.
8. Johnson, A.E.; Sidwick, K.L.; Pirgozliev, V.R.; Edge, A.; Thompson, D.F. Metabonomic Profiling of Chicken Eggs during Storage Using High-Performance Liquid Chromatography-Quadrupole Time-of-Flight Mass Spectrometry. *Anal. Chem.* **2018**, *90*, 7489–7494. [CrossRef] [PubMed]
9. Rogers, K.M.; Van Ruth, S.; Alewijn, M.; Philips, A.; Rogers, P. Verification of Egg Farming Systems from the Netherlands and New Zealand Using Stable Isotopes. *J. Agric. Food Chem.* **2015**, *63*, 8372–8380. [CrossRef] [PubMed]
10. Tres, A.; O'Neill, R.; Van Ruth, S.M. Fingerprinting of fatty acid composition for the verification of the identity of organic eggs. *Lipid Technol.* **2011**, *23*, 40–42. [CrossRef]
11. Van Ruth, S.; Alewijn, M.; Rogers, K.; Newton-Smith, E.; Tena, N.; Bollen, M.; Koot, A. Authentication of organic and conventional eggs by carotenoid profiling. *Food Chem.* **2011**, *126*, 1299–1305. [CrossRef]
12. Barbosa, R.M.; Nacano, L.R.; Freitas, R.; Batista, B.L.; Barbosa, F. The Use of Decision Trees and Naïve Bayes Algorithms and Trace Element Patterns for Controlling the Authenticity of Free-Range-Pastured Hens' Eggs. *J. Food Sci.* **2014**, *79*, C1672–C1677. [CrossRef] [PubMed]
13. Borges, E.M.; Volmer, D.A.; Gallimberti, M.; Ferreira De Souza, D.; Luiz De Souza, E.; Barbosa, F. Evaluation of macro- and microelement levels for verifying the authenticity of organic eggs by using chemometric techniques. *Anal. Methods* **2015**, *7*, 2577–2584. [CrossRef]

14. Bandoniene, D.; Walkner, C.; Zettl, D.; Meisel, T. Rare Earth Element Labeling as a Tool for Assuring the Origin of Eggs and Poultry Products. *J. Agric. Food Chem.* **2018**, *66*, 11729–11738. [CrossRef] [PubMed]
15. Zhang, J.; Zhang, X.; Dediu, L.; Victor, C. Review of the current application of fingerprinting allowing detection of food adulteration and fraud in China. *Food Control* **2011**, *22*, 1126–1135. [CrossRef]
16. Riedl, J.; Esslinger, S.; Fauhl-Hassek, C. Review of validation and reporting of non-targeted fingerprinting approaches for food authentication. *Anal. Chim. Acta* **2015**, *885*, 17–32. [CrossRef] [PubMed]
17. Cuadros-Rodríguez, L.; Ruiz-Samblás, C.; Valverde-Som, L.; Pérez-Castaño, E.; González-Casado, A. Chromatographic fingerprinting: An innovative approach for food "identitation" and food authentication—A tutorial. *Anal. Chim. Acta* **2016**, *909*, 9–23. [CrossRef] [PubMed]
18. Sobolev, A.P.; Circi, S.; Capitani, D.; Ingallina, C.; Mannina, L. Molecular fingerprinting of food authenticity. *Curr. Opin. Food Sci.* **2017**, *16*, 59–66. [CrossRef]
19. Esteki, M.; Shahsavari, Z.; Simal-Gandara, J. Use of spectroscopic methods in combination with linear discriminant analysis for authentication of food products. *Food Control* **2018**, *91*, 100–112. [CrossRef]
20. Chen, H.; Tan, C.; Lin, Z. Non-destructive identification of native egg by near-infrared spectroscopy and data driven-based class-modeling. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2019**, *206*, 484–490. [CrossRef] [PubMed]
21. Cavanna, D.; Catellani, D.; Dall'Asta, C.; Suman, M. Egg product freshness evaluation: A metabolomic approach. *J. Mass Spectrom.* **2018**, *53*, 849–861. [CrossRef] [PubMed]
22. Eigenvector Research Incorporated. Powerful Resources for Intelligent Data Analysis. Available online: http://www.eigenvector.com/software/solo.htm (accessed on 15 January 2019).
23. Massart, D.L.; Vandeginste, B.G.M.; Buydens, L.M.C.; de Jong, S.; Lewi, P.J.; Smeyers-Verbeke, J. *Handbook of Chemometrics and Qualimetrics*; Elsevier: Amsterdam, The Netherlands, 1997.

foods

MDPI

Article

Classification of Hen Eggs by HPLC-UV Fingerprinting and Chemometric Methods

Guillem Campmajó [1], Laura Cayero [1], Javier Saurina [1,2] and Oscar Núñez [1,2,3,*]

[1] Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, Martí i Franquès 1-11, E08028 Barcelona, Spain; campma03@gmail.com (G.C.); laura.cayero@gmail.com (L.C.); nereant7@gmail.com (N.N.), xavi.saurina@ub.edu (J.S.)

[2] Research Institute in Food Nutrition and Food Safety, University of Barcelona, Recinte Torribera, Av. Prat de la Riba 171, Edifici de Recerca (Gaudí), Santa Coloma de Gramenet, E08921 Barcelona, Spain

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

* Correspondence: oscar.nunez@ub.edu; Tel.: +34-93-403-3706

Received: date; Accepted: date; Published: date

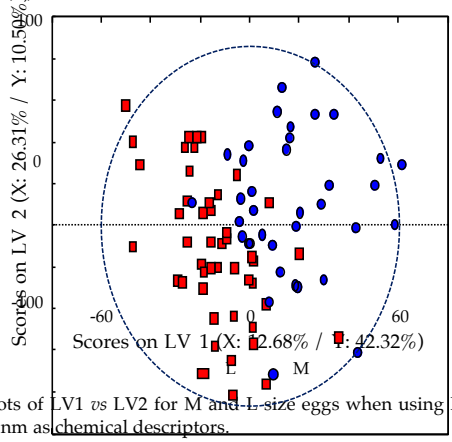


Figure S1. PLS-DA scores plots of LV1 *vs* LV2 for M and L size eggs when using HPLC-UV chromatographic fingerprints registered at 250 nm as chemical descriptors.

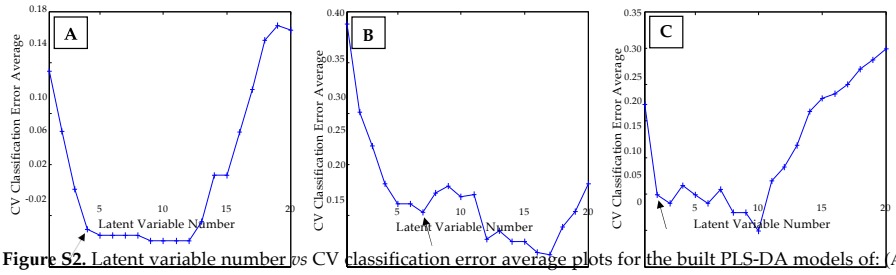


Figure S2. Latent variable number *vs* CV classification error average plots for the built PLS-DA models of: (**A**) O *vs* FR, B and C eggs, (**B**) FR *vs* B and C, and (**C**) B *vs* C.

3.2.3. PUBLICATION VI

High-performance liquid chromatography with fluorescence detection fingerprinting combined with chemometrics for nut classification and the detection and quantitation of almond-based product adulterations.

Campmajó, G.; Saez-Vigo, R.; Saurina, J.; Núñez, O.

Food Control. 2020, 114, 107265.

FRGSRQUR 114(2020) 107265

Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont

High-performance liquid chromatography with fluorescence detection fingerprinting combined with chemometrics for nut classification and the detection and quantitation of almond-based product adulterations

Guillem Campmajó[a,b,*], Ruben Saez-Vigo[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

ARTICLEINFO

Keywords:
Nuts
Almond
HPLC-FLD
Fingerprinting
Chemometrics
Food authentication
Food safety

ABSTRACT

Economically motivated food fraud has increased in recent years, with adulterations and substitutions of high-quality products being common practice. Moreover, this issue can affect food safety and pose a risk to human health by causing allergies through nut product adulterations. Therefore, in this study, high-performance liquid chromatography with fluorescence detection (HPLC-FLD) fingerprints were used for classification of ten types of nuts, using partial least squares regression-discriminant analysis (PLS-DA), as well as for the detection and quantitation of almond-based product (almond flour and almond custard cream) adulterations with hazelnut and peanut, using partial least squares regression (PLS). A satisfactory global nut classification was achieved with PLS-DA. Paired PLS-DA models of almonds in front of their adulterants were also evaluated, producing a classification rate of 100%. Moreover, PLS regression produced low prediction errors (below 6.1%) for the studied adulterant levels, with no significant matrix effect observed.

1. Introduction

Food fraud, which costs the global food industry approximately 30 billion euros a year, has increased because of the complex nature of the globalised world, where many individuals participate in the food chain between production and consumption. In the European Union (EU), the number of requests concerning fraud suspicions sent to the EU Administrative Assistance and Cooperation (AAC) system had increased by 49% from 2016 to 2018 (*European* Comission, 2018). There are different ways of perpetuating food fraud, such as deception during manufacturing, use of illicit supply chains, duplication, misrepresentation, and manipulation of the food product (e.g., adulteration, addition, substitution, etc.) (Manning & Soon, 2019). Although it is generally economically motivated, the addition or replacement of certain substances can be extremely dangerous for human health, for example, by causing allergies, thereby turning a food authentication issue into a food safety one (Fritsche, 2018).

Nuts and seeds, which are widely consumed mainly due to their beneficial effects on human health (De Souza, Schincaglia, Pimente, & Mota, 2017), encompass a wide range of food products such as almonds, Brazil nuts, cashew nuts, hazelnuts, macadamia nuts, peanuts, pecans, pine nuts, pistachios, pumpkin seeds, sunflower seeds, and walnuts. Some of them are at medium or high risk for food fraud (Food Fraud Risk Information Database, 2019), being susceptible to adulterations, replacements or substitutions with cheaper and lower-quality products, as well as to their characteristics being misrepresented (e.g., origin, year of the stock or organic production). For instance, almonds, which are one of the most expensive internationally produced nuts (more than 2 million tonnes produced in 2017, with USA the main producer (Food and Agriculture Organization of the United Nations, 2019)), as well as their byproducts (snacks, baked goods and pastry), can be partly or totally replaced with peanut or hazelnut, constituting not only an economic deception, but also a threat to human health by causing allergies (Mustafa et al., 2019). Therefore, there is an increasing need to develop new analytical methodologies to guarantee the authenticity and safety of almond and almond-based products.

To date, most of the analytical methods described in the literature for almond authentication deal with its agricultural origin, with only a few focusing on its adulteration. For instance, several analytical platforms based on thermal analysis (Beltrán-Sanahuja, Grané-Teruel,

* Corresponding author. Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, Martí i Franquès 1-11, E08028, Barcelona, Spain.
E-mail address: gcampmajo@ub.edu (G. Campmajó).

https://doi.org/10.1016/j.foodcont.2020.107265
Received 25 January 2020; Received in revised form 17 March 2020; Accepted 18 March 2020
[illegible]

Martín-Carratalá, & Garrigós-Selva, 2011), gas chromatography coupled to mass spectrometry for the determination of 12 targeted volatile compounds (Beltrán-Sanahuja, Ramos-Santonja, Grané-Teruel, Martín-Carratalá, & Garrigós-Selva, 2011), high-performance liquid chromatography with an evaporative light-scattering detector (HPLC-ELSD) for triacylglycerol profiling (Barreira et al., 2012), and approaches combining more than one technique (García, Beltrán Sanahuja, & Garrigós Selva, 2013; Čolić et al., 2017), have been successfully employed when combined with chemometric techniques for origin classification. However, to the best of our knowledge, there are very few studies investigating the adulteration of almond-based products. Multi-elemental profiling by inductively coupled plasma-optical emission spectrometry (ICP-OES) has been used to detect and quantitate the adulteration of almond powder with peanut (Esteki, Vander Heyden, Farajmand, & Kolahderazi, 2017), while fatty acid profiles obtained with gas chromatography with flame-ionisation detection (GC-FID) have been employed to study apricot kernel as an adulterant (Esteki, Farajmand, Kolahderazi, & Simal-Gandara, 2017). In both cases, multivariate data analysis was also used to quantify the adulterant level in the studied samples.

While most of the methods described in the literature for almond authentication are based on targeted profiling (a given group of known chemical compounds are determined), chromatographic fingerprinting involving non-targeted instrumental signals has emerged as a promising strategy in the food authentication field since it does not need specific biomarkers. This approach has already been proven in some studies on complex food matrices (Cuadros-Rodríguez, Ruiz-Samblás, Valverde-Som, Pérez-Castaño, & González-Casado, 2016). In fact, high-performance liquid chromatography with ultraviolet detection (HPLC-UV) fingerprinting has been demonstrated to be able to completely distinguish almond samples from peanut and hazelnut ones, although it could not discriminate the whole types of the studied nuts (Campmajó et al., 2019).

Therefore, this study aimed to classify nuts according to their typology, independently of their processing thermal treatment (natural, toasted or fried), by high-performance liquid chromatography with fluorescence detection (HPLC-FLD) fingerprinting, which is a more selective technique than HPLC-UV, and partial least squares regression-discriminant analysis (PLS-DA). Moreover, the chromatographic fingerprints were also used to detect and quantitate hazelnut and peanut adulterations of almond and almond-based products by partial least squares (PLS) regression.

2. Materials and methods

2.1. Reagents and solutions

Unless otherwise stated, all the reagents were of analytical grade. Purified water was obtained using an Elix® 3 coupled to a Milli-Q® system (Millipore Corporation, Bedford, MA, USA) and filtered through a 0.22-μm nylon membrane. Acetone, hexane and formic acid (96%) were obtained from Sigma-Aldrich (St. Louis, MO, USA), whereas UHPLC-gradient grade methanol was from Panreac (Barcelona, Spain).

2.2. Instrumentation

The chromatographic system consisted of an Agilent 1100 Series HPLC instrument equipped with a binary pump (G1312A), a degasser (G1379A), an automatic injection system (G1329B), a fluorescence detector (G1321A) and a computer with the Agilent ChemStation software, all from Agilent Technologies (Waldbronn, Germany). The HPLC-FLD fingerprints were obtained by employing a Kinetex C18 column (100 mm × 4.6 mm id., 2.6 μm particle size), which was purchased from Phenomenex (Torrance, CA, USA), and a previously developed gradient elution mode with 0.1% (*v/v*) formic acid aqueous solution (solvent A) and methanol (solvent B) constituting the components of the mobile phase (Campmajó et al., 2019). The flow rate was 0.4 mL min^{-1} and the injection volume 5 μL. For fluorescence acquisition, 280 nm and 350 nm were chosen as the excitation and emission wavelengths, respectively.

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

Nut type	Abbreviation	Number of samples		
		Natural	Fried	Toasted
Almonds	AL	10	10	10
Cashew Nuts	CN	–	10	–
Hazelnuts	HN	10	–	10
Macadamia Nuts	MN	10	–	–
Peanuts	PN	–	10	10
Pine Nuts	PI	10	–	–
Pistachios	PT	–	–	9
Pumpkin seeds	PS	–	10	10
Sunflower seeds	SS	–	–	9
Walnuts	WN	10	–	–

2.3. Samples and sample treatment

For nut classification, 149 nut samples obtained from Barcelona markets, belonging to various classes and some of them processed with different thermal treatments, were analysed (sample details are described in Table 1). Method repeatability and the robustness of the chemometric results were controlled by using a quality control (QC) sample, which was a mix prepared with 50 μL of each nut sample extract.

Hazelnuts and peanuts were studied as potential adulterants of almonds and almond-based products. Thus, they were added in proportions from 0 to 100%, as shown in Table 2, to two different almond matrices: natural almond flour and almond custard cream. The cream was made from hen eggs, milk, sugar, and corn flour. Afterwards, the almond custard cream and its adulterated samples were obtained by adding the adulterants as described above. Five replicates of each percentage of adulteration were prepared, giving a total of 105 samples for each studied almond-based product. In this study, an additional 50% adulterated sample was prepared for use as the QC sample.

A simple two-step sample treatment was performed following a previously described method (Campmajó et al., 2019) based on an extraction with acetone:water (70:30 *v/v*) followed by a defatting step with hexane. Briefly, 0.125 g of the nut product were extracted by stirring in a Vortex (Stuart, Stone, United Kingdom) and sonication (5510 Branson ultrasonic bath, Hampton, NH, USA) in 3 mL of the extracting solvent. Then, centrifugation was performed for 30 min at 3400 rpm (ROTANTA 460 RS Centrifuge, Hettich, Germany). the

Table 2
Samples used in the PLS adulteration studies as calibration or validation set. Hazelnut and peanut were proposed as adulterants of a natural almond flour and an almond custard cream.

	Almond, %	Adulterant, %
Calibration set	100	0
	80	20
	60	40
	40	60
	20	80
	0	100
Validation set	85	15
	75	25
	50	50
	25	75
	15	85

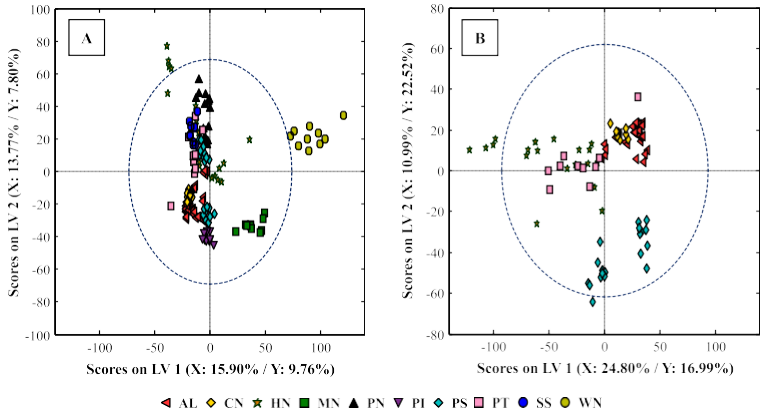


Fig. 1. (A) PLS-DA scores plot of LV1 *vs.* LV2, using the HPLC-FLD fingerprints acquired for all the nut samples tested. (B) PLS-DA scores plot of LV1 *vs.* LV2, using only the almond, cashew nut, hazelnut, pistachio, and pumpkin seed HPLC-FLD fingerprints.

resulting supernatant extract was defatted with 3 mL of hexane, also by stirring in a Vortex followed by centrifugation for 15 min. After filtering the sample extract with a 0.22-μm nylon filter (Scharlab, Sentmenat, Spain), it was stored at −18 °C in a 2-mL glass injection vial until HPLC-FLD analysis.

To avoid and control for systematic errors and cross-contamination during sample sequences, a QC sample and an extracting solvent blank were injected at the beginning and after every ten sample injections.

2.4. Data analysis

Depending on the aim of the multivariate data analysis, principal component analysis (PCA), PLS-DA or PLS regression was carried out by using the Solo 8.6 chemometrics software from Eigenvector Research (Manson, WA, USA) (Eigenvector Research Incorporated, 2019). Details of the theoretical background of these statistical methodologies are addressed elsewhere (Massart et al., 1997).

For the chemometric study, the construction of different data matrices was required. Thus, indistinctly of the chemometric method used, the X-data matrices of responses consisted of the HPLC-FLD chromatographic fingerprints acquired. Furthermore, PLS-DA Y-data matrices defined each sample class, whereas PLS ones defined each percentage of adulteration.

HPLC-FLD fingerprints were smoothed, baseline-corrected, aligned, and autoscaled before building the chemometric model to improve data quality by reducing noise interferences, baseline drifts and peak shifting. Afterwards, the most appropriate number of principal components (PCs) in PCA, and latent variables (LVs) in the PLS-DA and PLS was established at the first significant minimum point of the venetian blind cross validation (CV) error.

Moreover, the applicability of the built chemometric models was tested through their validation. For instance, the PLS-DA models were validated by using 70% of a sample group as the calibration set, and the remaining 30% as the validation set. In the case of the PLS models, Table 2 shows the percentages of adulteration used in the calibration and validation sets.

3. Results and discussion

3.1. Nut classification

Several types of nuts are vulnerable to food fraud practices such as being substituted with cheaper adulterants. Therefore, analytical methodologies capable of classifying nut samples according to their type are required. Although a previous study demonstrated that HPLC-UV fingerprints were good chemical descriptors for classifying certain types of nuts, they could not achieve complete nut classification (Campmajó et al., 2019). Thus, in this work, HPLC-FLD fingerprints were used as an alternative to obtain better descriptors.

3.1.1. HPLC-FLD fingerprints

As previously mentioned in Section 2.3, a wide variety of nut samples were assessed by HPLC-FLD for classification. As can be seen in Fig. S1 (Supplementary Material) showing the chromatographic fingerprints acquired for a selected sample, there were noteworthy differences in the abundance of the compounds detected (considering the retention time), as well as in the peak intensity. Moreover, since these features were reproducible among samples belonging to the same type of nut, these chemical descriptors were evaluated to classify nut types through a multivariate chemometric approach.

3.1.2. Chemometrics for classification

First, a preliminary exploratory chemometric PCA was performed to study QC sample behaviour. Therefore, a 164 × 4863 (samples × variables) dimension data matrix, with the emitted fluorescence intensity at 350 nm a function of time for the analysed nut and QC samples, was examined. As shown in Fig. S2, QC samples formed a compact group in the central part of the scores plot of PC1 *vs.* PC2 (two PCs were chosen for the PCA), indicating the absence of systematic errors during the sample injection sequence and demonstrating the validity of the chemometric results.

The supervised chemometric analysis for classification was conducted with PLS-DA. While the X-data matrix (149 × 4863) consisted of the same information as that used in the PCA without the QC samples, the Y-data matrix (149 × 2) indicated the membership of each nut sample. Due to the large number of nut classes under study, a total of ten LVs were required for the construction of the PLS-DA model, which clearly enabled the discrimination of some of them. For instance, the scores plot of LV1 *vs.* LV2 (Fig. 1A) shows a clear separation of walnuts and macadamia nuts, which are on the right side of the plot displaying positive LV1 values, whereas pine nuts are at the bottom of the plot with negative LV2 values. Although the combination of other LVs and the use of 3D plots also enabled the classification of peanuts (Fig. S3A) and sunflower seeds (Fig. S3B), LV construction was mainly influenced

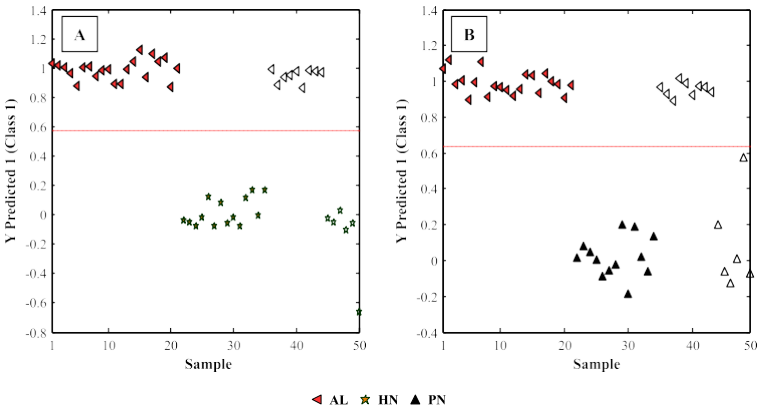


Fig. 2. Classification plot depicting Sample *vs.* Y predicted 1 score plot for (A) almond *vs.* hazelnut samples and (B) almond *vs.* peanut samples. Solid symbols, calibration samples; empty symbols, validation samples.

by these classes of nuts, with the scores plots not visually discriminating between the remaining five classes. For that reason, a new PLS-DA model for almond, cashew nut, hazelnut, pistachio, and pumpkin seed samples was built with four LVs. This resulted in better classification, especially for sunflower seeds, as can be seen in the corresponding scores plot of LV1 *vs.* LV2 in Fig. 1B.

As this work focused on the study of almond adulterations, which commonly constitute its substitution with cheaper nuts such as hazelnuts or peanuts, paired PLS-DA models with almond in front of hazelnut and peanut samples were constructed. As previously detailed in Section 2.4, 70% of the samples were used in the calibration set, whereas the remaining 30% were used in the validation set. Fig. 2 presents these classification plots, the red dashed line indicating the classification boundary. The calibration and validation samples are located on the left and right side of the plot, respectively. A classification rate of 100% was obtained when studying almonds in front of their most common adulterating nuts, [9, 0; 0, 6] being the confusion matrix for both almond *vs.* hazelnut and almond *vs.* peanut validations.

Although UV fingerprints at 280 nm are much richer in peak features than the FLD counterparts, results presented in this paper demonstrate the better descriptive performance of HPLC-FLD data compared with HPLC-UV (Campmajó et al., 2019), with higher classification rates and lower prediction errors for some of the systems under study. The selectivity of UV spectroscopy at 280 nm is poor and a wide range of compounds are detected, mainly consisting of phenolic acids (and flavonoids with lower sensitivity), which are components occurring in all kinds of samples. As a result, the nut discrimination is then based on cross selectivities (i.e., differences in concentration levels among classes), while more specific markers have not been encountered. In contrast, FLD fingerprints generally contain a fewer number of peaks since the selection of excitation and emission conditions provides more selective data (Bakhytkyzy, Nuñez, & Saurina, 2018). Moreover, signals from hydroxycinnamic acids, stilbenoids and various types of flavonoids are negligible; only hydroxybenzoic acids and flavanols are reasonably detectable under these conditions. In particular, the detection of flavanols is especially favored, thus achieving a great sensitivity for catechin, epicatechin, and related species. Therefore, despite having simpler chromatograms from FLD in terms of the number of features, the more selective detection of highly relevant descriptors may lead to better predictive figures.

3.2. Almond-based product adulterations

Following the satisfactory classification obtained with the PLS-DA models, HPLC-FLD fingerprints were also used for the detection and quantitation of adulterations in two types of almond-based matrices: natural almond flour and almond custard cream. PLS was applied as the most suitable chemometric approach to study them.

3.2.1. HPLC-FLD fingerprints

A set of almond-based product (natural almond flour and almond custard cream) samples, which were obtained by adding different percentages of the adulterant as specified in Section 2.3 and detailed in Table 2, were analysed with HPLC-FLD.

As shown in Fig. S1, both the pure hazelnut and peanut fingerprints showed significant differences compared to the almond ones in terms of the number of compounds detected, abundance, and intensity. For instance, the peanut and hazelnut samples presented a higher number of chromatographic peaks than the almond samples. In fact, an increase in the number of peaks could be seen when transitioning from pure almond to adulterated samples. Therefore, as the HPLC-FLD fingerprints seemed to vary according to the adulterant percentage, they were proposed as chemical descriptors to detect and quantitate adulterations, using PLS.

3.2.2. Chemometric detection and quantitation of adulterations

The ability of the HPLC-FLD fingerprints to detect and quantify almond adulterations with peanut or hazelnut was evaluated by PLS. Table 3 summarises the LVs used in each calibration PLS model, as well as the calibration and prediction error obtained in all the adulteration cases studied. The calibration models built were good, as indicated by

Table 3
Overall results for the evaluation of the adulteration of almond flour and almond custard cream with hazelnut and peanut by PLS. LVs, number to build each PLS mode; Cal. Error, error in the calibration step; Pred. Error, error in the prediction step.

	Almond flour			Almond custard cream		
	LVs	Cal. Error (%)	Pred. error (%)	LVs	Cal. Error (%)	Pred. error (%)
HAZELNUT	5	2.6	5.6	4	3.5	6.1
PEANUT	3	4.7	5.0	4	3.1	6.1

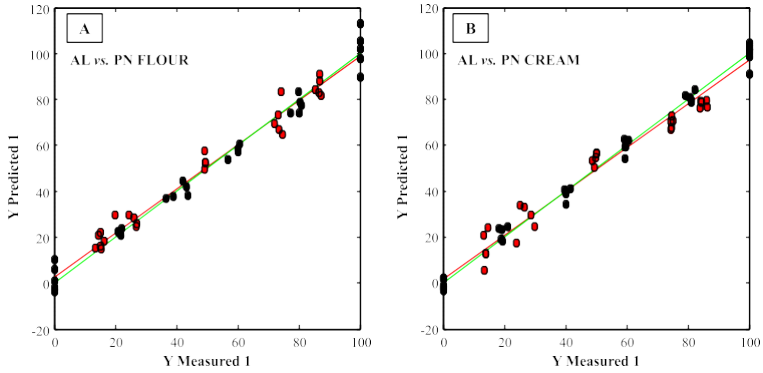


Fig. 3. Scatter plot of measured *vs.* predicted percentages of adulteration, using PLS. Results are shown for (A) almond flour and (B) almond custard cream adulterated with peanut.

the low calibration errors (≤4.7%), bias values tending towards zero and good linearity with $R^2 \geq 0.982$. When focusing on a specific matrix, similar prediction errors were obtained independently of the adulterant used. As can be seen in Fig. 3, the results achieved when predicting peanut levels in almond flour (Fig. 3A) and almond custard cream (Fig. 3B) were excellent, with no significant differences between the matrices (PLS results for the adulteration with hazelnut are shown in Fig. S4). Hence, although almond custard cream is a fatter matrix than almond flour, no interfering matrix effect was observed in the results.

4. Conclusions

HPLC-FLD chromatographic fingerprints, using an excitation wavelength of 280 nm and an emission wavelength of 350 nm, were suitable chemical descriptors for nut classification and authentication. Satisfactory discrimination of nut samples according to their type was achieved by PLS-DA. Moreover, when focusing on the specific adulteration of almond-based products with peanut or hazelnut, paired PLS-DA models showed complete sample distinction (classification rate of 100%), while PLS models produced low prediction errors below 6.1% for both matrices when predicting the percentages of adulteration. Thus, the HPLC-FLD fingerprinting method described in this study can classify nut samples according to their type, as well as detect and quantitate the levels of peanut or hazelnut adulteration of almond-based products. Therefore, it can be used as a simple and reliable method to prevent food fraud and guarantee food product safety.

Funding

The authors are grateful for the financial support received from the Spanish Ministry of Science, Innovation and Universities, through the grant PGC2018-095013-B-I00, and from the Agency for Administration of University and Research Grants (Generalitat de Catalunya), through the grants 2017SGR-171 and 2017SGR-310.

CRediT authorship contribution statement

Guillem Campmajó: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Supervision. Ruben Saez-Vigo: Methodology, Validation, Investigation. Javier Saurina: Conceptualization, Writing - review & editing, Supervision, Funding acquisition. Oscar Núñez: Conceptualization, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition.

Declaration of competing interest

There are no conflicts of interest to declare.

Acknowledgements

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodcont.2020.107265.

References

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.

Barreira, J. C. M., Casal, S., Ferreira, I. C. F. R., Peres, A. M., Pereira, J. A., & Oliveira, M. B. P. P. (2012). Supervised chemical pattern recognition in almond (Prunus dulcis) Portuguese PDO cultivars: PCA- and LDA-based triennial study. *Journal of Agricultural and Food Chemistry, 60*(38), 9697–9704. https://doi.org/10.1021/jf301402t.

Beltrán-Sanahuja, A., Grané-Teruel, N., Martín-Carratalá, M. L., & Garrigós-Selva, M. C. (2011a). Characterization of almond cultivars by the use of thermal analysis techniques. Application to cultivar authenticity. JAOCS. *Journal of the American Oil Chemists' Society, 88*(11), 1687–1693. https://doi.org/10.1007/s11746-011-1847-3.

Beltrán-Sanahuja, A., Ramos-Santonja, M., Grané-Teruel, N., Martín-Carratalá, M. L., & Garrigós-Selva, M. C. (2011b). Classification of almond cultivars using oil volatile compound determination by HS-SPME-GC-MS. JAOCS. *Journal of the American Oil Chemists' Society, 88*(3), 329–336. https://doi.org/10.1007/s11746-010-1685-8.

Campmajó, G., Navarro, G. J., Núñez, N., Puignou, L., Saurina, J., & Núñez, O. (2019). Non-targeted HPLC-UV fingerprinting as chemical descriptors for the classification and authentication of nuts by multivariate chemometric methods. *Sensors, 19*(6), https://doi.org/10.3390/s19061388.

Čolić, S. D., Fotirić Akšić, M. M., Lazarević, K. B., Zec, G. N., Gašić, U. M., Dabić Zagorac, D., et al. (2017). Fatty acid and phenolic profiles of almond grown in Serbia. *Food Chemistry, 234*, 455–463. https://doi.org/10.1016/j.foodchem.2017.05.006.

Cuadros-Rodríguez, L., Ruiz-Samblás, C., Valverde-Som, L., Pérez-Castaño, E., & González-Casado, A. (2016). Chromatographic fingerprinting: An innovative approach for food "identification" and food authentication - a tutorial. *Analytica Chimica Acta, 909*, 9–23. https://doi.org/10.1016/j.aca.2015.12.042.

De Souza, R. G. M., Schincaglia, R. M., Pimente, G. D., & Mota, J. F. (2017). Nuts and human health outcomes: A systematic review. *Nutrients, 9*(12), https://doi.org/10.3390/nu9121311.

Esteki, M., Farajmand, B., Kolahderazi, Y., & Simal-Gandara, J. (2017a). Chromatographic fingerprinting with multivariate data analysis for detection and quantification of apricot kernel in almond powder. *Food Analytical Methods, 10*(10), 3312–3320. https://doi.org/10.1007/s12161-017-0903-5.

Esteki, M., Vander Heyden, Y., Farajmand, B., & Kolahderazi, Y. (2017b). Qualitative and quantitative analysis of peanut adulteration in almond powder samples using multi-elemental fingerprinting combined with multivariate data analysis methods. *Food Control, 82*, 31–41. https://doi.org/10.1016/j.foodcont.2017.06.014.

Comission, E. (2018). *The EU food fraud network and the system for administrative assistance - food fraud Annual Report 2018*. (n.d.).

Food and Agriculture Organization of the United Nations. Food and agriculture data. (2019). http://www.fao.org/faostat/en/#data/QC/visualize/, Accessed date: 29 December 2019 ((n.d.)).

Risk Information, F. F.. Food fraud risk information database. (2019). https://trello.com/b/aoFO1UEf/food-fraud-risk-information/, Accessed date: 29 December 2019 ((n.d.)).

Fritsche, J. (2018). Recent developments and digital perspectives in food safety and authenticity. *Journal of Agricultural and Food Chemistry, 66*(29), 7562–7567. https://doi.org/10.1021/acs.jafc.8b00843.

García, A. V., Beltrán Sanahuja, A., & Garrigós Selva, M. del C. (2013). Characterization and classification of almond cultivars by using spectroscopic and thermal techniques. *Journal of Food Science, 78*(2), 138–144. https://doi.org/10.1111/1750-3841.12031.

Incorporated, E. R. (2019). Eigenvector Research incorporated. Powerful resources for intelligent data analysis (n.d.), Available online http://www.eigenvector.com/software/solo.htm, Accessed date: 27 December 2019.

Manning, L., & Soon, J. M. (2019). Food fraud vulnerability assessment: Reliable data sources and effective assessment approaches. *Trends in Food Science & Technology, 91*(July), 159–168. https://doi.org/10.1016/j.tifs.2019.07.007.

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.). Amsterdam: Elsevier ((n.d.)).

Mustafa, S. S., Vadamalai, K., Bingemann, T., Mortezavi, M., Aranez, V., & Ramsey, A. (2019). *Real world tree nut consumption in peanut-allergic individuals. Annals of Allergy, Asthma & Immunology.* Official Publication of the American College of Allergy, Asthma, & Immunologyhttps://doi.org/10.1016/j.anai.2019.11.027.

Supplementary Material

High-performance liquid chromatography with fluorescence detection fingerprinting combined with chemometrics for nut classification and the detection and quantitation of almond-based product adulterations

Guillem Campmajó[a,b,*], Ruben Saez-Vigo[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.: 34-93-403-3706, fax: 34-93-402-1233, e-mail: gcampmajo@ub.edu)

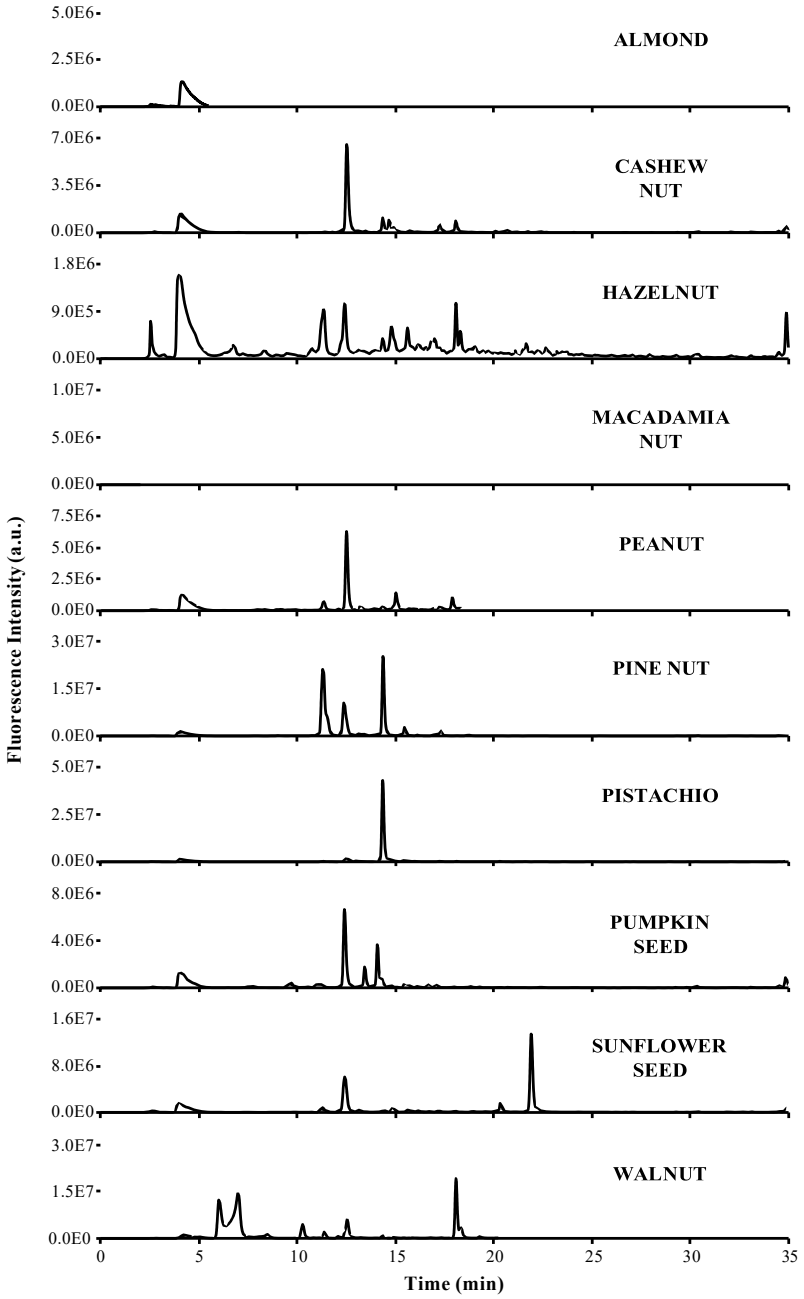


Figure S1. HPLC-FLD fingerprints (acquired with an excitation and emission wavelength of 280 and 350 nm, respectively) for a selected sample for each nut type under study.

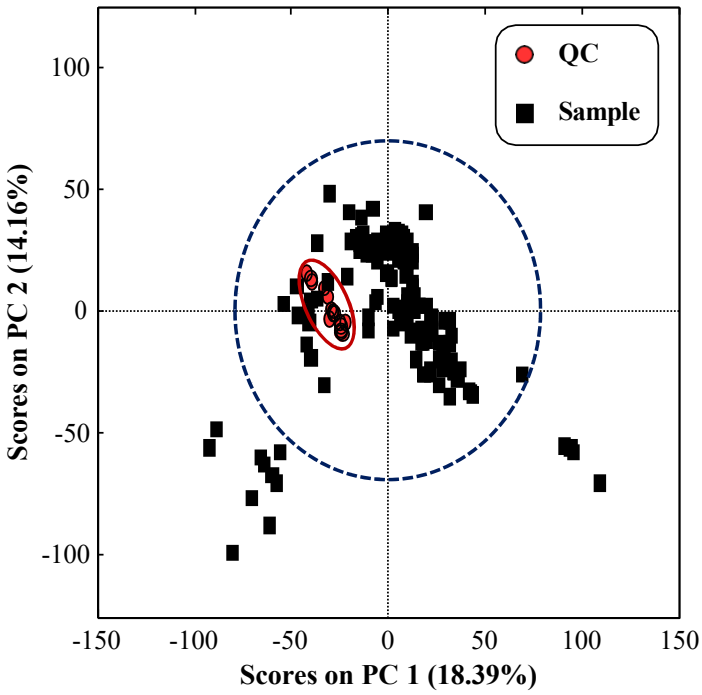


Figure S2. PCA scores plot of PC1 vs. PC2 showing the correct behaviour of QC samples.

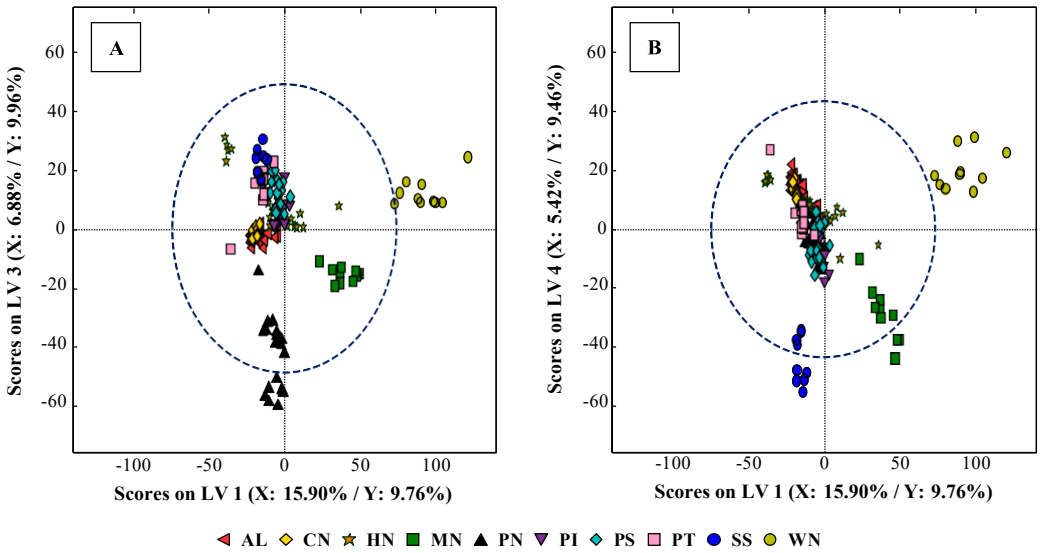


Figure S3. PLS-DA scores plot of (A) LV1 vs. LV3 and (B) LV1 vs. LV4, using the HPLC-FLD fingerprints acquired for all the nut samples assessed.

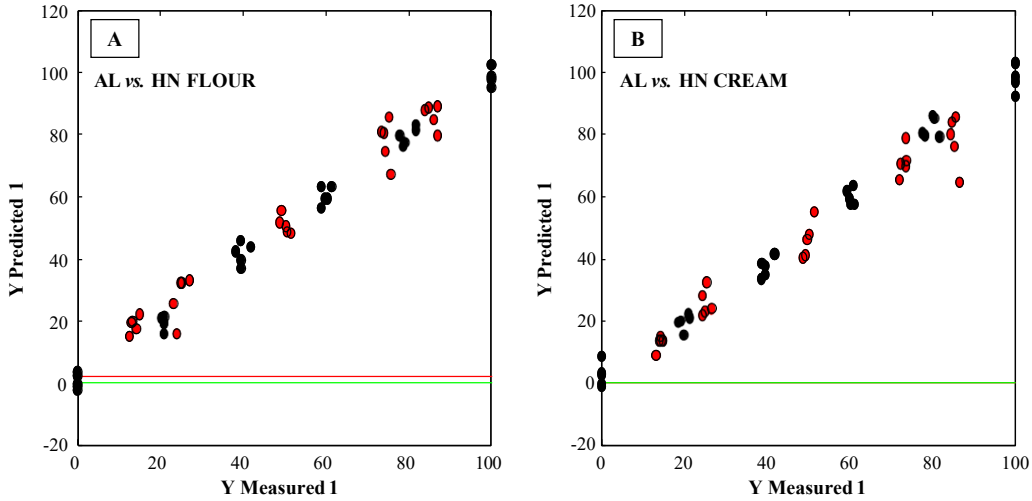


Figure S4. PLS results of (A) almond flour and (B) almond custard cream adulterated with hazelnut.

3.2.4. PUBLICATION VII

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

Campmajó, G.; Rodríguez-Javier, L. R.; Saurina, J.; Núñez, O.

Food Chemistry. 2021, 352, 129397.

Food Chemistry 352 (2021) 129397

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

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

Guillem Campmajó [a,b,*], Luis R. Rodr gue - avier [a], avier Saurina [a,b], Oscar N ñe [a,b,c]

[a] Department of Chemical Engineering and Analytical ChemistryJ University of BarcelonaJ Martí i Franquès 1-11J E08028 BarcelonaJ Spain
[b] Research Institute in Food Nutrition and Food SafetyJ University of BarcelonaJ Av. Prat de la Riba 171J Edifici Recerca (Gaudí)J E08921 Santa Coloma de GramenetJ Spain
[c] Serra Húnter FellowJ Generalitat de CatalunyaJ Rambla de Catalunya 19-21J E08007 BarcelonaJ Spain

ARTICLEINFO

eywords
Paprika
HPLC-FLD
Fingerprinting
Chemometrics
Food Authentication
Protected Designation of Origin

ABSTRACT

Paprika production under the protected designation of origin (PDO) standardi ed 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 C ech 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.

l. 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 nion (E) 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 nion, 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 (*alocsai fűszerpaprika-őrlemény* and *Szegedi fűszerpaprika-őrlemény*), and one from Slovakia (*˘itavská 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, usof, ahaya, Ro ali, Othman, 2019 Škrovánková, Mlček, Orsavová, ur ková, Dř malová, 2017 Topu , Dincer, Ö demir, Feng, Kushad, 2011). Nevertheless, herbs and spices are among the goods most vulnerable to fraudulent practices in the E (European Comission, 2019). In this line, paprika production under the PDO standardi ed procedures leads to more quality products with

* Corresponding author.
E-mail address gcampmajo@ub.edu (G. Campmajó).

https://doi.org/10.1016/j.foodchem.2021.129397
Received 21 December 2020 Received in revised form 2 February 2021 Accepted 14 February 2021
Available online 23 February 2021

higher retail prices (Dane is, 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, urado, Alcá ar, De Pablos, 2014) or mass spectrometry (ICP-MS) (Ördög et al., 2018), liquid chromatography coupled to low- (LC-MS) (Barbosa, Campmajó, Saurina, Puignou, N ñe , 2020) or high-resolution mass spectrometry (LC-HRMS) (Arrieabalaga-Larrañaga et al., 2020 Barbosa, Saurina, Puignou, N ñe , 2020b Mudrić et al., 2017) have been used to record the latter. On the other hand, several fingerprinting approaches (Cuadros-Rodr gue , Rui -Samblás, Valverde-Som, Pére -Castaño, Gon ále -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- V) (Cetó, Sánche , Serrano, D a -Cru , N ñe , 2020 Cetó et al., 2018), electrochemical detection (LC-ECD) (Serrano et al., 2018), or LC-HRMS (Barbosa, Saurina, Puignou, N ñe , 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 C ech 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, SA) was used to obtain purified water, correspondingly filtered with a 0.22-µm nylon membrane. Moreover, HPLC-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 si e) and guard column (2 mm × 4.6 mm id, 2.6 µm particle si e), both from Phenomenex (Torrance, CA, SA), 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 C ech Republic— and types —hot, bittersweet, and sweet— were analy ed 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 *alocsa* region, while the 15 C ech paprika samples' region was not labeled. *La Vera* samples were directly purchased from paprika production companies, whereas the others were bought in C ech, Hungarian, and Spanish commercial supermarkets and markets. Moreover, a quality control (QC) sample, constituted by pooling 50 µL of each analy ed 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 C ech 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 C ech 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, nited Kingdom) for 1 min, sonicated (5510 Branson ultrasonic bath, Hampton, NH, SA) 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 minimi e 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, SA) 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, SA). Details of the theoretical background of these statistical methodologies

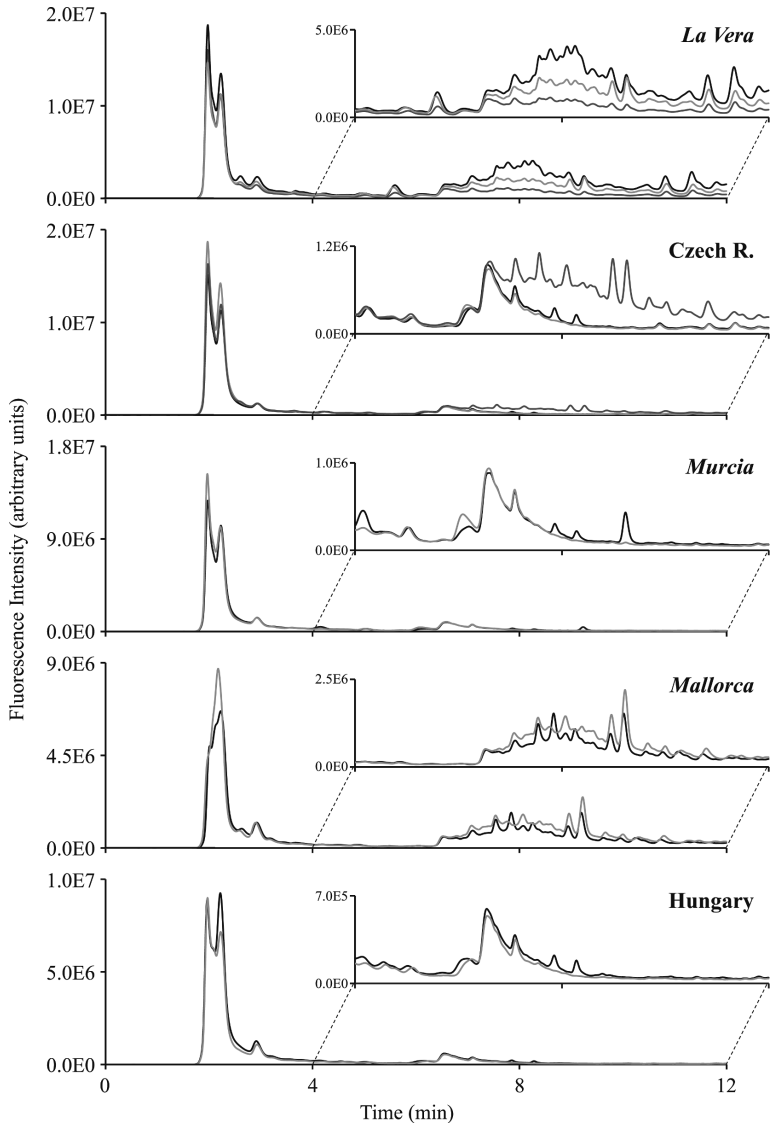


Fig. 1. HPLC-FLD fingerprints, acquired at an excitation wavelength of 310 nm and an emission wavelength of 380 nm, for a selected sample within each paprika region and type. Black and light grey indicate the hot and sweet type, respectively, while medium gray corresponds to bittersweet in *La Vera*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

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 -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-

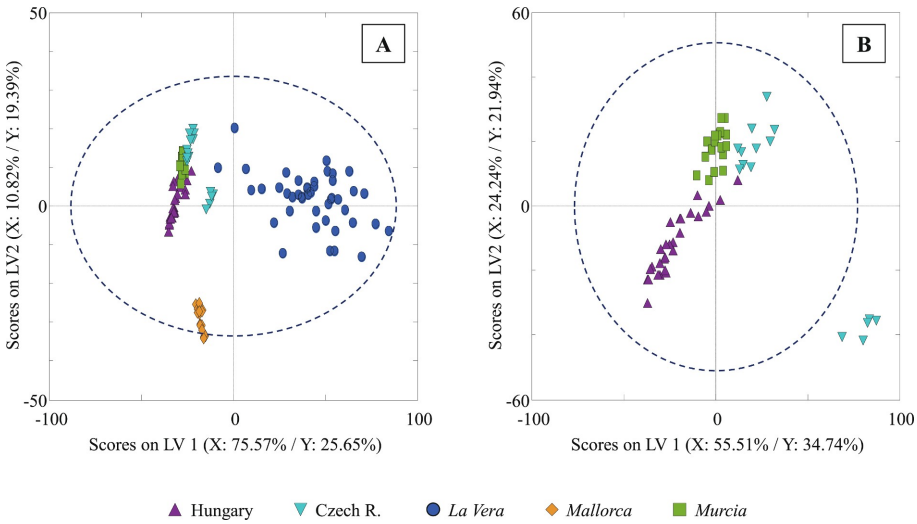


Fig. 2. (A) PLS-DA scores plot of LV1 *vs.* LV2, using the HPLC-FLD fingerprints acquired for all the paprika samples analy ed. (B) PLS-DA scores plot of LV1 *vs.* LV2, using only *Murcia*, Hungary, and the C ech Republic HPLC-FLD fingerprints.

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 V and MS detection of phenolic and polyphenolic molecules (Monasterio, Olmo-Garc a, Bajoub, Fernánde -Gutiérre , Carrasco-Pancorbo, 2016), since many of them —hydroxyben oic, hydroxyphenylacetic, and hydroxycinnamic acids, tyrosols (Godoy-Caballero, Acedo-Valen uela, Galeano-D a , 2012), lignans (Selvaggini et al., 2006), and flavanols (Bakhytky y, Nuñe , 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, N ñe , 2017), HPLC-FLD fingerprinting has recently gained interest in this field. Its application has already proven excellent descriptive performance when analy ing 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 (N ñe , Mart ne , Saurina, N ñe , 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- V (Campmajó, Sae -Vigo, Saurina, N ñe , 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 optimi ation 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, N ñe , 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 characteri ed by distinctive chromatographic fingerprints comparing to the remaining regions, which at first glance showed more similarities. In addition, these

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) C ech 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

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 -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 C ech paprika samples was built with six LVs, explaining a 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 C ech paprika appears separated from the other C ech 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 C ech 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 C ech 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 analy ed 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%.

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) C ech 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

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 C ech Republic *vs. Murcia*) by analy ing a set of mixed sweet samples (sweet smoked in the case of the C ech 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 C ech 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 -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%.

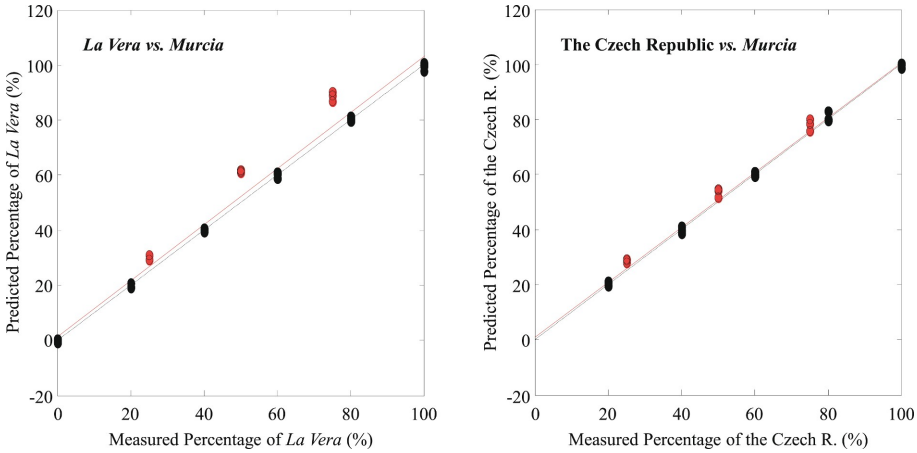


Fig. 3. Prediction test PLS results: *La Vera vs Murcia* (on the left side) and the C ech Republic *vs Murcia* (on the right side) scatter plot of measured *vs* predicted percentages of paprika blend level. Red and black symbols indicate prediction and calibration samples, respectively. Black dashed line corresponds to the theoretical diagonal line, while the red line to the experimental adjusted one. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Given these results, the built models' applicability was tested by analy ing 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 C ech 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 C ech 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

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

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

Funding

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2021.129397.

References

Arri abalaga-Larrañaga, A., Campmajó, G., Saurina, ., N ñe , O., Santos, F. ., & Moyano, E. (2020). Determination of capsaicinoids and carotenoids for the characteri ation and geographical origin authentication of paprika by HPLC–APCI–HRMS. *LWT*. https://doi.org/10.1016/j.lwt.2020.110533.

Bajoub, A., Medina-Rodr gue , S., Gómе -Romero, M., Ajal, E. A., Bagur-Gon ále , M. G., Fernánde -Gutiérre , 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,J 215*, 245–255. https://doi.org/10.1016/j.foodchem.2016.07.140.

Bakhytky y, I., Nuñe , O., & Saurina, . (2018). Determination of flavanols by liquid chromatography with fluorescence detection. Application to the characteri ation of cranberry-based pharmaceuticals through profiling and fingerprinting approaches.

Journal of Pharmaceutical and Biomedical AnalysisJ 156, 206−213. https://doi.org/10.1016/j.jpba.2018.04.031.

Ballin, N. ., & 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 TechnologyJ 86*, 537−543. https://doi.org/10.1016/j.tifs.2018.09.025.

Barbosa, S., Campmajó, G., Saurina, ., Puignou, L., & N ñe , 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 ChemistryJ 68*(2), 591−602. https://doi.org/10.1021/acs.jafc.9b06054.

Barbosa, S., Pardo-Mates, N., Puignou, L., & N ñe , O. (2017). The role of polyphenols and polyphenolic fingerprinting profiles in the characteri ation and authentication of natural food products. In . C. Taylor (Ed.), *Advances in Chemistry Research*. Nova Science Publishers, Inc.

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

Barbosa, S., Saurina, ., Puignou, L., & N ñe , O. (2020b). Targeted HPLC–HRMS (Orbitrap) polyphenolic and capsaicinoid profiling for the chemometric characteri ation and classification of paprika with protected designation of origin (PDO) attributes. *MoleculesJ 25*, 1623. https://doi.org/10.3390/molecules25071623.

Campmajó, G., Sae -Vigo, R., Saurina, ., & N ñe , 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 ControlJ 114*, 107265. https://doi.org/10.1016/j.foodcont.2020.107265.

Cetó, X., Sánche , C., Serrano, N., D a -Cru , . M., & N ñe , O. (2020). Authentication of paprika using HPLC- V fingerprints. *LwtJ 124*, 109153. https://doi.org/10.1016/j.lwt.2020.109153.

Cetó, X., Serrano, N., Aragó, M., Gáme , A., Esteban, M., D a -Cru , . M., & N ñe , O. (2018). Determination of HPLC- V fingerprints of spanish paprika (Capsicum annuum L.) for its classification by linear discriminant analysis. *Sensors (Switzerland)J 18*, 4479. https://doi.org/10.3390/s18124479.

Cuadros-Rodr gue , L., Rui -Samblás, C., Valverde-Som, L., Pére -Castaño, E., & Gon ále -Casado, A. (2016). Chromatographic fingerprinting: An innovative approach for food "identitation" and food authentication – A tutorial. *Analytica Chimica ActaJ 909*, 9−23. https://doi.org/10.1016/j.aca.2015.12.042.

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

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

European Comission. (2020). eAmbrosia - the E 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, ., 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 ControlJ 120*, 107496. https://doi.org/10.1016/j.foodcont.2020.107496.

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

Hassan, N. M., usof, N. A., ahaya, A. F., Ro ali, N. N. M., & Othman, R. (2019). Carotenoids of Capsicum fruits: Pigment profile and health-promoting functional attributes. *AntioxidantsJ 8*, 1−25. https://doi.org/10.3390/antiox8100469.

Lucci, P., Saurina, ., & N ñe , O. (2017). Trends in LC-MS and LC-HRMS analysis and characteri ation of polyphenols in food. *TrAC - Trends in Analytical ChemistryJ 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 ong, S., Lewi, P. ., & Smeyers-Verbeke, . (1997). *Handbook of chemometrics and qualimetrics* ((1st ed.). Amsterdam, The Netherlands: Elsevier.

Monasterio, R. P., Olmo-Garc a, L., Bajoub, A., Fernánde -Gutiérre , 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 SciencesJ 17*, 1627. https://doi.org/10.3390/ijms17101627.

Mudrić, S. ˇ., Gašić, . M., Dramićanin, A. M., Ćirić, I. ˇ., Milojković-Opsenica, D. M., Popović- ordević, . B., ... Tešić, ˇ. L. (2017). The polyphenolics and carbohydrates as indicators of botanical and geographical origin of Serbian autochthonous clones of red spice paprika. *Food ChemistryJ 217*, 705−715. https://doi.org/10.1016/j.foodchem.2016.09.038.

N ñe , N., Mart ne , C., Saurina, ., & N ñe , 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 AgricultureJ 101*, 65−73. https://doi.org/10.1002/jsfa.10615.

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

Palacios-Morillo, A., urado, . M., Alcá ar, Á., & De Pablos, F. (2014). Geographical characteri ation of Spanish PDO paprika by multivariate analysis of multielemental content. *TalantaJ 128*, 15−22. https://doi.org/10.1016/j.talanta.2014.04.025.

Selvaggini, R., Servili, M., rbani, 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 ChemistryJ 54*(8), 2832−2838. https://doi.org/10.1021/jf0527596.

Serrano, N., Cetó, X., N ñe , O., Aragó, M., Gáme , A., Ariño, C., & D a -Cru , . M. (2018). Characteri ation and classification of Spanish paprika (Capsicum annuum L.) by liquid chromatography coupled to electrochemical detection with screen-printed carbon-based nanomaterials electrodes. *TalantaJ 189*, 296−301. https://doi.org/10.1016/j.talanta.2018.06.085.

Škrovánková, S., Mlček, ., Orsavová, ., ur ková, T., & Dř malová, P. (2017). Polyphenols content and antioxidant activity of paprika and pepper spices. *PotravinarstvoJ 11*, 52−57. https://doi.org/10.5219/695.

The European Parliament and the Council of the European nion. (2012). Regulation (E) 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 UnionJ L343*, 1−29.

Topu , A., Dincer, C., Ö demir, K. S., Feng, H., & Kushad, M. (2011). Influence of different drying methods on carotenoids and capsaicinoids of paprika (Cv., alapeno). *Food ChemistryJ 129*, 860−865. https://doi.org/10.1016/j.foodchem.2011.05.035.

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

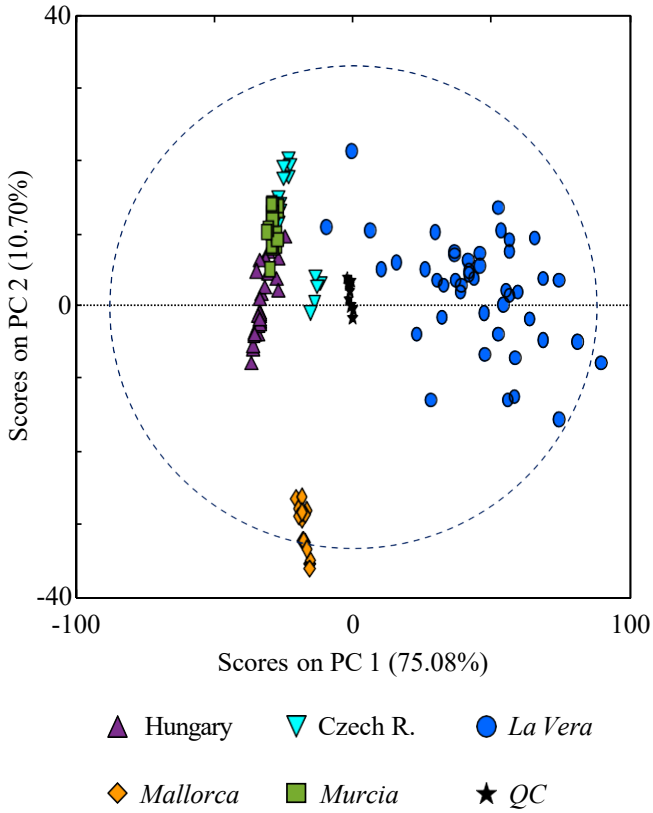


Figure S1. PCA scores plot showing the QC samples' correct behavior in the classification study and some trends associated to sample geographical origin.

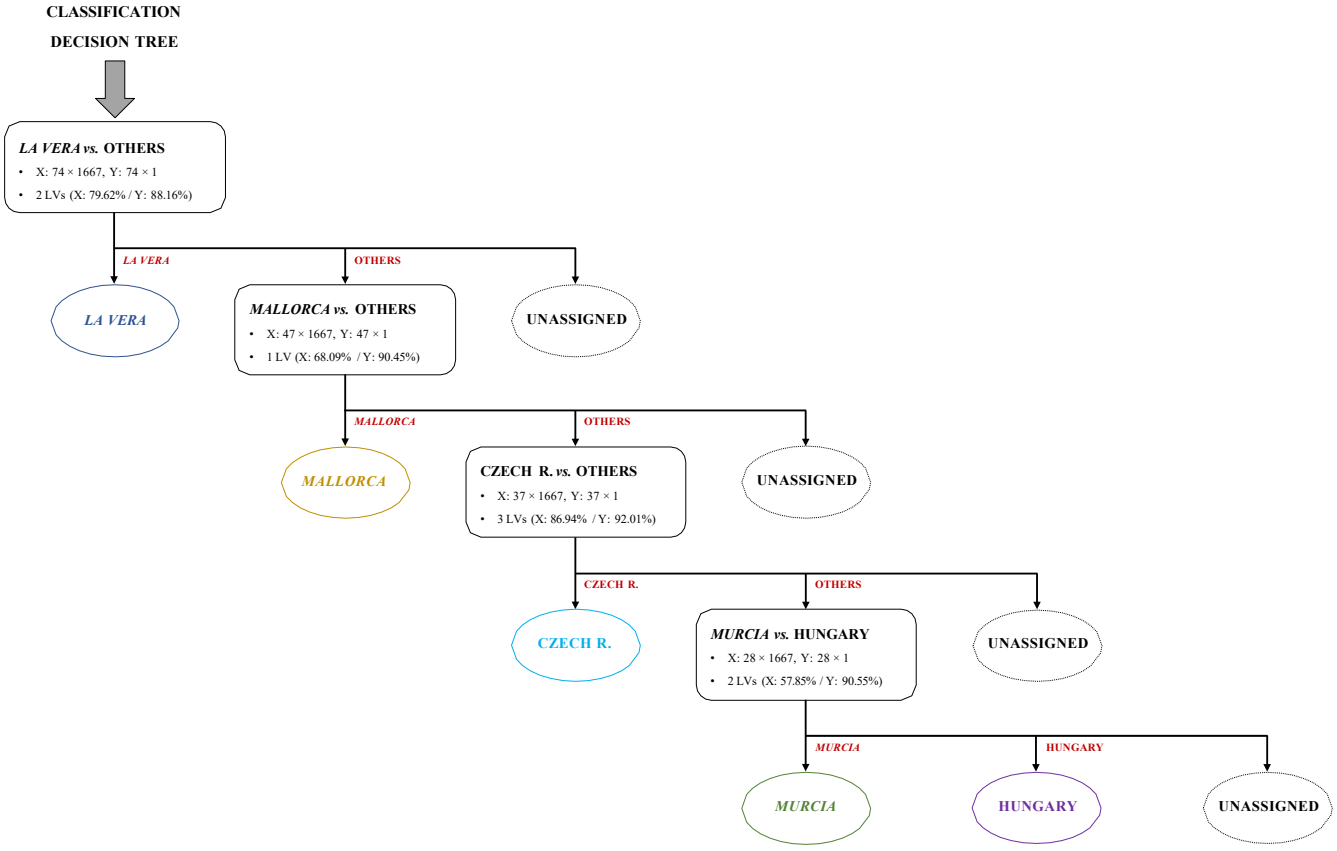


Figure S2. Flow-chart of the designed classification decision tree build using PLS-DA models as the rule nodes. Data matrices dimensions and LVs used to construct the calibration models are detailed.

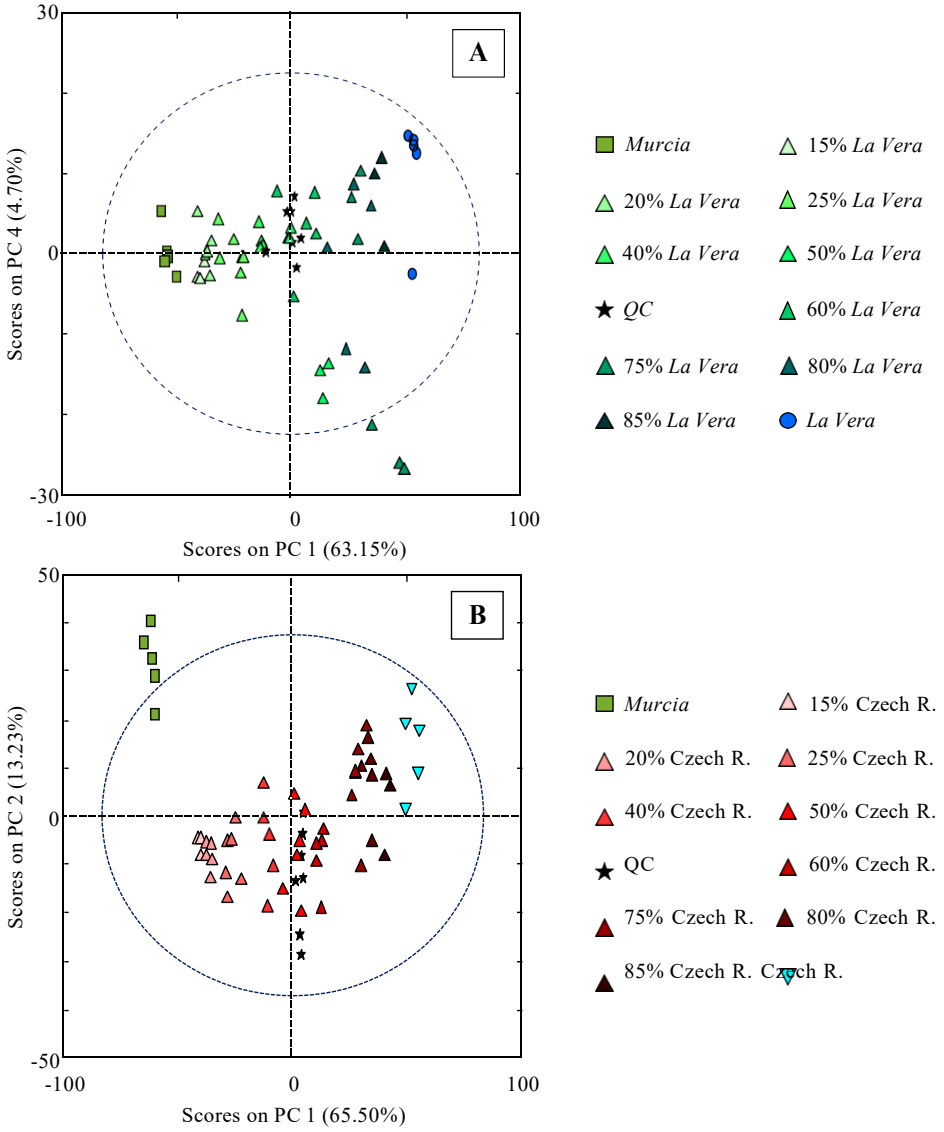


Figure S3. PCA scores plot showing the QC samples' correct behavior in the study to detect and quantitate paprika geographical origin fraud: (A) La Vera *vs*. Murcia PLS case and (B) the Czech Republic *vs*. Murcia case.

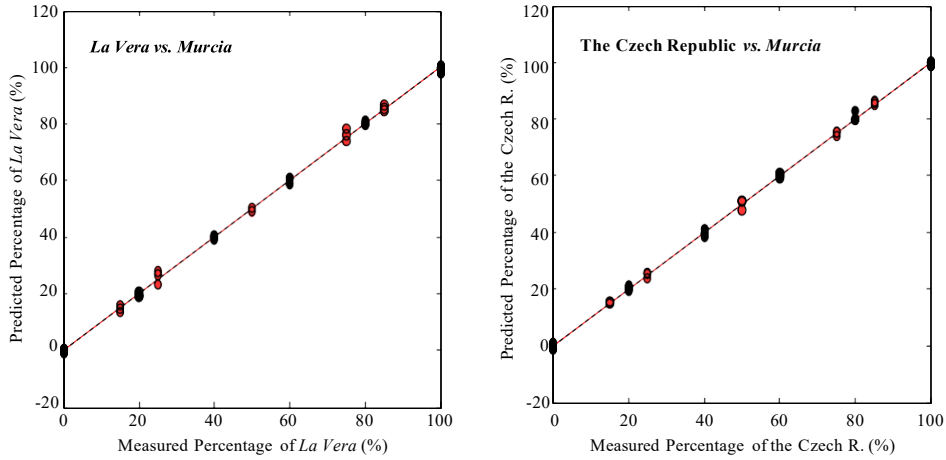


Figure S4. External validation PLS results: *La Vera vs. Murcia* (on the left side) and the Czech Republic *vs. Murcia* (on the right side) scatter plot of measured *vs* predicted percentages of paprika blend level. Red and black symbols indicate external validation and calibration samples, respectively.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

3.2.5. PUBLICATION VIII

Liquid chromatography coupled to high-resolution mass spectrometry for nut classification and marker identification.

Campmajó, G.; Saurina, J.; Núñez, O.

Food Control. 2023, 152, 109834.

Food Control 152 (2023) 109834

Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont

Liquid chromatography coupled to high-resolution mass spectrometry for nut classification and marker identification

Guillem Campmajó [a,b,*], Javier Saurina [a,b], Oscar Núñez [a,b,c]

[a] Department of Chemical Engineering and Analytical ChemistryJ Universitat de BarcelonaJ Martí i Franquès 1-11J E08028J BarcelonaJ Spain
[b] Research Institute in Food Nutrition and Food SafetyJ Universitat de BarcelonaJ Av. Prat de la Riba 171J Edifici Recerca (Gaudí)J E08921J Santa Coloma de GramenetJ Spain
[c] Serra Húnter Fellow ProgrammeJ Generalitat de CatalunyaJ Via Laietana 2J E08003J BarcelonaJ Spain

ARTICLEINFO

eywords
Nuts
Non-targeted
LC−HRMS
Chemometrics
Adulteration
Food authentication

ABSTRACT

Fraud in nut and seed products poses an economic deception and a threat to human health because of their allergens. This study comprehensively evaluated the metabolomic diversity of ten different nut types through non-targeted liquid chromatography coupled to high-resolution mass spectrometry (LC−HRMS). First, LC−HRMS fingerprints were subjected to partial least squares regression-discriminant analysis (PLS-DA), and the developed multi-class model reached a classification accuracy of 100% after external validation. Then, variable importance in projection (IP) scores obtained from two-input class PLS-DA models (*i.e.*, a specific nut type against all the other samples) allowed the selection of 136 discriminant compounds that were tentatively annotated/identified through HRMS data. Finally, as a case of study, successful detection and quantitation of almond-based products adulteration (with hazelnut or peanut) was achieved through a targeted LC−HRMS study, using some of the found markers and partial least squares (PLS) regression. In this context, new profiling approaches could be further implemented based on the reported markers using cheaper techniques.

Analytical strategies based on *omics* approaches —*genomics*, *proteomics*, *metabolomics*, and *metallomics/isotopolomics*— have been widely proposed to solve food authenticity control. In particular, *metabolomics*, which is the closest omics discipline to the phenotype of biological systems, focuses on the analysis of small molecules (<1500 Da) (Creydt & Fischer, 2018). In this context, the use of *metabolomics* non-targeted methods, where instrumental responses (*i.e.*, mainly analytical signals obtained through chromatography and related techniques, spectroscopy, mass spectrometry, or electronic sensors) are analysed without assuming any previous knowledge, has proved its potential in this field (Medina, Perestrelo, Silva, Pereira, & Câmara, 2019).

Liquid chromatography coupled to high-resolution mass spectrometry (LC–HRMS) is a reliable instrumental platform to perform non-targeted analysis with high molecular coverage of the non-volatile metabolome. Using time-of- ight (TOF) or Orbitrap instruments, HRMS provides high resolving power, allowing accurate *miz* measurements. Furthermore, structural information can be obtained through fragmentation data when using hybrid configurations like quadrupole-Orbitrap (-Orbitrap) or quadrupole-time-of- ight (-TOF). Besides, the hyphenation of LC with HRMS enhances both the selectivity and sensitivity of the analytical approach. Therefore, because of these instrumental capabilities and despite its higher cost and longer analysis time, it is usually preferred over spectroscopic techniques for tentative compound identification. In this line, LC¯HRMS has been widely proposed in diverse applications such as the screening of chemical contaminants in food (Fu, Zhao, Lu, & Xu, 2017) or human biomonitoring (Caballero-Casero et al., 2021), the characterisation of natural plants (Alvarez-Rivera, Ballesteros- ivas, Parada-Alfonso, Ibañez, & Cifuentes, 2019), or clinical research (Rochat, 2016). Particularly in the food fraud field, it has also been used to investigate markers related to specific authentication issues (Lacalle-Bergeron et al., 2021; Zhong et al., 2022).

Nuts and seeds are usually consumed as a snack, although they can also be added to salads, sausages, stews, or bakery products. It is well-known that their regular intake promotes beneficial health effects on humans (Bitok & Sabaté, 2018). However, according to the Food Fraud Risk Information database (Food Fraud Advisors, 2017), some nut-based products are at medium or high risk for fraud practices, such as adulterations or replacements with cheaper and lower-quality ingredients. In

* Corresponding author. Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, Martí i Franquès 1-11, E08028, Barcelona, Spain.
E-mail address gcampmajo@ub.edu (G. Campmajó).

https://doi.org/10.1016/j.foodcont.2023.109834
Received 10 March 2023; Received in revised form 21 April 2023; Accepted 29 April 2023
Available online 29 April 2023

this context, while fraud in raw nuts is unusual due to evident visual differences, it is more common in processed nut products such as ours or pastes, where its detection is more complicated. Moreover, fraud in these products implies an economic deception and a threat to human health because of their allergens (Luparelli et al., 2022), being a food authentication and safety issue. Thus, developing analytical methodologies to detect these practices is required.

Several approaches and analytical techniques have been proposed to detect nut species adulteration. Although some of the methods described in the literature rely on *genomics*, *proteomics*, or *metallomics* (Ding et al., 2020; Esteki, ander Heyden, Farajmand, & Kolahderazi, 2017; Monaci, De Angelis, Bavaro, & Pilolli, 2015), most focus has been on *metabolomics*. In this line, spectroscopic fingerprinting approaches are commonly used for this aim. For instance, Taylan et al. employed Raman spectroscopy to detect green-pea adulteration in pistachio (Taylan et al., 2021), while Rovira et al. evaluated two infrared spectroscopic techniques —near-infrared (NIR) and attenuated total re ection-Fourier transform infrared (ATR-FTIR)— to assess cashew nut authenticity in front of four different adulterants (Rovira et al., 2022).

Instead, to date, most of the targeted and non-targeted methods based on chromatographic and related techniques have focused on the authentication of the cultivar or geographical origin of nut samples (Campmajó & Núñez, 2021; Suman, Cavanna, Sammarco, Lambertini, & Loffi, 2021). Indeed, to our knowledge, few studies using these techniques have dealt with nut species adulteration and classification. For instance, almond powder adulteration with apricot kernel was evaluated using the fatty acid profile obtained with gas chromatography with ame ionisation detection (GC-FID) (Esteki, Farajmand, Kolahderazi, & Simal-Gandara, 2017), while liquid chromatography with ultraviolet (LC-U) and uorescent detection (LC-FLD) fingerprinting assessed nut type classification (Campmajó et al., 2019; Campmajó, Saez- igo, Saurina, & Núñez, 2020). In the latter case, almond-based product adulteration was also studied. Moreover, only the adulteration of pistachio nut powder with green-pea has been evaluated by non-targeted LC–HRMS (<avus, Us, & Güzelsoy, 2018).

Therefore, in this study, non-targeted LC–HRMS was used to address the classification of ten nut-type samples (almonds, cashew nuts, hazelnuts, macadamia nuts, peanuts, pine nuts, pistachios, pumpkin seeds, sun ower seeds, and walnuts) through partial least squares regression-discriminant analysis (PLS-DA). Moreover, two-input class PLS-DA models (*i.e.*, a specific nut type in front of others) were built, and the evaluation of the corresponding variable importance in projection (IP) scores allowed finding the most discriminant molecular features for each nut. After tentatively identifying these markers, targeted LC–HRMS (focusing on the found discriminant molecules) was proposed to detect and quantitate hazelnut and peanut adulterations of almond-based products by partial least squares (PLS) regression.

2 Ma e als a me h s

2.1. Reagents and solutions

Purified water obtained with Elix® 3 coupled to a Milli- ® system (Millipore Corporation, Bedford, MA, USA) and filtered through a 0.22-µm nylon membrane, and acetone and hexane purchased from Merck (Darmstadt, Germany), were used for the sample treatment. Regarding the LC–HRMS analysis, LC–MS grade water and methanol, as well as formic acid (96%), were provided by Merck.

The analytical reagent grade compound standards used for confirmation were: (−)-epicatechin and citric, malic, quinic, and tartaric acids from Merck, and (+)-catechin from Fluka (Steinheim, Germany).

2.2. Instrumentation

The chromatographic system consisted of an ultra-high-performance liquid chromatography (UHPLC) system equipped with an Accela 1250 quaternary pump and an Accela autosampler (Thermo Fisher Scientific, San Jose, CA, USA). The chromatographic separation was performed using a 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), from Phenomenex (Torrance, CA, USA), and using 0.1% (*v/v*) formic acid aqueous solution (solvent A) and methanol (solvent B) as the constituents of the mobile phase. Hence, the developed chromatographic method started with linear gradient elution from 5 to 75% solvent B in 30 min, continued with a 2.5 min-lineal increase up to 95%, and ended with an isocratic step at 95% for 2.5 min. Finally, 0.1 min-lineal decrease back to the initial conditions and 4.9 min of isocratic elution for column re-equilibration were set. The mobile phase ow rate was 400 µL min^{-1}, and the injection volume 10 µL (partial loop mode).

The UHPLC system was coupled to a hybrid -Orbitrap mass spectrometer (-Exactive Orbitrap, Thermo Fisher Scientific) equipped with a heated electrospray ionisation (H-ESI II) source operating in the negative ion mode. Nitrogen with a purity of 99.98%, purchased from Linde (Barcelona, Spain), was used for the ESI sheath, sweep, and auxiliary gas at ow rates of 60, 0, and 10 a.u. (arbitrary units), respectively. Other H-ESI parameters were established as follows: spray voltage, - 2.5 k ; probe heater temperature, 350 °C; capillary temperature, 320 °C; and S-lens RF level at 50 . Full-scan HRMS data were acquired over an *m/z* range of 100−1500 at a mass resolution of 70,000 full width at half maximum (FWHM) at *m/z* 200. In addition, an automatic gain control (AGC) of 1.0×10^6 and a maximum injection time (IT) of 200 ms were established.

For the MS/HRMS experiments, targeted data-dependent scan mode, requiring accurate mass inclusion lists, was used to obtain the product ion scans of specific ions of interest. The acquisition was performed at a mass resolution of 17,500 FWHM at *m/z* 200. Precursor ions, isolated by the quadrupole with an isolation window of 0.5 *m/z*, were fragmented in the higher-energy collisional dissociation (HCD) cell using three-stepped normalised collision energies (NCE) ranging from 10 to 50%. Moreover, the targeted data-dependent acquisition was subordinated to an intensity threshold fixed at 1.0×10^5, and the AGC and IT values were established at 2.0×10^5 and 200 ms, respectively.

The -Orbitrap system was tuned and calibrated every three days, using commercially available calibration solutions for both negative and positive ion modes (Thermo Fisher Scientific). Moreover, the Xcalibur software v 4.1 (Thermo Fisher Scientific) was used to control the LC−HRMS system and acquire and process data.

2.3. Samples

2.3.1. Nut samples for the classification study

A set of 149 raw nut samples bought in Spanish commercial supermarkets were analysed for classification purposes. Samples encompassed various nut classes–30 almonds, 10 cashew nuts, 20 hazelnuts, 10 macadamia nuts, 20 peanuts, 10 pine nuts, 10 pistachios, 20 pumpkin seeds, 9 sun ower seeds, and 10 walnuts—, some of them processed with different thermal treatments —natural, fried, and toasted— (see Table S1 for nut sample details). Before the sample treatment, samples were crushed and homogenised. Moreover, a quality control (C) sample consisting of a mix prepared by pooling 50 µL of each sample extract was employed.

2.3.2. Adulterated almond samples for the quantitative study

Two different almond adulteration scenarios (almond *vs.* hazelnut and almond *vs.* peanut) were considered to evaluate the suitability of the identified biomarkers to address its authentication in two matrices: natural almond our and homemade almond custard cream. Both matrices were obtained from a random almond sample previously used in the classificatory study. Thus, the sample was crushed and homogenised to obtain the almond our (as done in Section 2.3.1), whereas the custard cream was made from hen eggs, milk, sugar, and corn our.

Thus, the adulterants were added in different proportions −0, 20, 40,

60, 80, and 100% for calibration; and 15, 25, 50, 75, and 85% for external validation— to both almond-based products. Five replicates of each blend were prepared, giving 55 samples for each case (adulteration scenario and matrix). Besides, a 50% adulterated sample was used in this case as the corresponding C sample.

2.4. Sample treatment and analysis

A previously developed two-step sample treatment for phytochemical extraction from nut samples was carried out (Campmajó et al., 2019), consisting of ultrasound-assisted solid-liquid extraction (USLE) with acetone:water (70:30, *v*/*v*) followed by a defatting step with hexane.

Regarding the analysis procedure, samples were randomly injected along each sequence to minimise the in uence of any instrumental drift in the chemometric results. Moreover, an extracting solvent blank and a C sample were injected at the beginning and after every ten sample injections to control cross-contamination and avoid systematic errors.

2.5. Data treatment

LC–HRMS generates massive datasets, requiring software programs to properly reduce and handle the obtained data. Thus, aiming to obtain a matrix consisting of ion peak area values as a function of *miz* and retention times, the application of MSConvertGUI from the ProteoWizard Toolkit (Chambers et al., 2012) and mzMine 2.53 (Pluskal, Castillo, illar-Briones, & Orešič, 2010) software was required.

2.5.1. Non-targeted approach

A non-targeted approach was proposed for nut classification. First, raw data were reduced by establishing an absolute intensity threshold peak filter of 5.0×10^5 and transformed to mzXML format through MSConvertGUI. Then, the resulting LC–HRMS data were submitted to the mzMine 2.53 software for peak detection (exact mass detection, chromatogram detection, and chromatogram deconvolution), isotopic peak grouper, and retention time alignment. Brie y, exact mass lists for each scan in a sample were generated, establishing a noise level of 5.0×10^5. Afterwards, the ADAP chromatogram builder allowed the joining of the exact masses found in contiguous scans in a sample that fulfilled the following conditions: peak time range of 0–35 min, 15 minimum scans above an intensity threshold set at 1.0×10^6, and an *miz* tolerance of 5 ppm. Next, individual chromatographic peaks were achieved through chromatogram deconvolution. Thus, for this purpose, the baseline cut-off algorithm was selected for peak recognition with a baseline level of 5.0×10^5, a minimum peak height of 1.0×10^6, and a peak duration range of 0.1–1 min. Subsequently, isotope removal was carried out considering that the most representative isotope was the most intense and setting an *miz* and retention time tolerance of 5 ppm and 0.3 min, respectively. Finally, retention time alignment was carried out using the random sample consensus (RANSAC) peak list aligner method, following the following requisites: retention time tolerance before and after correction of 2.2 and 1.2 min, *miz* tolerance of 5 ppm, 10^5 maximum RANSAC iterations to find the suitable model, and 80% as the minimum value to consider the model valid.

At the end of this work ow, a data matrix was exported to an Excel File containing the obtained LC–HRMS fingerprints: samples × variables. The samples' column included the 149 nut samples and the 16 C samples, whereas the variables' row comprised all the detected molecular features (an exact *miz* value at a specific retention time). Only the molecular features detected at least in 80% of the samples belonging to a nut class were selected to reduce the matrix dimensions. As a result, a 165 ×278 dimension data matrix was obtained, containing the chromatographic peak areas for each molecular feature in all samples.

2.5.2. Targeted approach

A targeted approach was applied to detect and quantify the adulteration in almond-based products, focusing on the discriminant markers identified for almond, hazelnut, and peanut samples encountered in the classificatory study.

First, raw data were processed with MSConvertGUI, applying an absolute intensity threshold peak filter of 1.0×10^5. Then, targeted peak detection was performed with mzMine 2.53 software using a list of targeted molecular features (an exact *miz* value at a specific retention time) for each almond adulteration scenario: 28 for almond *vs.* hazelnut and 35 for almond *vs.* peanut. Besides, a noise level of 1.0×10^5, an intensity tolerance of 10% (maximum allowed deviation from the expected shape of a chromatographic peak), an *miz* tolerance of 5 ppm, and a retention time tolerance of 0.5 min were established. Finally, in this case, the join aligner allowed matching of the detected molecular features across samples, setting a mass tolerance of 5 ppm, a retention time tolerance of 0.5 min, 80% of weight for *miz*, and 20% of weight for retention time. Again, at the end of the work ow, a data matrix was exported to an Excel File containing the obtained LC–HRMS profiles.

2.6. Chemometric and statistical analysis

The chemometric analysis by principal component analysis (PCA), PLS-DA, and PLS regression was carried out using Solo 8.6 chemometric software from Eigenvector Research (Manson, WA, USA). Details of their theoretical background are addressed elsewhere (Massart et al., 1997).

LC- HRMS data matrices (normalised and autoscaled) were used as X-data matrices indistinctly of the chemometric method used, which depended on the aim of the study. In this line, PCA assessed a first exploratory analysis to check the absence of systematic errors through C sample behaviour and allowed visualising sample trends. PLS-DA, particularly PLS1-DA (Brereton & Lloyd, 2014), was used in the classificatory study, requiring a Y-data matrix that defined the nut type of each sample. PLS was employed in the quantitation of almond-based product adulteration, demanding a Y-data matrix that expounded sample adulteration percentages. Moreover, the proper number of latent variables (L s) for building PLS-DA and PLS models was selected, after enetian blinds cross-validation (C), at the first minimum of the cross-validation classification error (C CE) and the root-mean-square error of cross-validation (RMSEC), respectively.

PLS-DA and PLS models' performance was checked by external validation. On the one hand, for the classificatory study, samples were stratified and randomly chosen: 60% were used as the calibration set, whereas the remaining 40% as the external validation set. Then, overall accuracy and each class sensitivity (capability to detect true positives) and specificity (capability to detect true negatives) were used to evaluate the classification models. While the former is calculated by dividing the number of well-classified samples by the total number of samples, class sensitivity (Eq. (1)) and specificity (Eq. (2)) are calculated as follows (Riedl, Esslinger, & Fauhl-Hassek, 2015):

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (1)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (2)$$

where TP is true positive samples, TN is true negative samples, FP is false positive samples, and FN is false negative samples.

On the other hand, PLS regression was done using some adulteration percentages for calibration and others for external validation, as detailed in Section 2.3.2. The model performance was evaluated through the root-mean-square error of calibration (RMSEC), RMSEC , and prediction (RMSEP), as well as the corresponding R^2 (determination coefficient) values. Eq. (3) shows how RMSEs are calculated. Moreover, relative error in each external validation to estimate the adulterant percentage was also assessed.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \quad (3)$$

where $\hat{y}_i$ is the predicted value, y_i is the actual value, n is the number of samples, and N is the number of predictions.

Finally, to find the most discriminant molecular features for each nut class, individual PLS-DA models of a specific nut class against all the others were built. Then, after external validation of the models, variables with the highest IP scores were selected for further annotation and identification steps. Regression vector coefficients were also evaluated, with positive values indicating that variable contribution is related to the target class. Moreover, the significance of the differences in their peak area values between nut classes was evaluated statistically. Thus, after a Fisher test of variances, the student *t*-test for comparing the means of two classes was carried out. A confidence level of 0.99 was assumed, so when *p* (probability) values were lower than 0.01, differences in the molecular feature peak areas between the classes were considered significant.

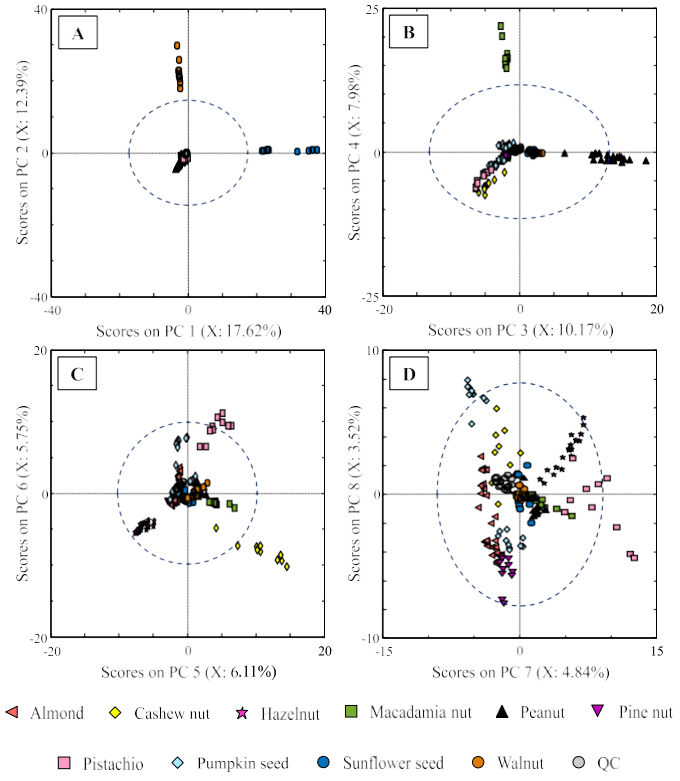
2.7. Annotation and identification of the most discriminant compounds

The most discriminant molecular features for each nut class, selected through IP scores, were putatively identified following Schymanski et al. HRMS identification levels (Schymanski et al., 2014). The established parameters to assess this identification step were: 5 ppm of exact mass tolerance, >85% of isotopic pattern fit, MS^2 data similarity, and retention time agreement. For the MS^2 comparison, public databases such as mzCloud (HighChem LLC, Bratislava, Slovakia), The Human Metabolome Database (Wishart et al., 2018), and LIPID MAPS Structure Database (Sud et al., 2006) were employed. Besides, Phenol-Explorer (Rothwell et al., 2013), a database including polyphenolic content in food, was also consulted. Finally, in some specific cases, MetFrag software (Wolf, Schmidt, Müller-Hannemann, & Neumann, 2010) was also used for tentative in-silico elucidation.

3 Res l s a s ss

3.1. Non-targeted LC–HRMS nut classification

As previously mentioned, the present study aimed to develop a non-targeted LC–HRMS method to classify nut samples according to their type and identify the most discriminant molecular features. Despite the non-targeted nature of the developed method, instrumental conditions were oriented to favour phenolic and polyphenolic compound detection since they have already been successfully proposed as potential markers in several food authentication issues (Lucci, Saurina, & Núñez, 2017; Proestos & Pesic, 2022). Thus, a total of 149 nut samples belonging to 10

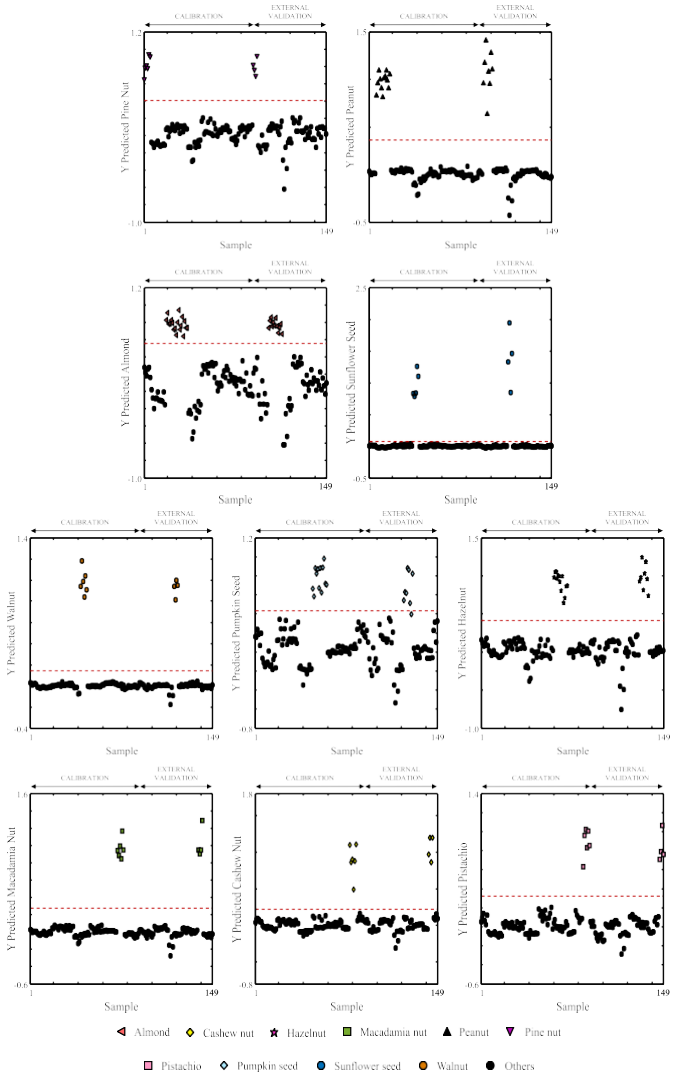


F g PCA scores plot obtained for the analysed nut samples according to their type, using the non-targeted LC–HRMS data, of (A) PC1 *vs*. PC2, (B) PC3 *vs*. PC4, (C) PC5 *vs*. PC6, and (D) PC7 *vs*. PC8.

different nut classes were analysed following the proposed method. As an example, Fig. S1 depicts the total ion current (TIC) LC–HRMS chromatogram for a selected sample within each nut type. In this context, remarkable qualitative differences regarding peak distribution and signal intensity can be visually detected.

LC–HRMS data were subjected to PCA to appraise their discriminating capability. However, PCA was first employed to select the most appropriate data treatment, which is crucial for subsequent unequivocal results. In this case, normalisation (scaling each sample to the sum of the corresponding peak areas) and autoscaling (mean centring and variable scaling to unit standard deviation) were assessed to try to improve the data quality. As a result, it was found that performing a normalisation step before autoscaling provided a better sample grouping, reducing the effect of the HRMS detection variance.

In this context, Fig. 1 shows the PCA scores plot obtained after applying this data pretreatment and using the 165×278 dimension data matrix containing both nut and C samples. A total of eight principal components (PCs), describing 68.37% of the variance, were chosen for the PCA analysis. As a result, the non-supervised chemometric plots showed C samples grouped in the centre, indicating the lack of



F g 2 Sample *vs*. Y Predicted plot for the two-input class PLS-DA models after external validation.

systematic errors affecting the reliability of the results. Moreover, the complexity of the studied issue —encompassing a significant number of sample classes (nut type) and factors (different geographical origins and thermal treatments)— was re ected in the low variance explained by the PCs (*e.g.*, 17.62% for PC1). Nevertheless, good sample distinction was achieved for almost all nut types. For instance, in Fig. 1A, where the plot of scores of PC1 *vs*. PC2 is depicted, sun ower seed and walnut samples were visibly separated through PC1 and PC2, respectively. PC3 and PC4 were highly related to peanut and macadamia nut samples, respectively (see Fig. 1B). Besides, the plot of scores of PC5 *vs*. PC6 (Fig. 1C) allowed the discrimination of cashew nut (displaying positive PC5 and negative PC6 values), hazelnut (displaying negative PC5 and PC6 values), and pistachio (displaying positive PC5 and PC6 values) samples. Finally, the scatter plot for scores of PC7 and PC8 (Fig. 1D) allowed a slight discrimination of pine nut and almond samples along the PC8, presenting negative PC7 values.

Given the excellent results observed in the PCA, with a remarkable separation of samples according to nut classes (except for pumpkin seed samples), PLS-DA was applied to the non-targeted LC–HRMS data. For this, C samples were removed from the dataset, and a 142 × 278 dimension data matrix was subjected to the supervised classificatory analysis. In this case, nine L s were selected to build the PLS-DA model, which described 71.35% and 95.53% of X-variance and Y-variance, respectively. As a result, visual sample classification was reached for all nut types investigated.

Therefore, to evaluate the classificatory ability of the non-targeted LC–HRMS data through PLS-DA, external validation was performed as described in Section 2.6. In this line, a PLS-DA calibration model, built with 60% of the analysed samples, was composed of nine L s explaining 73.53% of X-variance and 95.62% of Y-variance. C results —sensitivities of 100%, specificities above 97.6%, and classification accuracies above 98.8%, for each nut class under study— anticipated the excellent results obtained in the external validation —sensitivities, specificities, and classification accuracies of 100%, for each nut class under study—, proving the excellent discriminant capacity of the non-targeted LC–HRMS data. The external validation graphical results (Sample *vs*. Y predicted score plot) for each analysed nut type are shown in Fig. S2.

3.2. Annotation and identification of nut type markers

As previously mentioned, one of this study's main goals was to identify characteristic discriminant molecular features for each studied nut type. Thus, with this purpose, two-input class PLS-DA models were built: the first input corresponded to a specific nut class, while the second encompassed all the others. In this context, as shown in Fig. 2, the performance of each binary PLS-DA model was assessed through external validation, obtaining complete sample classification in all the cases, except for the Pumpkin seed *vs*. Others model, where 99.3% classification accuracy, 87.5% sensitivity, and 100% specificity were obtained.

Then, in each PLS-DA model, IP loadings scores allowed the selection of the most discriminant molecular features for each nut under study. For instance, Fig. S3 presents the results obtained for the PLS-DA model of Walnut *vs*. Others: the classification plot depicting Sample *vs*. Y Predicted Walnut and the corresponding IP scores plot. To obtain MS/HRMS data of the selected discriminant molecular features, an arbitrary sample of each nut type was analysed by LC–HRMS using a targeted data-dependent acquisition method, built with an inclusion list containing them.

Table S2 summarises the tentative annotation and identification of the discriminant markers found for each nut type. Most of the compounds were detected in their deprotonated form $[M-H]^-$, as expected considering that the HRMS acquisition was performed in the negative mode, although in some cases, their adduct with formic acid $[M+FA-H]^-$ or chlorine $[M+Cl]^-$, or even their deprotonated dimeric form $[2M-H]^-$ corresponded to the base peak. Some of the annotated/identified compounds are discussed below since some of them had been previously reported in the literature.

In the case of almond discriminant compounds, amygdalin —a cyanogenic diglucoside responsible for the bitterness of almonds— and amygdaloside were found (Lee, Zhang, Wood, Rogel Castillo, & Mitchell, 2013; Sang et al., 2003). Furthermore, sugars and derivatives such as *miz* 341.1083 and 683.2243, annotated as disaccharide and tetrasaccharide ‡H$_2$O, respectively, were also detected (Gil Solsona, Boix, Ibáñez, & Sancho, 2018; Huang, Robinson, Dias, de Moura Bell, & Barile, 2022). Besides, Gil-Solsona et al. previously identified the first, annotating it as inulobiose, as a discriminant marker related to the Spanish almond variety.

Among the molecular features presenting high IP scores in the cashew nut PLS-DA classification, the isomers with the molecular formula $C_{15}H_{14}O_6$, observed at the retention times of 12.80 and 15.50 min, were identified as the avanols (+)-catechin and (–)-epicatechin that have been previously seen in cashew nut testa (Trox et al., 2011).

Several indoleacetic acid glycoside isomers were found to be discriminant markers for hazelnut classification. In this line, *miz* 541.1458 (at a retention time of 20.91 min) and 540.1719 (at a retention time of 21.05 min) were assigned as isomers of 2-(3-hydroxy-2-oxoindolin-3-yl) acetic acid 3-O-6′-galactopyranosyl-2″-(2‴oxoindolin-3‴yl) and hazelnutin D, respectively, which have been previously detected in hazelnut kernel (Shataer et al., 2021). Besides, two other indoleacetic acid glycoside chiral isomers with *miz* 368.0984 —named 3-(O-β-D-glycosyl)dioxindole-3-acetic acid and hazelnutin E by Singl- dinger et al. (Singldinger et al., 2018) and Shataer et al. (Shataer et al., 2021), respectively— were found to be discriminant. These compounds presented chromatographic peaks at 7.31 and 8.39 min, although their retention time assignment was not possible.

Regarding macadamia nut markers, among others, various phenolic and polyphenolic compounds such as phenolic acid derivatives (hydroxybenzoic acid glucoside and apiosylglucosyl 4-hydroxybenzoate isomers), guaiacol hexose-pentose isomers, and oleoside dimethyl ester were tentatively identified.

In the case of peanut, several hydroxycinnamic acids (*i.e.*, *cis*- and *trans*-*p*-coumaroyl tartaric acids, feruloyl tartaric acid isomer, *p*-coumaric acid, coumaroyl-O-pentoside isomer, di-*p*-coumaroyl tartaric acid isomer, and *p*-coumaroylferuloyl tartaric acid isomer) and a derivative (such as *p*-coumaroylnicotinoyl tartaric acid), as well as an hydroxybenzoic acid isomer, were detected agreeing with literature and considered as discriminant (Ma et al., 2014).

In relation to pine nut molecular features, ascorbalamic acid isomer ($C_9H_{13}NO_8$) and vanillic acid glucoside isomers ($C_{14}H_{18}O_9$) were annotated as relevant for pine nut classification.

Moreover, accordingly to previous studies (Ers, an, Güclü-Üstündag, Carle, & Schweiggert, 2016), protocatechuic acid and quercetin 3-O-glucoside were observed in their deprotonated form at 10.44 and 23.03 min, respectively, providing high IP scores for pistachio classification. Organic acids, such as malic and isocitric acid, and nucleotides, as uridine monophosphate and adenosine 5′-monophosphate, also appeared to be discriminant for pistachio.

In the case of pumpkin seed, sugars —a trisaccharide in its $[M+Cl]^-$ form and sedoheptulose—, guanosine, and tyrosol diglycoside isomer, were tentatively identified, among others.

Finally, regarding sun ower seed discriminant compounds, as reported by Romani et al. (Romani, Pinelli, Moschini, & Heimler, 2017), several hydroxycinnamic acids were found and related to sun ower seed classification: 4-caffeoylquinic acid, 5-caffeoylquinic acid, 3,4-dicaffeoylquinic acid, and 4,5-dicaffeoylquinic acid. Besides, other phenolic compounds, such as avonoid O-glycoside isomers and phenylacetic acid, were also annotated.

Instead, among the discriminant compounds presenting high IP scores for walnut classification, several hydroxybenzoic acids and derivatives —galloyl-hexahydroxydiphenoyl-glucose isomer, digalloyl-

Table
Calibration, cross-validation, and external validation results obtained for each of the PLS regression models used to determine the almond products adulteration percentage.

Adulteration case		Matrix	Data matrix (samples × variables)	CALIBRATION			CROSS-ALIDATION		EXTERNAL ALIDATION						
				L s	RMSEC (%)	R^2	RMSEC (%)	R^2	RMSEP (%)	R^2	Relative error in each adulterant percentage (%)				
											15%	25%	50%	75%	85%
Im	**A EL**	Flour	55 × 28	1	5.994	0.969	6.304	0.966	5.037	0.977	6.06	4.93	2.70	3.54	1.88
		Cream	55 × 28	2	7.773	0.946	10.154	0.909	8.887	0.933	6.07	7.14	9.24	7.77	7.55
Im	**ea**	Flour	55 × 35	2	3.865	0.986	6.665	0.960	8.202	0.950	4.30	0.97	8.62	10.86	5.86
		Cream	55 × 35	3	5.237	0.977	8.648	0.937	9.852	0.893	12.89	8.54	8.15	8.20	5.61

L : latent variable; R^2: determination coefficient; RMSEC: root-mean-square error of calibration; RMSEC : root-mean-square error of cross-validation; RMSEP: root-mean-square error of prediction.

hexahydroxydiphenoyl-glucose isomer, ellagic acid, and ellagic acid pentoside— and a saccharolipid as glansreginin A were found, in accordance to Regueiro et al. (Regueiro et al., 2014).

It should be mentioned that for a molecular feature to be discriminant, it is not necessary to be detected or not detected in an exclusive nut type. Sometimes it presents a higher or lower content compared to other matrices. Moreover, considering the complexity of the studied issue, where ten nut types were studied, some markers were discriminant just in front of some others. Therefore, to better understand the significance of the differences of each discriminant marker content in their corresponding nut type compared to each other, student *t*-tests (preceded by a Fisher test of variances) were performed. Thus, Table S3 summarises the obtained results. As an example, focusing on hazelnut markers, *miz* 131.0462 (at a retention time of 2.27 min), tentatively identified as asparagine, showed lower content in hazelnut than in almond, cashew nut, macadamia nut, and peanut, whereas no differences were observed with the remaining types. Instead, *miz* 368.0984 (at a retention time of 7.31 min), tentatively identified as a 3-(O-β-D-glycosyl)dioxindole-3-acetic acid isomer, presented higher values in hazelnut than in any other nut.

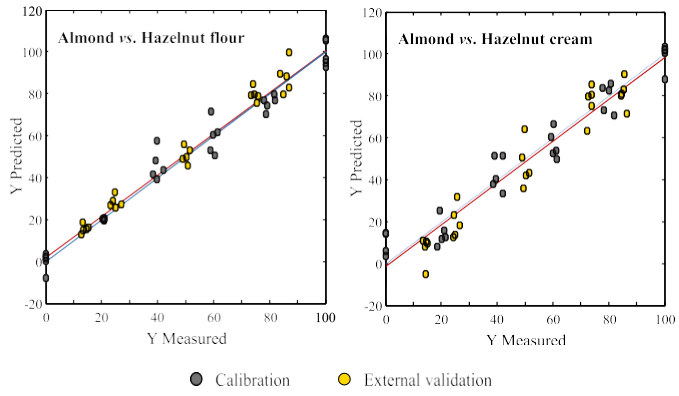
In addition, since some of the analysed nut types presented different processing treatments —natural, fried, and toasted for almond; natural and toasted for hazelnut and pumpkin seed; and fried and toasted for peanut—, differences in the identified markers because of the processing were studied. In this context, Table S4 presents the results obtained after performing univariate statistical analysis to evaluate the significance of differences. In general, differences were not significant except for some given cases (*e.g.*, quinic acid content was significantly higher in toasted almonds than natural and fried ones). Besides, differences between nut matrices prevailed over processing treatment ones.

3.3. Detection and quantitation of almond-based product adulterations through targeted LC–HRMS

To validate the applicability of the identified molecular features as a discriminant profile for nut authentication, the adulteration of almond-based products (natural almond our and homemade almond custard cream) was evaluated through PLS regression. These products are at medium risk of adulteration with cheaper nuts (Food Fraud Advisors, 2017). Therefore, hazelnut and peanut were chosen as adulterants due to the difficulty of visually detecting them (*i.e.*, mainly due to physical similarities such as granulometry or colour) in the studied matrices and their lower price. Furthermore, peanut was especially selected because of their serious threat to food safety (*i.e.*, they can cause severe allergy episodes).

Hence, as detailed in Section 2.3.2, different blend percentages were prepared and analysed following the developed LC–HRMS method. Afterwards, LC–HRMS data were processed using the targeted approach described in Section 2.5.2, focusing on the discriminant markers identified in the supervised study (Table S2): 12, 16, and 23 molecular features for almond, hazelnut, and peanut, respectively. Therefore, 28 molecular features were monitored for the almond *vs*. hazelnut adulteration scenario, while 35 were for the almond *vs*. peanut one (the corresponding feature lists file is provided in the Supplementary Material as a CS file).

Before PLS regression analysis, PCA was performed using the



F g 3 External validation PLS results for the prediction of the percentage of adulteration of almond our (on the left side) and almond custard cream (on the right side) with hazelnut. The blue line corresponds to the theoretical diagonal line, while the red line to the experimental adjusted one.

obtained targeted LC–HRMS data to observe non-supervised sample clustering and trends, as well as to check C sample behaviour. For instance, Fig. S4 illustrates the scatter plot for scores of PC1 and PC2 (describing 70.8% of the variance) for the almond our adulteration with peanut case. As observed, C samples, which corresponded to a 50% adulterated sample, appeared in the centre of the plot, ensuring a good instrumental performance. Moreover, pure almond and peanut samples were distributed on opposite sides of the plot, displaying negative and positive PC1 scores, respectively. Intuitively, adulterated samples were ordered according to their adulterant percentage from the left (low adulterant percentages) to the right (high adulterant percentages) of the plot.

Then, PLS regression was applied to quantitate the adulteration level in each case under study. Thus, complementary to the X-data matrix, a Y-data matrix specifying the blend degree was required. Table 1 summarises the original data matrices used, the number of L s employed to build each calibration PLS regression model, and the results obtained for the calibration, C , and external validation. Good calibration models were built with low RMSEC≤(7.773%) and R²≥0.946. Moreover, the similarity between calibration and C parameter values indicated good internal consistency, preventing overfitting in the subsequent external validation. Regarding the external validation results, RMSEP and R^2 values (≤9.852% and ≥0.893, respectively) indicated that the built PLS regression models showed a satisfactory ability to detect and quantitate almond adulterations. As an example, Fig. 3 shows the external validation PLS results for adulterating almond-based products with hazelnut (see Fig. S5 for the PLS results when adulterating with peanut). Results indicated that although more accurate quantitation was obtained of the almond our matrix, no significant differences were observed between the studied matrices.

C l s s

This study applied LC–HRMS, combined with chemometrics, to analyse nut product samples. In this line, 149 samples belonging to 10 nut types were analysed through non-targeted LC–HRMS aiming to find markers to prevent nut fraud (*e.g.*, adulteration, substitution, or replacement). PLS-DA allowed complete sample classification by the developed multi-class model (classification accuracy of 100% after external validation) and the identification of the most discriminant markers for each type of nut. In this regard, 136 molecular features were tentatively annotated/identified, taking benefit of the power of the MS/HRMS detection (*i.e.*, high sensitivity and selectivity, leading to good molecular coverage). For instance, organic acids, phenolic compounds, sugars, amino acids, and some derivatives were found among the compounds identified. Besides, although some of these markers' content varied due to thermal processing (*i.e.*, natural, toasted, or fried), differences between nut matrices prevailed. Moreover, to validate the use of the found markers for nut authentication, the adulteration of almond-based products (almond our and homemade custard cream) with hazelnut or peanut was addressed. Thus, after targeted LC–HRMS analysis focusing on the corresponding markers, the obtained PLS results demonstrated their applicability to detect and quantitate the blend percentage.

Therefore, this study provides a set of nut and seed markers that could be further used in developing profiling approaches (for instance, using low-resolution mass spectrometers), which are more established in routine analysis, to detect adulteration in processed products.

F g

This research was supported by the project PID2020-114401RB-C22 financed by the Agencia Estatal de Investigación (AEI/10.13039/501100011033) and by the Agency for Administration of University and Research Grants (Generalitat de Catalunya, Spain) under the project 2021SGR-00365.

CRe T a h sh p b s a eme

G llem Campmaj ´: Conceptualization, Methodology, alidation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Jav e Sa a:** Conceptualization, Writing – review & editing, Supervision, Funding acquisition. **Os a Nú ˜E :** Conceptualization, Writing – review & editing, Supervision, Funding acquisition.

De la a f mpe g e es

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to in uence the work reported in this paper.

Da a ava lab l y

Data will be made available on request.

k wle geme s

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

ppe x S ppleme a y a a

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodcont.2023.109834.

Refe e es

Alvarez-Rivera, G., Ballesteros- ivas, D., Parada-Alfonso, F., Ibañez, E., & Cifuentes, A. (2019). Recent applications of high resolution mass spectrometry for the characterization of plant natural products. *TrACJ Trends in Analytical ChemistryJ 112*, 87–101. https://doi.org/10.1016/j.trac.2019.01.002

Bitok, E., & Sabaté, J. (2018). Nuts and cardiovascular disease. *Progress in Cardiovascular DiseasesJ 61*, 33–37. https://doi.org/10.1016/j.pcad.2018.05.003

Brereton, R. G., & Lloyd, G. R. (2014). Partial least squares discriminant analysis: Taking the magic away. *Journal of ChemometricsJ 28*, 213–225. https://doi.org/10.1002/cem.2609

Caballero-Casero, N., Belova, L., ervliet, P., Antignac, J. P., Castaño, A., Debrauwer, L., et al. (2021). Towards harmonised criteria in quality assurance and quality control of suspect and non-target LC-HRMS analytical work ows for screening of emerging contaminants in human biomonitoring. *TrACJ Trends in Analytical ChemistryJ 136*, Article 116201. https://doi.org/10.1016/j.trac.2021.116201

Campmajó, G., Navarro, G. J., Núñez, N., Puignou, L., Saurina, J., & Núñez, O. (2019). Non-targeted HPLC-U fingerprinting as chemical descriptors for the classification and authentication of nuts by multivariate chemometric methods. *SensorsJ 19*, 1388. https://doi.org/10.3390/s19061388

Campmajó, G., & Núñez, O. (2021). Authentication of nut and seed products. In G. Campmajó, & O. Núñez (Eds.), *Chromatographic and related separation techniques in food integrity and authenticityJ B* pp. 129–155). https://doi.org/10.1142/9781786349972_0005

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

<avus, F., Us, M. F., & Güzelsoy, N. A. (2018). Assessing pistachio nut (Pistacia vera L.) adulteration with green pea (Pisum sativum L.) by untargeted liquid chromatography-(quadrupole-time of ight)-mass spectrometry method and chemometrics. *Journal of Food and Feed Science - TechnologyJ 19*, 34–41.

Chambers, M. C., MacLean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., et al. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nature BiotechnologyJ 30*, 918–920. https://doi.org/10.1038/nbt.2377

Creydt, M., & Fischer, M. (2018). Omics approaches for food authentication. *ElectrophoresisJ 39*, 1569–1581. https://doi.org/10.1002/elps.201800004

Ding, Y., Jiang, G., Huang, L., Chen, C., Sun, J., & Zhu, C. (2020). DNA barcoding coupled with high-resolution melting analysis for nut species and walnut milk beverage authentication. *Journal of the Science of Food and AgricultureJ 100*, 2372–2379. https://doi.org/10.1002/jsfa.10241

Erşan, S., Güclü Üstündag, Ö., Carle, R., & Schweiggert, R. M. (2016). Identification of phenolic compounds in red and green pistachio (Pistacia vera l.) hulls (exo- and mesocarp) by HPLC-DAD-ESI-(HR)-MSn. *Journal of Agricultural and Food ChemistryJ 64*, 5334–5344. https://doi.org/10.1021/acs.jafc.6b01745

Esteki, M., Farajmand, B., Kolahderazi, Y., & Simal-Gandara, J. (2017). Chromatographic fingerprinting with multivariate data analysis for detection and quantification of

apricot kernel in almond powder. *Food Analytical MethodsJ 10*, 3312−3320. https://doi.org/10.1007/s12161-017-0903-5

Esteki, M., ander Heyden, Y., Farajmand, B., & Kolahderazi, Y. (2017). ualitative and quantitative analysis of peanut adulteration in almond powder samples using multi-elemental fingerprinting combined with multivariate data analysis methods. *Food ControlJ 82*, 31−41. https://doi.org/10.1016/j.foodcont.2017.06.014

Food Fraud Advisors. (2017). *Food fraud risk information*. Retrieved October 10, 2020, from https://trello.com/b/aoFO1UEf/food-fraud-risk-information.

Fu, Y., Zhao, C., Lu, X., & Xu, G. (2017). Nontargeted screening of chemical contaminants and illegal additives in food based on liquid chromatography−high resolution mass spectrometry. *TrACJ Trends in Analytical ChemistryJ 96*, 89−98. https://doi.org/10.1016/j.trac.2017.07.014

Gil Solsona, R., Boix, C., Ibáñez, M., & Sancho, J. . (2018). The classification of almonds (Prunus dulcis) by country and variety using UHPLC-HRMS-based untargeted metabolomics. *Food Additives & Contaminants Part A ChemistryJ AnalysisJ ControlJ Exposure and Risk AssessmentJ 35*, 395−403. https://doi.org/10.1080/19440049.2017.1416679

Huang, Y. P., Robinson, R. C., Dias, F. F. G., de Moura Bell, J. M. L. N., & Barile, D. (2022). Solid-phase extraction approaches for improving oligosaccharide and small peptide identification with liquid chromatography-high-resolution mass spectrometry: A case study on proteolyzed almond extract. *FoodsJ 11*, 4. https://doi.org/10.3390/foods11030340

Lacalle-Bergeron, L., Izquierdo-Sandoval, D., Sancho, J. ., López, F. J., Hernández, F., & Portolés, T. (2021). Chromatography hyphenated to high resolution mass spectrometry in untargeted metabolomics for investigation of food (bio)markers. *TrACJ Trends in Analytical ChemistryJ 135*, Article 116161. https://doi.org/10.1016/j.trac.2020.116161

Lee, J., Zhang, G., Wood, E., Rogel Castillo, C., & Mitchell, A. E. (2013). uantification of amygdalin in nonbitter, semibitter, and bitter almonds (prunus dulcis) by UHPLC-(ESI) q MS/MS. *Journal of Agricultural and Food ChemistryJ 61*, 7754−7759. https://doi.org/10.1021/jf402295u

Lucci, P., Saurina, J., & Núñez, O. (2017). Trends in LC-MS and LC-HRMS analysis and characterization of polyphenols in food. *TrACJ Trends in Analytical ChemistryJ 88*, 1−24. https://doi.org/10.1016/j.trac.2016.12.006

Luparelli, A., Losito, I., De Angelis, E., Pilolli, R., Lambertini, F., & Monaci, L. (2022). Tree nuts and peanuts as a source of beneficial compounds and a threat for allergic consumers: Overview on methods for their detection in complex food products. *FoodsJ 11*, 728. https://doi.org/10.3390/foods11050728

Ma, Y., Kosińska-Cagnazzo, A., Kerr, W. L., Amarowicz, R., Swanson, R. B., & Pegg, R. B. (2014). Separation and characterization of phenolic compounds from dry-blanched peanut skins by liquid chromatography-electrospray ionization mass spectrometry. *Journal of Chromatography AJ 1356*, 64−81. https://doi.org/10.1016/j.chroma.2014.06.027

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

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 TechnologyJ 85*, 163−176. https://doi.org/10.1016/j.tifs.2019.01.017

Monaci, L., De Angelis, E., Bavaro, S. L., & Pilolli, R. (2015). High-resolution Orbitrap™-based mass spectrometry for rapid detection of peanuts in nuts. *Food Additives & Contaminants Part A ChemistryJ AnalysisJ ControlJ Exposure and Risk AssessmentJ 32*, 1607−1616. https://doi.org/10.1080/19440049.2015.1070235

Pluskal, T., Castillo, S., illar-Briones, A., & Orešič, M. (2010). MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC BioinformaticsJ 11*, 395. https://doi.org/10.1186/1471-2105-11-395

Proestos, C., & Pesic, M. (2022). Editorial: Polyphenols in food authenticity through the application of advanced omics technologies. *Frontiers in NutritionJ 9*, Article 998635. https://doi.org/10.3389/fnut.2022.998635

Regueiro, J., Sánchez-González, C., allverdú- ueralt, A., Simal-Gándara, J., Lamuela-Raventós, R., & Izquierdo-Pulido, M. (2014). Comprehensive identification of walnut polyphenols by liquid chromatography coupled to linear ion trap-Orbitrap mass spectrometry. *Food ChemistryJ 152*, 340−348. https://doi.org/10.1016/j.foodchem.2013.11.158

Riedl, J., Esslinger, S., & Fauhl-Hassek, C. (2015). Review of validation and reporting of non-targeted fingerprinting approaches for food authentication. *Analytica Chimica ActaJ 885*, 17−32. https://doi.org/10.1016/j.aca.2015.06.003

Rochat, B. (2016). From targeted quantification to untargeted metabolomics: Why LC-high-resolution-MS will become a key instrument in clinical labs. *TrACJ Trends in Analytical ChemistryJ 84*, 151−164. https://doi.org/10.1016/j.trac.2016.02.009

Romani, A., Pinelli, P., Moschini, ., & Heimler, D. (2017). Seeds and oil polyphenol content of sun ower (Helianthus annuus l.) grown with different agricultural management. *Advances in Horticultural ScienceJ 31*, 85−88. https://doi.org/10.13128/ahs-20608

Rothwell, J. A., Perez-Jimenez, J., Neveu, ., Medina-Remón, A., M'Hiri, N., García-Lobato, P., et al. (2013). *Phenol-explorer 3.0 A ma or update of the phenol-explorer database to incorporate data on the effects of food processing on polyphenol content*. Database. https://doi.org/10.1093/database/bat070, 2013, bat070.

Rovira, G., Miaw, C. S. W., Martins, M. L. C., Sena, M. M., de Souza, S. . C., Ruisánchez, I., et al. (2022). In-depth chemometric strategy to detect up to four adulterants in cashew nuts by IR spectroscopic techniques. *Microchemical JournalJ 181*, Article 107816. https://doi.org/10.1016/j.microc.2022.107816

Sang, S., Li, G., Tian, S., Lapsley, K., Stark, R. E., Pandey, R. K., et al. (2003). An unusual diterpene glycoside from the nuts of almond (Prunus amygdalus Batsch). *Tetrahedron LettersJ 44*, 1199−1202. https://doi.org/10.1016/S0040-4039(02)02794-6

Schymanski, E. L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H. P., et al. (2014). Identifying small molecules via high resolution mass spectrometry: Communicating confidence. *Environmental Science and TechnologyJ 48*, 2097−2098. https://doi.org/10.1021/es5002105

Shataer, D., Li, J., Duan, X. M., Liu, L., Xin, X. L., & Aisa, H. A. (2021). Chemical composition of the hazelnut kernel (corylus avellana L.) and its anti-in ammatory, antimicrobial, and antioxidant activities. *Journal of Agricultural and Food ChemistryJ 69*, 4111−4119. https://doi.org/10.1021/acs.jafc.1c00297

Singldinger, B., Dunkel, A., Bahmann, D., Bahmann, C., Kadow, D., Bisping, B., et al. (2018). New taste-active 3-(O-β- d -Glucosyl)-2-oxoindole-3-acetic acids and diarylheptanoids in cimiciato-infected hazelnuts. *Journal of Agricultural and Food ChemistryJ 66*, 4662−4673. https://doi.org/10.1021/acs.jafc.8b01216

Sud, M., Fahy, E., Cotter, D., Brown, A., Dennis, E. A., Glass, C. K., et al. (2006). LMSD: LIPID MAPS structure database. *Nucleic Acids ResearchJ 35*, 527−532. https://doi.org/10.1093/nar/gkl838

Suman, M., Cavanna, D., Sammarco, G., Lambertini, F., & Loffi, C. (2021). Fighting food frauds exploiting chromatography-mass spectrometry technologies: Scenario comparison between solutions in scientific literature and real approaches in place in industrial facilities. *TrACJ Trends in Analytical ChemistryJ 142*, Article 116305. https://doi.org/10.1016/j.trac.2021.116305

Taylan, O., Cebi, N., Yilmaz, M. T., Sagdic, O., Ozdemir, D., & Balubaid, M. (2021). Rapid detection of green-pea adulteration in pistachio nuts using Raman spectroscopy and chemometrics. *Journal of the Science of Food and AgriculturеJ 101*, 1699−1708. https://doi.org/10.1002/jsfa.10845

Trox, J., adivel, ., etter, W., Stuetz, W., Kammerer, D. R., Carle, R., et al. (2011). Catechin and epicatechin in testa and their association with bioactive compounds in kernels of cashew nut (Anacardium occidentale L.). *Food ChemistryJ 128*, 1094−1099. https://doi.org/10.1016/j.foodchem.2011.04.018

Wishart, D. S., Feunang, Y. D., Marcu, A., Guo, A. C., Liang, K., ázquez-Fresno, R., et al. (2018). HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids ResearchJ 46*, D608−D617. https://doi.org/10.1093/nar/gkx1089

Wolf, S., Schmidt, S., Müller-Hannemann, M., & Neumann, S. (2010). In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC BioinformaticsJ 11*, 148. https://doi.org/10.1186/1471-2105-11-148

Zhong, P., Wei, X., Li, X., Wei, X., Wu, S., Huang, W., et al. (2022). Untargeted metabolomics by liquid chromatography-mass spectrometry for food authentication: A review. *Comprehensive Reviews in Food Science and Food SafetyJ 21*, 2455−2488. https://doi.org/10.1111/1541-4337.12938

Supplementary Material

Liquid chromatography coupled to high-resolution mass spectrometry for nut classification and marker identification

Guillem Campmajó[a,b,*], Javier Saurina[a,b], Oscar Núñez[a,b,c]

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

[b] Research Institute in Food Nutrition and Food Safety, Universitat de Barcelona, Av. Prat de la Riba 171, Edifici Recerca (Gaudí), E08921 Santa Coloma de Gramenet, Spain.

[c] Serra Húnter Fellow Programme, Generalitat de Catalunya, Via Laietana 2, E08003, Barcelona, Spain.

* Corresponding author: Guillem Campmajó

(e-mail: gcampmajo@ub.edu)

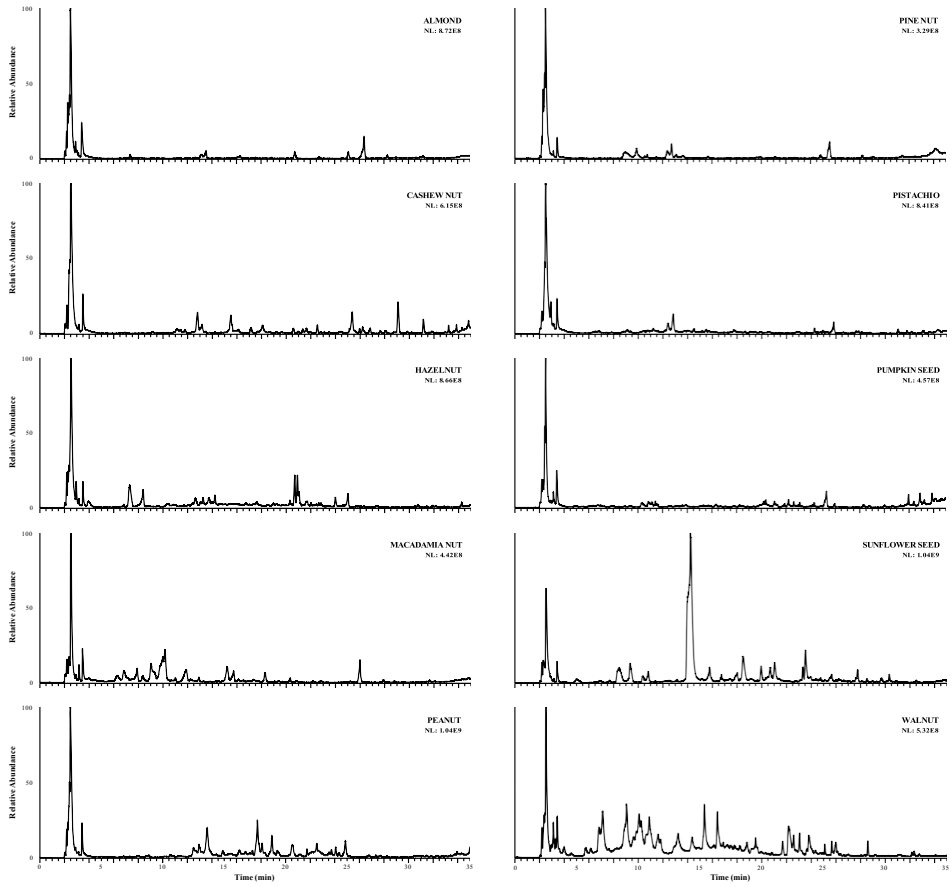


Fig. S1. Total ion current (TIC) LC-HRMS chromatogram for a selected sample within each nut type.

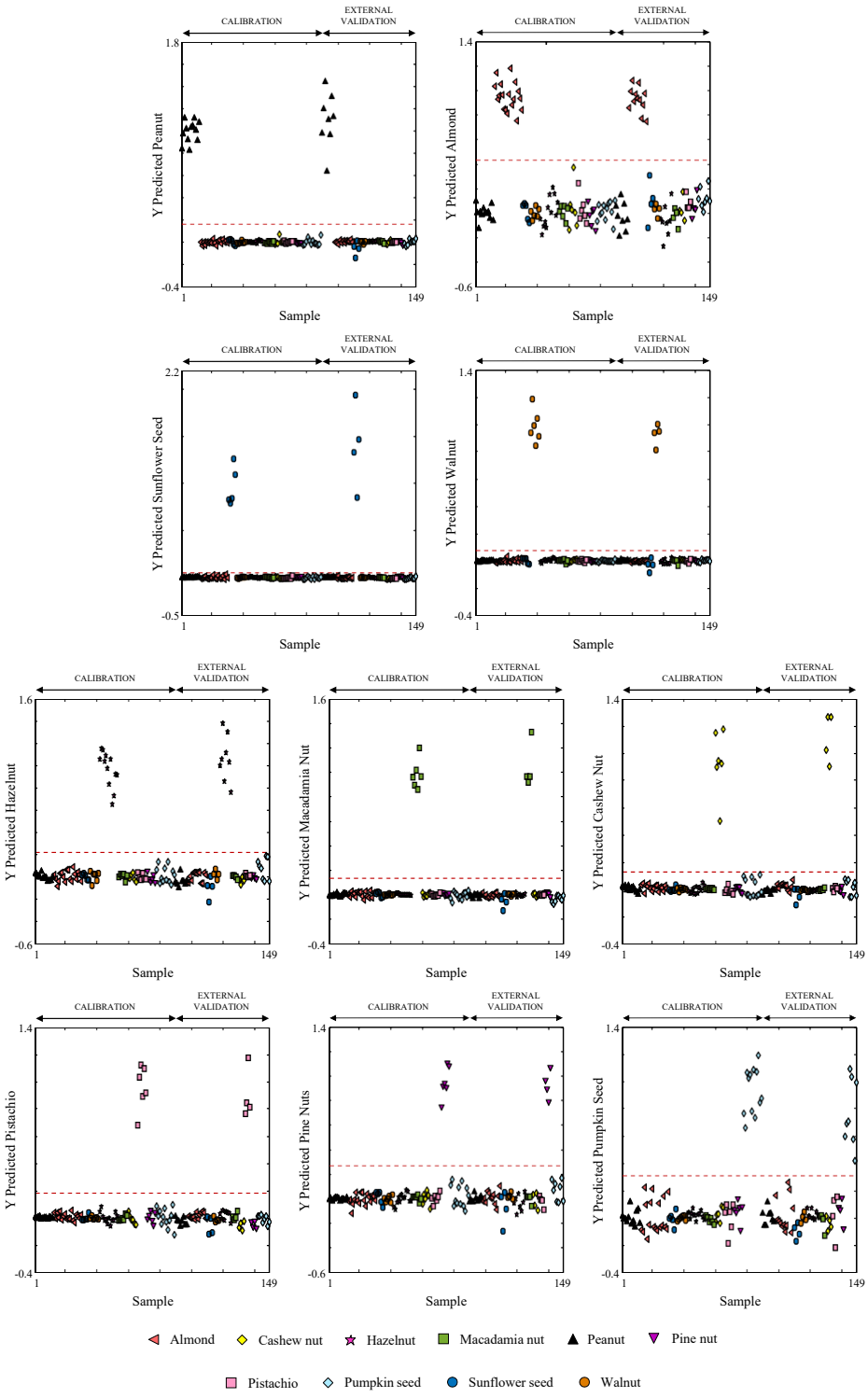


Fig. S2. External validation classification plot depicting Sample *vs*. Y predicted score plot for each nut type analysed.

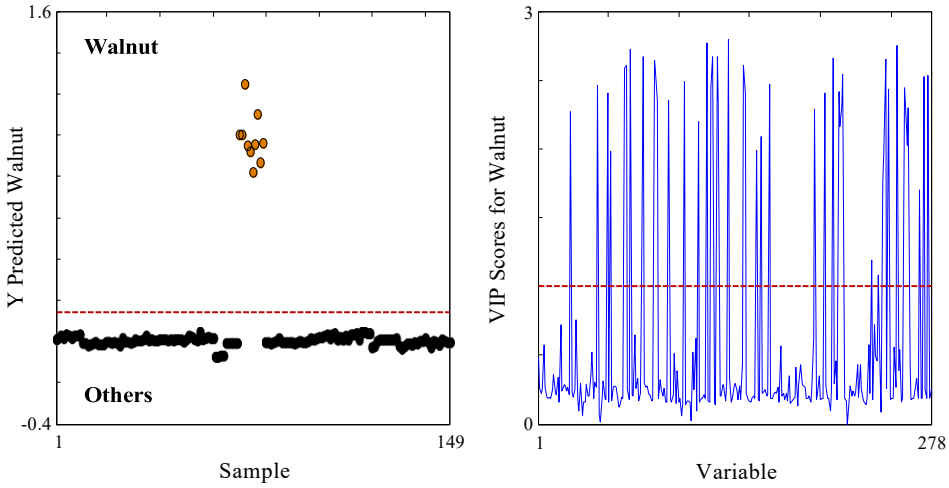


Fig. S3. On the left side, Sample *vs*. Y Predicted plot for the two-input class Walnut *vs*. Others PLS-DA model. On the right side, the corresponding VIP scores plot.

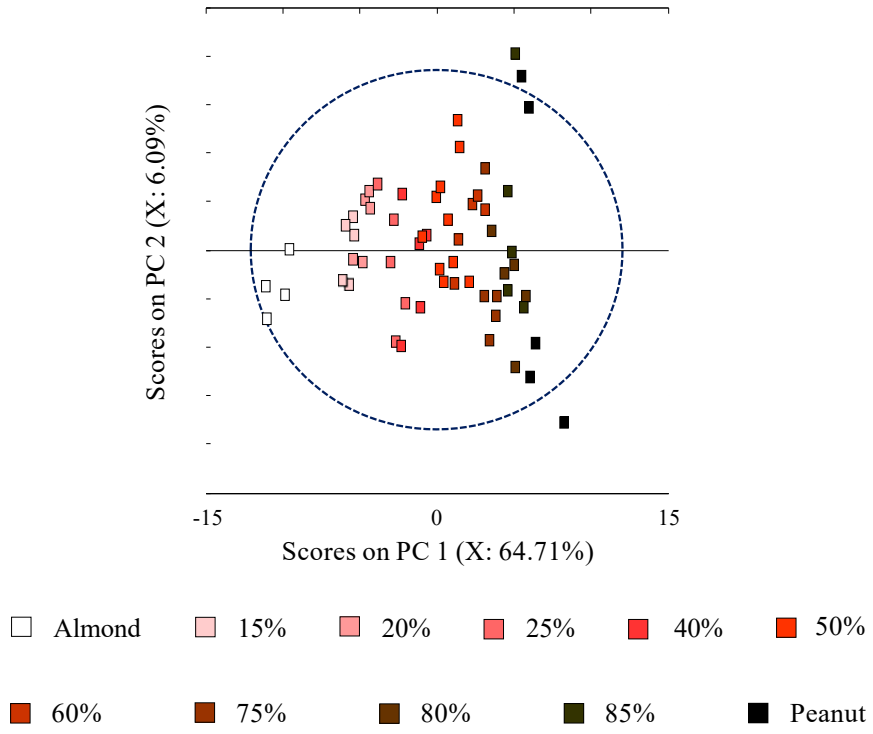


Fig. S4. PCA scores plot of PC1 *vs*. PC2 obtained for the samples analysed in the study of almond flour adulteration with peanut.

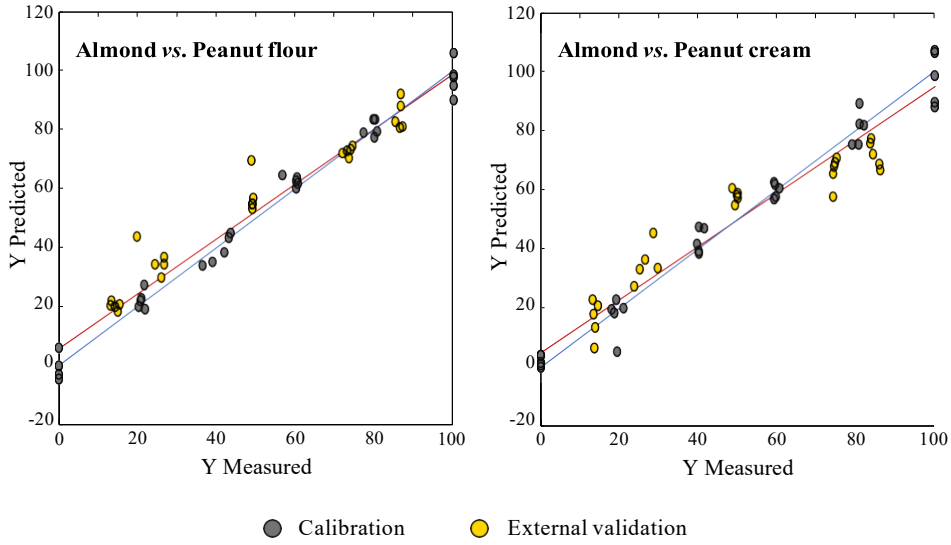


Fig. S5. External validation PLS results for the prediction of the percentage of adulteration of almond flour (on the left side) and almond custard cream (on the right side) with peanut. The blue line corresponds to the theoretical diagonal line, while the red line to the experimental adjusted one.

NUT	THERMAL TREATMENT	ORIGIN	SAMPLES
Almond	Natural	USA	AL1 – AL5
		USA	AL6 – AL10
	Toasted	Spain	AL11 – AL12
		Spain	AL13 – AL14
		Spain	AL15 – AL16
		-	AL17 – AL18
		-	AL19 – AL20
	Fried	-	AL21 – AL23
		Spain	AL24 – AL26
		-	AL27 – AL30
Cashew nut	Fried	Brazil	CN1 – CN5
		-	CN6 – CN10
Hazelnut	Natural	Turkey	HN1 – HN5
		-	HN6 – HN10
	Toasted	Spain	HN11 – HN13
		Turkey	HN14 – HN16
		Spain	HN17 – HN18
		-	HN19 – HN20
Macadamia nut	Natural	South Africa	MN1 – MN10
Peanut	Fried	Brazil	PN1 – PN3
		China	PN4 – PN6
		USA	PN7 – PN10
	Toasted	Spain	PN11 – PN13
		-	PN14 – PN16
		-	PN17 – PN20
Pine nut	Natural	Spain	PI1 – PI3
		Spain	PI4 – PI6
		Spain	PI7 – PI8
		China	PI9 – PI10
Pistachio	Toasted	Germany	PT1 – PT2
		Spain	PT3 – PT4
		Iran	PT5 – PT6
		Spain	PT7
		Iran	PT8
		Iran	PT9
		-	PT10
Pumpkin seed	Natural	Austria	PS1 – PS10
	Toasted	China	PS11 – PS20
Sunflower seed	Toasted	Spain	SS1
		Spain	SS2
		Spain	SS3
		Spain	SS4
		Spain	SS5
		Spain	SS6
		Spain	SS7
		Spain	SS8
		Spain	SS9
Walnut	Natural	USA	WN1 – WN2
		USA	WN3 – WN4
		USA	WN5 – WN6
		Chile	WN7
		-	WN8
		USA	WN9 – WN10

Table S1. Description of the nut samples analysed in the classification study.

ALMOND											
Measured accurate mass	**RT (min)**	**VIP**	**RV**	**Ion assignment**	**Molecular formula**	**RDB eq.**	**Accurate mass error (ppm)**	**Isotopic pattern score (%)**	**MS/HRMS**	**Putative Identification**	**Identification confidence level**
386.9390	2.03	2.93	0.038	-	-	-	-	-	158.9781 / 272.9588	Unknown	V
165.0403	2.35	2.20	0.029	$[M-H]^-$	$C_5H_{10}O_6$	1.5	-1.038	99.48	75.0087 / 85.0294 / 99.0085 / 129.0192 / 147.0297	2-(2-hydroperoxyethoxymethoxy)acetic acid	II
209.0300	2.38	2.72	0.035	$[M-H]^-$	$C_6H_{10}O_8$	2.5	-1.150	99.98	85.0293 / 111.0085 / 129.0192 / 133.0141 / 147.0298 / 173.0090 / 191.0196	Glucaric acid	II
341.1083	2.48	2.02	0.026	$[M-H]^-$	$C_{12}H_{22}O_{11}$	2.5	-1.743	99.45	101.0242 / 113.0242 / 119.0347 / 125.0241 / 131.0347 /	Disaccharide	III
387.1139		2.17	0.028	$[M+FA-H]^-$		2.5	-1.327	99.25	143.0347 / 149.0453 / 161.0453 / 179.0559 / 281.0873		
683.2243	2.48	2.13	0.028	$[M-H]^-$	$C_{24}H_{44}O_{22}$	3.5	-1.311	94.90	341.1084 / 647.2046	Tetrasaccharide + H_2O	III
729.2304		2.08	0.027	$[M+FA-H]^-$		3.5	-0.350	96.32			
191.0564	2.51	2.05	0.027	$[M-H]^-$	$C_7H_{12}O_6$	2.5	-1.263	99.49	85.0293 / 93.0344 / 109.0293 / 111.0449 / 127.0398 / 171.0298 / 173.0453	Quinic acid	I
533.1718	2.51	4.10	0.053	$[M-H]^-$	$C_{19}H_{34}O_{17}$	3.5	-0.905	98.50	191.0559	Quinic acid derivative	III
133.0143	2.91	3.28	0.043	$[M-H]^-$	$C_4H_6O_5$	2.5	0.101	99.96	71.0138 / 115.0036	Malic acid	II
191.0196	3.39	3.26	0.042	$[M-H]^-$	$C_6H_8O_7$	3.5	-0.554	99.41	85.0294 / 87.0087 / 111.0088 / 129.0192 / 173.0091	Citric acid	I
413.1658	3.46	-	-	$[M-H]^-$	$C_{16}H_{30}O_{12}$	2.5	-1.693	98.74	161.0453 / 251.1132	Unknown	IV
459.1716		4.50	0.058	$[M+FA-H]^-$		2.5	-0.803	97.38			
456.1506	13.52	-	-	$[M-H]^-$	$C_{20}H_{27}NO_{11}$	8.5	-1.126	99.38	[A] 323.0978 / 339.1192 / 340.1230 / 456.1509	Amygdalin	II
502.1564		2.17	0.028	$[M+FA-H]^-$		8.5	-0.424	97.91			
511.2538	25.10	-	-	$[M-H]^-$	$C_{26}H_{40}O_{10}$	7.5	-2.192	88.43	[A] 349.2013 / 511.2548	Amygdaloside	II
557.2602		4.73	0.061	$[M+FA-H]^-$		7.5	-0.323	96.06			
CASHEW NUT											
Measured accurate mass	**RT (min)**	**VIP**	**RV**	**Ion assignment**	**Molecular formula**	**RDB eq.**	**Accurate mass error (ppm)**	**Isotopic pattern score (%)**	**MS/HRMS**	**Putative Identification**	**Identification confidence level**
289.0719	12.80	2.54	0.033	$[M-H]^-$	$C_{15}H_{14}O_6$	9.5	0.445	98.39	125.0242 / 137.0242 / 165.0192 / 179.0348 / 203.0713 /	(+)-catechin	I
579.1511		2.88	0.038	$[2M-H]^-$		17.5	0.554	97.80	205.0504 / 231.0297 / 245.0818 / 271.0610		
457.1357	13.17	3.51	0.046	$[M-H]^-$	$C_{20}H_{26}O_{12}$	8.5	1.270	96.43	163.0398 / 205.0502 / 325.0923	7-hydroxy-4-methylphthalide O-[arabinosyl-(1->6)-glucoside]	II
289.0718	15.50	3.60	0.047	$[M-H]^-$	$C_{15}H_{14}O_6$	9.5	0.722	98.56	125.0242 / 137.0242 / 165.0192 / 179.0348 / 203.0713 / 205.0504 / 231.0297 / 245.0818 / 271.0610	(-)-epicatechin	I
401.1456	15.61	3.17	0.041	$[M-H]^-$	$C_{18}H_{26}O_{10}$	6.5	0.598	99.09	161.0453 / 269.1027	Benzyl dihexoside isomer	III
241.1084	17.15	3.21	0.042	$[M-H]^-$	$C_{12}H_{18}O_5$	4.5	1.174	98.89	153.1283 / 179.1077 / 197.1183 / 223.0974	Unknown	IV
241.1084	20.59	3.09	0.040	$[M-H]^-$	$C_{12}H_{18}O_5$	4.5	1.174	99.09	153.1283 / 179.1076 / 197.1182 / 223.0973	Unknown	IV
529.2659	22.54	3.79	0.049	$[M-H]^-$	$C_{26}H_{42}O_{11}$	6.5	0.821	99.11	285.1856 / 303.1961 / 305.2119 / 331.1911 / 347.1861 /	Unknown	IV

									349.2017 / 365.1967 / 367.2122 / 511.2550		
463.2190	25.96	3.67	0.048	$[M-H]^-$	$C_{21}H_{36}O_{11}$	4.5	1.155	98.54	161.0453 / 221.0663 / 331.1757 / 353.1084 / 445.2075	Unknown	IV
463.2189	26.21	3.97	0.052	$[M-H]^-$	$C_{21}H_{36}O_{11}$	4.5	0.961	98.07	161.0454 / 221.0665 / 331.1760 / 353.1086 / 445.2080	Unknown	IV
463.2190	26.83	3.76	0.049	$[M-H]^-$	$C_{21}H_{36}O_{11}$	4.5	1.198	99.10	161.0453 / 221.0664 / 331.1758 / 353.1085 / 445.2079	Unknown	IV
507.2449	29.09	4.03	0.053	$[M-H]^-$	$C_{23}H_{40}O_{12}$	4.5	0.493	97.99	149.0452 / 161.0452 / 179.0557 / 191.0557 / 233.0663 /	Unknown	IV
1015.4972		3.49	0.046	$[2M-H]^-$		7.5	0.467	89.33	253.4846 / 331.2120 / 357.1939 / 375.2018 / 463.2546 / 489.2334		
507.2448	31.14	3.69	0.048	$[M-H]^-$	$C_{23}H_{40}O_{12}$	4.5	0.198	98.24	[B] 149.0452 / 161.0450 / 191.0557 / 233.0663 / 293.0873	Unknown	IV
543.2220		3.74	0.049	$[M+Cl]^-$		3.5	1.091	99.19	/ 311.0980 / 507.2441		
553.2506		3.46	0.045	$[M+FA-H]^-$		4.5	0.652	94.83			
HAZELNUT											
Measured accurate mass	**RT (min)**	**VIP**	**RV**	**Ion assignment**	**Molecular formula**	**RDB eq.**	**Accurate mass error (ppm)**	**Isotopic pattern score (%)**	**MS/HRMS**	**Putative Identification**	**Identification confidence level**
131.0462	2.27	3.64	0.049	$[M-H]^-$	$C_4H_8N_2O_3$	2.5	-0.118	99.54	87.0449 / 95.0249 / 113.0354 / 114.0195	Asparagine	II
223.0459	2.44	3.25	0.044	$[M-H]^-$	$C_7H_{12}O_8$	2.5	-1.168	99.85	125.0242 / 143.0347 / 147.0297 / 205.0351	Unknown	IV
262.0565	2.48	3.59	0.049	$[M-H]^-$	$C_9H_{13}NO_8$	4.5	-1.601	99.76	142.0142 / 244.0458	Ascorbalamic acid isomer	III
205.0351	2.57	2.47	0.033	$[M-H]^-$	$C_7H_{10}O_7$	3.5	-1.297	99.64	81.0345 / 125.0243 / 143.0348	Unknown	IV
203.0197	3.49	3.73	0.050	$[M-H]^-$	$C_7H_8O_7$	4.5	0.119	99.48	69.0344 / 97.0293 / 115.0034 / 141.0191	Unknown	IV
368.0984	7.31	4.11	0.056	$[M-H]^-$	$C_{16}H_{19}NO_9$	8.5	-0.718	97.64	101.0241 / 113.0241 / 119.0347 / 143.0347 / 144.0452 / 161.0453 / 179.0558 / 188.0350	3-(O- -D-glycosyl)dioxindole-3-acetic acid isomer	III
368.0985	8.39	4.04	0.055	$[M-H]^-$	$C_{16}H_{19}NO_9$	8.5	-0.555	99.53	101.0241 / 113.0242 / 119.0347 / 143.0348 / 144.0453 / 161.0453 / 179.0559 / 188.0351	3-(O- -D-glycosyl)dioxindole-3-acetic acid isomer	III
443.1920	12.39	2.56	0.035	$[M-H]^-$	$C_{21}H_{32}O_{10}$	6.5	-0.633	97.02	143.0346 / 161.0452 / 189.1282 / 201.1282 / 219.1387 / 237.1491 / 263.1285 / 281.1391 / 425.1815	Cynaroside A isomer	III
679.1813	12.62	3.42	0.046	-	-	-	-	-	232.0612 / 252.0696 / 274.0717 / 282.0801 / 334.0929 / 344.0805 / 506.1338 / 543.1985 / 574.1389 / 617.1813 / 661.1709	Unknown	V
443.1920	13.04	2.42	0.033	$[M-H]^-$	$C_{21}H_{32}O_{10}$	6.5	-0.633	96.97	143.0347 / 161.0452 / 189.1282 / 201.1279 / 219.1388 / 237.1493 / 263.1287 / 281.1394 / 425.1818	Cynaroside A isomer	III
679.1809	13.25	2.94	0.040	-	-	-	-	-	232.0612 / 252.0696 / 274.0717 / 282.0801 / 334.0928 / 344.0804 / 506.1335 / 543.1983 / 574.1387 / 617.1809 / 661.1706	Unknown	V
514.1561	13.99	3.15	0.043	$[M-H]^-$	$C_{22}H_{29}NO_{13}$	9.5	-1.017	96.57	146.0608 / 172.0402 / 190.0507 / 232.0611 / 262.0718 / 292.0821 / 334.0929 / 341.1085	2-[2-[3,4-Dihydroxy-6-(hydroxymethyl)-5-[3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxyoxan-2-yl]oxyethyl]isoindole-1,3-dione isomer	III
679.1810	14.22	2.89	0.039	-	-	-	-	-	232.0613 / 252.0696 / 274.0718 / 282.0802 / 334.0929 /	Unknown	V

									344.0805 / 506.1337 / 543.1984 / 574.1391 / 617.1813 / 661.1710		
541.1458	20.91	3.79	0.051	$[M-H]^-$	$C_{26}H_{26}N_2O_{11}$	15.5	-1.151	95.49	190.0509 / 232.0613 / 262.0720 / 292.0824 / 352.1037	2-(3-hydroxy-2-oxoindolin-3-yl) acetic acid 3-O-6'-galactopyranosyl-2"-(2"oxoindolin-3"yl) isomer	III
540.1719	21.05	3.99	0.054	$[M-H]^-$	$C_{24}H_{31}NO_{13}$	10.5	-0.765	97.05	157.0504 / 171.0660 / 188.0350 / 189.0766 / 466.1353	Hazelnutin D isomer	III
259.1911	34.26	3.53	0.048	$[M-H]^-$	$C_{14}H_{28}O_4$	1.5	-1.515	99.15	141.1283 / 171.1389 / 185.1545 / 195.1752 / 213.1858 / 241.1807	Ipurolic acid	II
MACADAMIA NUT											
Measured accurate mass	**RT (min)**	**VIP**	**RV**	**Ion assignment**	**Molecular formula**	**RDB eq.**	**Accurate mass error (ppm)**	**Isotopic pattern score (%)**	**MS/HRMS**	**Putative Identification**	**Identification confidence level**
439.0757	2.62	2.62	0.033	$[M-H]^-$	$C_{14}H_{21}N_2O_{12}P$	6.5	-0.487	94.71	259.0221 / 277.0325 / 341.1070 / 377.0758 / 387.1011	Unknown	IV
261.0731	3.20	3.18	0.040	$[M-H]^-$	$C_9H_{14}N_2O_7$	4.5	1.172	99.95	74.0246 / 104.0350 / 115.0034 / 115.0511 / 116.0351 / 132.0299 / 145.0617 / 146.0457 / 187.0722 / 199.0720 / 217.0827 / 231.0620 / 243.0619	Unknown	IV
417.1409	6.28	-	-	$[M-H]^-$	$C_{18}H_{26}O_{11}$	6.5	1.547	97.58	293.0873	Guaiacol hexose-pentose isomer	III
463.1465		2.91	0.036	$[M+FA-H]^-$		6.5	1.762	96.27			
251.0677	6.32	3.46	0.043	$[M-H]^-$	$C_{11}H_{12}N_2O_5$	7.5	1.734	99.52	92.0504 / 114.0194 / 115.0034 / 135.0562 / 136.0402 / 189.0668 / 207.0773 / 233.0565	Unknown	IV
299.0777	6.82	3.54	0.044	$[M-H]^-$	$C_{13}H_{16}O_8$	6.5	1.402	98.92	137.0247	Hydroxybenzoic acid glucoside isomer	III
345.0832		2.63	0.033	$[M+FA-H]^-$		6.5	1.246	98.32			
299.0776	7.89	3.58	0.045	$[M-H]^-$	$C_{13}H_{16}O_8$	6.5	1.235	99.04	137.0247	Hydroxybenzoic acid glucoside isomer	III
417.1408	7.90	-	-	$[M-H]^-$	$C_{18}H_{26}O_{11}$	6.5	1.355	97.22	293.0875	Guaiacol hexose-pentose isomer	III
463.1464		2.84	0.035	$[M+FA-H]^-$		6.5	1.568	96.18			
431.1202	8.38	3.33	0.042	$[M-H]^-$	$C_{18}H_{24}O_{12}$	7.5	1.649	98.35	137.0242 / 191.0559 / 233.0664 / 293.0875	Apiosylglucosyl 4-hydroxybenzoate isomer	III
431.1202	9.29	3.35	0.042	$[M-H]^-$	$C_{18}H_{24}O_{12}$	7.5	1.533	96.76	137.0241 / 191.0557 / 233.0663 / 293.0874	Apiosylglucosyl 4-hydroxybenzoate isomer	III
266.0673	10.01	3.52	0.044	$[M-H]^-$	$C_{12}H_{13}NO_6$	7.5	1.013	98.74	88.0401 / 107.0500 / 114.0194 / 115.0034 / 132.0300 / 150.0559 / 151.0398 / 222.0768 / 248.0561	(2S)-3-(2-acetyloxybenzoyl)oxy-2-aminopropanoic acid isomer	III
252.0517	10.16	3.18	0.040	$[M-H]^-$	$C_{11}H_{11}NO_6$	7.5	1.229	99.95	93.0343 / 114.0194 / 115.0034 / 132.0300 / 136.0402 / 137.0242 / 190.0508 / 208.0614 / 234.0404 / 252.0456	Unknown	IV
417.1410	11.00	2.68	0.033	$[M-H]^-$	$C_{18}H_{26}O_{11}$	6.5	1.930	95.95	255.0505 / 357.0820	Oleoside dimethyl ester	II
239.0565	15.20	3.11	0.039	$[M-H]^-$	$C_{11}H_{12}O_6$	6.5	1.626	99.90	91.0551 / 131.0500 / 133.0656 / 149.0606 / 159.0449 / 163.0398 / 177.0554 / 193.0503 / 195.0660	2-benzyl-3-carboxyoxy-2-hydroxy-propanoic acid isomer	III
415.1617	18.28	-	-	$[M-H]^-$	$C_{19}H_{28}O_{10}$	6.5	1.637	97.46	131.0351 / 149.0449 / 191.0565 / 251.0771	Phenylethyl diglycoside	III
461.1666		3.54	0.044	$[M+FA-H]^-$		6.5	0.240	96.44			
527.1417	20.31	3.03	0.038	$[M-H]^-$	$C_{23}H_{28}O_{14}$	10.5	1.938	94.76	157.0292 / 158.0370 / 159.0449 / 171.0448 / 173.0605 /	Unknown	IV

									185.0604 / 195.0448 / 201.0190 / 201.0553 / 202.0268 / 203.0346 / 213.0553 / 215.0345 / 225.0553 / 227.0346 / 231.0658 / 243.0659 / 245.0450 / 269.0451 / 275.0556 / 365.0873 / 509.1296		
PEANUT											
Measured accurate mass	**RT (min)**	**VIP**	**RV**	**Ion assignment**	**Molecular formula**	**RDB eq.**	**Accurate mass error (ppm)**	**Isotopic pattern score (%)**	**MS/HRMS**	**Putative Identification**	**Identification confidence level**
149.0090	2.54	2.82	0.030	$[M-H]^-$	$C_4H_6O_6$	2.5	-0.880	99.98	87.0086 / 103.0035 / 130.9984	Tartaric acid	I
295.0455	12.50	2.86	0.031	$[M-H]^-$	$C_{13}H_{12}O_8$	8.5	-1.663	99.66	149.0087 / 163.0399	*cis*-*p*-coumaroyl tartaric acid	II
431.1553	12.95	2.03	0.022	$[M-H]^-$	$C_{19}H_{28}O_{11}$	6.5	-1.426	98.20	131.0346 / 143.0346 / 149.0452 / 161.0452 / 179.0558 /	Darendoside A isomer	III
477.1609		2.13	0.023	$[M+FA-H]^-$		6.5	-0.994	97.02	191.0557 / 233.0662 / 269.1026 / 293.0873 / 299.1132		
293.1142	13.60	2.81	0.030	$[M-H]^-$	$C_{14}H_{18}N_2O_5$	7.5	-0.358	97.50	128.0352 / 131.0712 / 164.0716 / 257.0930 / 275.1036	Unknown	IV
295.0454	13.63	2.22	0.024	$[M-H]^-$	$C_{13}H_{12}O_8$	8.5	-1.832	99.57	112.9881 / 119.0502 / 163.0398	*trans*-*p*-coumaroyl tartaric acid	II
325.0563	14.42	2.66	0.028	$[M-H]^-$	$C_{14}H_{14}O_9$	8.5	-0.539	99.38	112.9878 / 193.0505	Feruloyl tartaric acid isomer	III
250.0719	14.90	2.65	0.028	$[M-H]^-$	$C_{12}H_{13}NO_5$	7.5	-0.943	99.71	132.0301 / 135.0449 / 206.0820 / 232.0613	N-phenylacetylaspartate	II
400.0672	15.55	2.25	0.024	$[M-H]^-$	$C_{19}H_{15}NO_9$	13.5	-0.435	97.98	112.9879 / 119.0501 / 163.0398 / 203.0351 / 215.0349 / 277.0355	*p*-coumaroylnicotinoyl tartaric acid	II
295.0455	16.22	2.22	0.024	$[M-H]^-$	$C_{13}H_{12}O_8$	8.5	-1.425	99.56	103.0037 / 119.0502 / 130.9987 / 163.0398	Coumaroyl-O-pentoside isomer	III
575.1975	17.70	2.88	0.031	$[M-H]^-$	$C_{25}H_{36}O_{15}$	8.5	-1.171	95.84	413.1454 / 431.1561 / 473.1668 / 513.1982	Unknown	IV
1151.4023		2.62	0.028	$[2M-H]^-$		15.5	-1.123	88.78			
501.1513	17.83	2.60	0.028	$[M-H]^-$	$C_{24}H_{26}N_2O_{10}$	13.5	-0.415	98.43	58.0297 / 91.0553 / 147.0446 / 164.0715 / 206.0822 / 250.0720	Unknown	IV
473.1658	18.09	2.47	0.026	$[M-H]^-$	$C_{21}H_{30}O_{12}$	7.5	-1.457	98.97	59.0138 / 71.0138 / 73.0294 / 89.0244 / 99.0087 / 101.0243 / 113.0243 / 119.0502 / 191.0563	Unknown	IV
579.1729	18.89	2.56	0.027	$[M-H]^-$	$C_{28}H_{28}N_4O_{10}$	17.5	-0.563	94.53	203.0823 / 245.0928 / 289.0825	Unknown	IV
163.0400	18.92	2.93	0.031	$[M-H]^-$	$C_9H_8O_3$	6.5	-0.352	99.84	119.0501	*p*-coumaric acid	II
245.0929	19.31	2.72	0.029	$[M-H]^-$	$C_{13}H_{14}N_2O_3$	8.5	-1.043	99.87	74.0246 / 98.0245 / 116.0351 / 201.1032 / 203.0824	N-acetyltryptophan	II
279.0507	19.33	2.50	0.027	$[M-H]^-$	$C_{13}H_{12}O_7$	8.5	-1.347	99.77	71.0138 / 115.0036 / 119.0501 / 163.0397	*p*-coumaroyl malic acid	II
248.0926	19.46	2.76	0.030	$[M-H]^-$	$C_{13}H_{15}NO_4$	7.5	-0.851	98.87	82.0297 / 100.0402 / 147.0450 / 164.0717 / 186.0923	Unknown	IV
501.1038	21.97	2.35	0.025	$[M-H]^-$	$C_{24}H_{22}O_{12}$	14.5	-0.198	98.85	119.0501 / 145.0289 / 149.0239 / 163.0395 / 164.0477 / 203.0348 / 263.0558 / 277.0349 / 337.0559	Malonyldaidzin isomer	III
441.0824	22.19	2.67	0.029	$[M-H]^-$	$C_{22}H_{18}O_{10}$	14.5	-0.816	96.11	119.0502 / 145.0291 / 147.0048 / 163.0399 / 175.0400 / 203.0352 / 277.0355 / 295.0460	di-*p*-coumaroyl tartaric acid isomer	III
471.0927	22.27	2.29	0.025	$[M-H]^-$	$C_{23}H_{20}O_{11}$	14.5	-1.198	96.28	119.0501 / 134.0373 / 175.0399 / 203.0350 / 233.0453 / 277.0352 / 307.0460	*p*-coumaroylferuloyl tartaric acid isomer	III
803.3337	22.54	2.66	0.028	$[M-H]^-$	$C_{37}H_{56}O_{19}$	10.5	-0.688	93.43	365.1966 / 641.2813 / 659.2920 / 701.3026	Unknown	IV
501.1036	22.62	2.39	0.026	$[M-H]^-$	$C_{24}H_{22}O_{12}$	14.5	-0.398	95.62	119.0501 / 145.0291 / 149.0238 / 163.0398 / 164.0477 / 203.0346 / 263.0559 / 277.0349 / 337.0567	Malonyldaidzin isomer	III

137.0244	22.73	2.47	0.026	$[M-H]^-$	$C_7H_6O_3$	5.5	-0.127	99.39	65.0397 / 93.0345 / 94.0378 / 108.0216	Hydroxybenzoic acid isomer	III
PINE NUT											
Measured accurate mass	**RT (min)**	**VIP**	**RV**	**Ion assignment**	**Molecular formula**	**RDB eq.**	**Accurate mass error (ppm)**	**Isotopic pattern score (%)**	**MS/HRMS**	**Putative Identification**	**Identification confidence level**
262.0571	2.62	5.45	0.090	$[M-H]^-$	$C_9H_{13}NO_8$	4.5	0.879	99.66	96.0453 / 128.0351 / 140.0351 / 158.0457 / 172.0250 / 200.0563 / 202.0356 / 218.0669 / 219.0509 / 244.0462	Ascorbalamic acid isomer	III
525.1216		4.70	0.078	$[2M-H]^-$		7.5	1.265	97.46			
329.0884	8.89	4.59	0.076	$[M-H]^-$	$C_{14}H_{18}O_9$	6.5	1.837	98.72	167.0347 / 209.0450	Vanillic acid glucoside isomer	III
329.0884	9.90	4.67	0.077	$[M-H]^-$	$C_{14}H_{18}O_9$	6.5	1.655	98.72	167.0348 / 181.0504 / 209.0452 / 239.0559 / 269.0664	Vanillic acid glucoside isomer	III
451.1366	12.74	2.56	0.042	$[M-H]^-$	$C_{20}H_{24}N_2O_{10}$	10.5	1.822	96.80	173.0718 / 292.1187 / 335.1246 / 336.1084 / 407.1457 / 433.1249	Unknown	IV
347.1717	13.15	-	-	$[M-H]^-$	$C_{16}H_{28}O_8$	3.5	1.668	95.88	[A] 179.0557 / 347.1705	Unknown	IV
393.1771		4.51	0.074	$[M+FA-H]^-$		3.5	2.110	89.42			
361.1513	13.62	3.87	0.064	$[M-H]^-$	$C_{17}H_{22}N_4O_5$	9.5	-1.254	96.47	101.0241 / 113.0241 / 161.0452 / 317.1600	Unknown	IV
477.2346	24.83	-	-	$[M-H]^-$	$C_{22}H_{38}O_{11}$	4.5	1.016	98.35	[A] 163.0609 / 307.1028 / 325.1136 / 331.1758 / 477.2344	Unknown	IV
523.2406		4.28	0.071	$[M+FA-H]^-$		4.5	1.884	97.77			
PISTACHIO											
Measured accurate mass	**RT (min)**	**VIP**	**RV**	**Ion assignment**	**Molecular formula**	**RDB eq.**	**Accurate mass error (ppm)**	**Isotopic pattern score (%)**	**MS/HRMS**	**Putative Identification**	**Identification confidence level**
179.0561	2.37	-	-	$[M-H]^-$	$C_6H_{12}O_6$	1.5	-0.119	99.24	[B] 89.0242 / 119.0348 / 143.0347 / 161.0454 / 179.0559	Monosaccharide	III
215.0328		2.13	0.028	$[M+Cl]^-$		0.5	-0.135	89.66			
323.0284	2.71	3.18	0.042	$[M-H]^-$	$C_9H_{13}N_2O_9P$	5.5	-0.742	99.66	96.9694 / 138.9799 / 150.9799 / 192.9904 / 211.0010 / 280.0224	Uridine monophosphate	II
133.0143	2.91	2.36	0.031	$[M-H]^-$	$C_4H_6O_5$	2.5	0.778	98.48	71.0138 / 115.0034	Malic acid	I
191.0198	2.92	3.19	0.042	$[M-H]^-$	$C_6H_8O_7$	3.5	0.388	99.95	111.0085 / 129.0191 / 154.9983 / 173.0089	Isocitric acid	II
346.0557	3.09	2.85	0.038	$[M-H]^-$	$C_{10}H_{14}N_5O_7P$	7.5	-0.253	99.25	150.9798 / 192.9903 / 211.0009	Adenosine 5'-monophosphate	II
153.0194	10.44	2.98	0.040	$[M-H]^-$	$C_7H_6O_4$	5.5	0.771	99.87	109.0293	Protocatechuic acid	II
431.1557	12.23	3.54	0.047	$[M-H]^-$	$C_{19}H_{28}O_{11}$	6.5	-0.521	99.58	131.0346 / 143.0345 / 149.0452 / 161.0453 / 179.0554 / 191.0557 / 233.0664 / 251.0768 / 293.0877 / 299.1132	Darendoside A isomer	III
451.1357	12.40	-	-	$[M-H]^-$	$C_{20}H_{24}N_2O_{10}$	10.5	-0.240	98.40	173.0719 / 292.1188 / 335.1248 / 336.1085 / 407.1459 / 433.1251	Unknown	IV
903.2790		3.24	0.043	$[2M-H]^-$		19.5	0.074	89.47			
451.1357	12.82	3.35	0.045	$[M-H]^-$	$C_{20}H_{24}N_2O_{10}$	10.5	-0.195	98.39	173.0719 / 292.1188 / 335.1248 / 336.1085 / 407.1459 / 433.1251	Unknown	IV
903.2790		2.74	0.036	$[2M-H]^-$		19.5	0.085	88.27			
395.1558	13.19	3.84	0.051	$[M-H]^-$	$C_{16}H_{28}O_{11}$	3.5	-0.088	98.37	125.0241 / 233.1025 / 251.1130 / 293.1237 / 333.1551 / 351.1655	Unknown	IV
395.1558	13.74	3.53	0.047	$[M-H]^-$	$C_{16}H_{28}O_{11}$	3.5	-0.189	99.01	125.0241 / 233.1026 / 251.1132 / 293.1238 / 333.1550 / 351.1654	Unknown	IV

323.1347	14.02	3.60	0.048	$[M-H]^-$	$C_{13}H_{24}O_9$	2.5	-0.011	99.39	89.0242 / 95.0136 / 101.0241 / 113.0241 / 119.0347 / 125.0241 / 131.0347 / 143.0347 / 149.0452 / 161.0453 / 161.0816 / 179.0558	Unknown	IV
323.1346	14.51	3.98	0.053	$[M-H]^-$	$C_{13}H_{24}O_9$	2.5	-0.419	99.17	89.0242 / 95.0136 / 101.0241 / 113.0241 / 119.0347 / 125.0241 / 131.0346 / 143.0347 / 149.0452 / 161.0453 / 161.0816 / 179.0558	Unknown	IV
463.0881	23.03	3.22	0.043	$[M-H]^-$	$C_{21}H_{20}O_{12}$	12.5	-0.106	98.04	301.0351 / 343.0454	Quercetin 3-O-glucoside	II
507.2439	24.26	-	-	$[M-H]^-$	$C_{23}H_{40}O_{12}$	4.5	-1.616	97.31	375.2022	Unknown	IV
553.2500		3.61	0.048	$[M+FA-H]^-$		4.5	-0.414	96.28			

PUMPKIN SEED

Measured accurate mass	RT (min)	VIP	RV	Ion assignment	Molecular formula	RDB eq.	Accurate mass error (ppm)	Isotopic pattern score (%)	MS/HRMS	Putative Identification	Identification confidence level
503.1623	2.52	-	-	$[M-H]^-$	$C_{18}H_{32}O_{16}$	3.5	1.117	98.70	[B] 179.0557 / 323.0977 / 341.1083 / 377.0852 / 503.1617	Trisaccharide	III
539.1379		2.02	0.034	$[M+Cl]^-$		2.5	-0.556	95.07			
209.0670	2.54	3.28	0.062	$[M-H]^-$	$C_7H_{14}O_7$	1.5	1.645	99.63	85.0293 / 87.0085 / 99.0084 / 111.0084 / 129.0190 / 141.0190 / 159.0295	Sedoheptulose	II
337.0784	3.06	2.90	0.057	$[M-H]^-$	$C_{12}H_{18}O_{11}$	4.5	2.360	87.14	277.0563	Unknown	IV
213.0885	3.40	4.58	0.088	$[M-H]^-$	$C_9H_{14}N_2O_4$	4.5	1.829	99.68	98.0608 / 110.0608 / 169.0980 / 171.0772	Unknown	IV
282.0848	3.49	3.92	0.076	$[M-H]^-$	$C_{10}H_{13}N_5O_5$	7.5	1.305	98.53	133.0154 / 150.0420	Guanosine	II
431.1566	10.32	2.81	0.055	$[M-H]^-$	$C_{19}H_{28}O_{11}$	6.5	1.682	97.86	137.0604 / 161.0452 / 221.0814 / 299.1131 / 413.1447	Tyrosol diglycoside isomer	III
673.3090	22.22	4.76	0.092	$[M-H]^-$	$C_{32}H_{50}O_{15}$	8.5	1.999	94.80	205.0714 / 317.1754 / 347.1860 / 365.1966 / 465.2490 / 527.2498 / 611.3067 / 643.2967 / 645.3123 / 655.2980	3-[2-[5,8a-dimethyl-2-methylene-6-[3,4,5-trihydroxy-6-(hydroxymethyl)tetrahydropyran-2-yl]oxy-5-[[3,4,5-trihydroxy-6-(hydroxymethyl)tetrahydropyran-2-yl]oxymethyl]decalin-1-yl]ethylidene]-4-hydroxytetrahydrofuran-2-one isomer	III

SUNFLOWER SEED

Measured accurate mass	RT (min)	VIP	RV	Ion assignment	Molecular formula	RDB eq.	Accurate mass error (ppm)	Isotopic pattern score (%)	MS/HRMS	Putative Identification	Identification confidence level
209.0567	3.41	2.34	0.020	$[M-H]^-$	$C_9H_{10}N_2O_4$	6.5	-0.574	99.93	121.0404 / 122.0244 / 163.0510 / 191.0459	Unknown	IV
575.2607	8.48	2.33	0.020	$[M-H]^-$	$C_{29}H_{40}N_2O_{10}$	11.5	-0.554	97.77	250.1080 / 310.1290 / 367.2119 / 515.2394	Unknown	IV
575.2614	9.32	2.28	0.019	$[M-H]^-$	$C_{29}H_{40}N_2O_{10}$	11.5	-0.675	96.47	250.1080 / 310.1293 / 515.2397	Unknown	IV
353.0875	14.22	-	-	$[M-H]^-$	$C_{16}H_{18}O_9$	8.5	-1.006	99.55	135.0448 / 173.0453 / 179.0348 / 191.0559	5-caffeoylquinic acid	II
1061.2780		2.37	0.020	$[3M-H]^-$		22.5	0.029	89.99			
353.0875	15.79	2.28	0.019	$[M-H]^-$	$C_{16}H_{18}O_9$	8.5	-0.751	99.44	135.0449 / 173.0453 / 179.0348 / 191.0560	4-caffeoylquinic acid	II

491.1192	16.74	2.31	0.019	$[M-H]^-$	$C_{23}H_{24}O_{12}$	12.5	-0.589	99.90	161.0242 / 179.0348 / 323.0768 / 447.1296 / 473.1090	Flavonoid O-glycoside isomer	III
641.1510	18.47	2.29	0.019	$[M-H]^-$	$C_{31}H_{30}O_{15}$	17.5	-0.348	97.34	191.0558 / 287.0557 / 297.0401 / 353.0872 / 423.1080 / 449.0874	5,7-dihydroxy-2-[3-(4-hydroxy-3-methoxyphenyl)-2-(hydroxymethyl)-2,3-dihydro-1,4-benzodioxin-6-yl]-3-[(2S,3R,4S,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxychromen-4-one	II
1283.3071		2.34	0.020	$[2M-H]^-$		33.5	-1.982	83.40			
542.1299	18.55	2.31	0.019	$[M-H]^-$	$C_{26}H_{25}NO_{12}$	15.5	-0.901	96.44	173.0453 / 206.0455 / 233.0663 / 335.0767 / 380.0984 / 395.0979	Unknown	IV
1085.2668		2.36	0.020	$[2M-H]^-$		29.5	-1.146	87.52			
491.1190	18.64	2.32	0.019	$[M-H]^-$	$C_{23}H_{24}O_{12}$	12.5	-0.976	98.54	161.0245 / 323.0768 / 329.0875 / 447.1297 / 473.1091	Flavonoid O-glycoside isomer	III
542.1300	19.95	2.29	0.019	$[M-H]^-$	$C_{26}H_{25}NO_{12}$	15.5	-0.735	96.04	173.0453 / 206.0456 / 233.0663 / 335.0768 / 380.0984 / 395.0979	Unknown	IV
1085.2671		2.35	0.020	$[2M-H]^-$		29.5	-0.869	83.99			
135.0452	20.49	2.30	0.019	$[M-H]^-$	$C_8H_8O_2$	5.5	0.646	99.99	91.0553 / 107.0503	Phenylacetic acid	II
515.1190	21.04	-	-	$[M-H]^-$	$C_{25}H_{24}O_{12}$	14.5	-0.911	97.27	191.0556 / 353.0873	3,4-dicaffeoylquinic acid	II
1031.2466		2.31	0.019	$[2M-H]^-$		27.5	-0.664	89.37			
563.1766	21.43	2.29	0.019	$[M-H]^-$	$C_{27}H_{32}O_{13}$	12.5	-0.682	98.78	161.0242 / 179.0347 / 269.1026 / 401.1449 / 431.1341	Naringenin 5,7-dimethyl ether 4'-O-xylosyl-(1->4)-arabinoside isomer	III
563.1765	23.33	2.37	0.020	$[M-H]^-$	$C_{27}H_{32}O_{13}$	12.5	-0.842	96.58	161.0241 / 179.0347 / 269.1026 / 401.1447 / 431.1347	Naringenin 5,7-dimethyl ether 4'-O-xylosyl-(1->4)-arabinoside isomer	III
515.1190	23.53	-	-	$[M-H]^-$	$C_{25}H_{24}O_{12}$	14.5	-0.872	96.94	173.0453 / 179.0347 / 191.0558 / 203.0346 / 255.0659 / 299.0557 / 317.0662 / 353.0874	4,5-dicaffeoylquinic acid	II
1031.2457		2.32	0.019	$[2M-H]^-$		27.5	-0.567	89.62			

WALNUT

Measured accurate mass	RT (min)	VIP	RV	Ion assignment	Molecular formula	RDB eq.	Accurate mass error (ppm)	Isotopic pattern score (%)	MS/HRMS	Putative Identification	Identification confidence level
785.0854	10.35	2.65	0.027	$[M-H]^-$	$C_{34}H_{26}O_{22}$	22.5	1.331	91.68	249.0399 / 275.0191 / 300.9984 / 331.0663 / 419.0612 / 483.0774 / 589.0827 / 615.0620 / 633.0723	Digalloyl-hexahydroxydiphenoyl-glucose isomer	III
633.0744	11.82	2.54	0.026	$[M-H]^-$	$C_{27}H_{22}O_{18}$	17.5	1.648	95.82	275.0193 / 300.9987 / 331.0666 / 463.0515	Galloyl-hexahydroxydiphenoyl-glucose isomer	III
305.0783	13.78	2.57	0.026	$[M-H]^-$	$C_{14}H_{14}N_2O_6$	9.5	1.313	97.85	115.0034 / 132.0299 / 146.0608 / 171.0561 / 189.0667 / 190.0506 / 243.0771 / 261.0877 / 269.0564 / 287.0670	N-[4-(2,5-dioxopyrrolidin-1-yl)phenyl]-L-aspartic acid isomer	III
259.1191	16.11	2.59	0.026	$[M-H]^-$	$C_{12}H_{20}O_6$	3.5	1.306	99.06	159.0660 / 197.1180 / 215.1285 / 241.1078	Dicarboxylic acid derivative isomer	III
259.1191	17.04	2.61	0.026	$[M-H]^-$	$C_{12}H_{20}O_6$	3.5	1.345	99.25	197.1183 / 199.0974 / 241.1079	Dicarboxylic acid derivative isomer	III
289.0834	17.30	2.67	0.027	$[M-H]^-$	$C_{14}H_{14}N_2O_5$	9.5	1.333	99.93	88.0402 / 114.0194 / 115.0034 / 130.0660 / 132.0300 / 156.0453 / 173.0719 / 174.0558 / 227.0823 / 245.0929 / 271.0720	Indole-3-acetyl-L-aspartic acid	II
261.1346	19.64	2.72	0.027	$[M-H]^-$	$C_{12}H_{22}O_6$	2.5	1.028	99.78	125.0969 / 169.0868 / 181.1232 / 187.0975 / 199.1338 / 201.1130 / 225.1130 / 243.1236	Dicarboxylic acid derivative isomer	III
433.0415	22.20	2.80	0.028	$[M-H]^-$	$C_{19}H_{14}O_{12}$	13.5	0.464	98.50	299.9910 / 300.9986	Ellagic acid pentoside	III

867.0909		2.75	0.028	$[2M-H]^-$		25.5	1.332	93.43			
403.1613	22.59	2.76	0.028	$[M-H]^-$	$C_{18}H_{28}O_{10}$	5.5	0.868	98.75	161.0452 / 179.1074 / 223.0972 / 359.1705	2,7-dimethyl-2,4-diene-deca- , -diacid 8-O- -glucoside	II
592.2042	23.05	2.66	0.027	$[M-H]^-$	$C_{28}H_{35}NO_{13}$	12.5	0.991	95.42	197.1180 / 241.1078 / 283.1182 / 313.1288 / 343.1394 / 403.1605	Glansreginin A	II
371.2076	23.67	-	-	$[M-H]^-$	$C_{19}H_{32}O_7$	4.5	0.171	90.6	[A] 161.0451 / 300.9985 / 371.2069	Unknown	IV
417.2133		2.67	0.027	$[M+FA-H]^-$		4.5	0.705	98.83			
300.9991	23.81	2.65	0.027	$[M-H]^-$	$C_{14}H_6O_8$	12.5	0.265	99.52	185.0240 / 201.0189 / 213.0189 / 229.0138 / 257.0086 / 273.0041 / 283.9957	Ellagic acid	II
413.2183	25.10	2.67	0.027	$[M-H]^-$	$C_{21}H_{34}O_8$	5.5	0.433	98.91	353.1964 / 371.2064	Methyl epidioxy dihydroperoxy eicosatrienoate isomer	III
413.2184	25.67	-	-	$[M-H]^-$	$C_{21}H_{34}O_8$	5.5	0.699	99.07	353.1962 / 371.2070	Methyl epidioxy dihydroperoxy eicosatrienoate isomer	III
459.2240		2.61	0.026	$[M+FA-H]^-$		5.5	0.979	98.65			

[A] MS/HRMS data corresponding to the formic acid adduct; [B] MS/HRMS data corresponding to the chlorine adduct; FA: formic acid; RDB: rings and double bounds; RT: retention time; RV: regression vector

Table S2. Tentative annotation/identification by LC-HRMS of the most discriminant molecular features found in the nut matrices under study.

ALMOND			
Molecular feature	**Significantly higher than**	**Significantly lower than**	**No significant differences**
386.9390 (2.03 min)	PN** / HN, PI, PS, WN***	-	CN, MN, PT, SS
165.0403 (2.35 min)	SS* / PI** / CN, HN, PN, PS, PT, WN***	-	MN
209.0300 (2.38 min)	PS** / CN, MN, PI, PN, PT, SS, WN***	-	HN
341.1083 (2.48 min)	PS***	-	CN, HN, MN, PI, PN, PT, SS, WN
683.2243 (2.48 min)	PS, SS **	-	CN, HN, MN, PI, PN, PT, WN
191.0564 (2.51 min)	CN, HN, MN, PI, PN, PS, PT, WN***	SS***	-
533.1718 (2.51 min)	CN, HN, MN, PI, PN, PS, PT, WN***	SS***	-
133.0143 (2.91 min)	MN, PN** / CN, PI, PS***	PT***	HN, SS, WN
191.0196 (3.39 min)	HN, PI, PS, SS, WN***	PT***	CN, MN, PN
413.1658 (3.46 min)	CN** / HN, MN, PI, PN, PS, PT, SS, WN***	-	-
456.1506 (13.52 min)	CN, MN, PI, PN, PS, PT, SS, WN*	-	HN
511.2538 (25.10 min)	CN, HN, MN, PI, PN, PS, PT, WN***	-	SS
CASHEW NUT			
Molecular feature	**Significantly higher than**	**Significantly lower than**	**No significant differences**
289.0719 (12.80 min)	AL, HN, MN, PI, PN, PS, SS**	-	PT, WN
457.1357 (13.17 min)	AL, HN, MN, PI, PN, PS, PT, SS, WN**	-	-
289.0718 (15.50 min)	AL, HN, MN, PI, PN, PS, PT, SS, WN**	-	-
401.1456 (15.61 min)	AL, HN, MN, PI, PN, PS, PT, WN**	-	SS
241.1084 (17.15 min)	AL, HN, MN, PI, PN, PS, PT, SS, WN***	-	-
241.1084 (20.59 min)	AL, HN, MN, PI, PN, PS, PT, SS, WN***	-	-
529.2659 (22.54 min)	AL, HN, MN, PI, PN, PS, PT, SS, WN**	-	-
463.2190 (25.96 min)	AL, HN, MN, PI, PN, PS, PT, SS, WN*	-	-
463.2189 (26.21 min)	AL, HN, MN, PI, PN, PS, PT, SS, WN**	-	-
463.2190 (26.83 min)	AL, HN, MN, PI, PN, PS, PT, SS, WN**	-	-
507.2449 (29.09 min)	AL, HN, MN, PI, PN, PS, PT, SS, WN**	-	-
507.2448 (31.14 min)	PT* / AL, HN, MN, PI, PN, PS, SS, WN**	-	-
HAZELNUT			
Molecular feature	**Significantly higher than**	**Significantly lower than**	**No significant differences**
131.0462 (2.27 min)	-	CN, MN, PN** / AL***	PI, PS, PT, SS, WN
223.0459 (2.44 min)	AL, CN, MN, PI, PN, PS, SS, WN ***	-	PT
262.0565 (2.48 min)	PI* / AL, CN, MN, PN, PS, SS**	-	PT, WN

205.0351 (2.57 min)	CN, PN, WN* / AL, MN, PI, PS, SS**	PT**	-
203.0197 (3.49 min)	AL, CN, MN, PI, PN, PS, PT, SS, WN***	-	-
368.0984 (7.31 min)	WN* / AL, CN, MN, PI, PN, PS, PT, SS***	-	-
368.0985 (8.39 min)	WN* / AL, CN, MN, PI, PN, PS, PT, SS***	-	-
443.1920 (12.39 min)	MN, PN, PS, WN** / AL, PI***	-	CN, PT, SS
679.1813 (12.62 min)	AL, CN, MN, PI, PN, PS, PT, SS, WN***	-	-
443.1920 (13.04 min)	MN, PN, WN**/ AL, PI, PS***	-	CN, PT, SS
679.1809 (13.25 min)	AL, CN, MN, PI, PN, PS, PT, SS, WN***	-	-
514.1561 (13.99 min)	AL, CN, MN, PI, PN, PS, PT, SS, WN***	-	-
679.1810 (14.22 min)	AL, CN, MN, PI, PN, PS, PT, SS, WN***	-	-
541.1458 (20.91 min)	AL, CN, MN, PI, PN, PS, PT, SS, WN***	-	-
540.1719 (21.05 min)	AL, CN, MN, PI, PN, PS, PT, SS, WN***	-	-
259.1911 (34.26 min)	AL, CN, MN, PI, PN, PS, PT, SS, WN***	-	-
MACADAMIA NUT			
Molecular feature	**Significantly higher than**	**Significantly lower than**	**No significant differences**
439.0757 (2.62 min)	CN, PI, PN, PS, SS, WN*	-	AL, HN, PT
261.0731 (3.20 min)	AL, CN, HN, PI, PN, PS, PT, SS, WN***	-	-
417.1409 (6.28 min)	AL, CN, HN, PI, PN, PT, SS, WN**	-	PS
251.0677 (6.32 min)	AL, CN, HN, PI, PN, PS, PT, SS, WN***	-	-
299.0776 (6.82 min)	AL, CN, HN, PI, PN, PS, PT, SS, WN***	-	-
299.0776 (7.89 min)	AL, CN, HN, PI, PN, PS, PT, SS, WN***	-	-
417.1408 (7.90 min)	AL, CN, HN, PI, PN, PT, SS, WN**	-	PS
431.1202 (8.38 min)	AL, HN, PI, PN, PS, PT, SS, WN*	-	CN
431.1202 (9.29 min)	AL, HN, PI, PN, PS, PT, SS, WN*	-	CN
266.0673 (10.01 min)	HN, PN, WN** / AL, CN, PI, PS, PT, SS***	-	-
252.0517 (10.16 min)	AL, CN, HN, PI, PN, PS, PT, SS, WN**	-	-
417.1410 (11.00 min)	AL, CN, HN, PI, PN, PS, PT, SS, WN*	-	-
239.0565 (15.20 min)	AL, CN, HN, PI, PN, PS, PT, SS, WN***	-	-
415.1617 (18.28 min)	AL, WN* / CN, HN, PI, PN, PS**	-	PT, SS
527.1417 (20.31 min)	AL, CN, HN, PI, PN, PS, PT, SS, WN*	-	-
PEANUT			
Molecular feature	**Significantly higher than**	**Significantly lower than**	**No significant differences**
149.0090 (2.54 min)	HN ** / AL, CN, MN, PI, PS, PT, SS, WN***	-	-

295.0455 (12.50 min)	AL, CN, HN, MN, PI, PS, PT, SS, WN***	-	-
431.1553 (12.95 min)	AL, CN, HN, MN, PI, PS, PT, SS, WN**	-	-
293.1142 (13.60 min)	AL, CN, HN, MN, PI, PS, PT, SS, WN***	-	-
295.0454 (13.63 min)	AL, CN, HN, MN, PI, PS, PT, SS, WN***	-	-
325.0563 (14.42 min)	AL, CN, HN, MN, PI, PS, PT, SS, WN***	-	-
250.0719 (14.90 min)	AL, CN, HN, MN, PI, PS, PT, SS, WN***	-	-
400.0672 (15.55 min)	AL, CN, HN, MN, PI, PS, PT, SS, WN***	-	-
295.0455 (16.22 min)	AL, CN, HN, MN, PI, PS, PT, SS, WN***	-	-
575.1975 (17.70 min)	AL, CN, HN, MN, PI, PS, PT, SS, WN**	-	-
501.1513 (17.83 min)	AL, CN, HN, MN, PI, PS, PT, SS, WN**	-	-
473.1658 (18.09 min)	AL, CN, HN, MN, PI, PS, PT, SS, WN**	-	-
579.1729 (18.89 min)	AL, CN, HN, MN, PI, PS, PT, SS, WN**	-	-
163.0400 (18.92 min)	AL, CN, HN, MN, PI, PS, PT, SS, WN***	-	-
245.0929 (19.31 min)	AL, CN, HN, MN, PI, PS, PT, SS, WN***	-	-
279.0507 (19.33 min)	AL, CN, HN, MN, PI, PS, PT, SS, WN***	-	-
248.0926 (19.46 min)	AL, CN, HN, MN, PI, PS, PT, SS, WN***	-	-
501.1038 (21.97 min)	AL, CN, HN, MN, PI, PS, PT, SS, WN***	-	-
441.0824 (22.19 min)	AL, CN, HN, MN, PI, PS, PT, SS, WN***	-	-
471.0927 (22.27 min)	AL, CN, HN, MN, PI, PS, PT, SS, WN***	-	-
803.3337 (22.54 min)	AL, CN, HN, MN, PI, PS, PT, SS, WN**	-	-
501.1036 (22.62 min)	AL, CN, HN, MN, PI, PS, PT, SS, WN***	-	-
137.0244 (22.73 min)	PS** / AL, CN, HN, MN, PI, PT, SS, WN***	-	-
PINE NUT			
Molecular feature	**Significantly higher than**	**Significantly lower than**	**No significant differences**
262.0571 (2.62 min)	PT* / AL, CN, HN, MN, PN, PS, SS, WN**	-	-
329.0884 (8.89 min)	MN* / AL, CN, HN, PN, PS, PT, SS, WN**	-	-
329.0884 (9.90 min)	MN* / AL, CN, HN, PN, PS, PT, SS, WN**	-	-
451.1366 (12.74 min)	AL, CN, HN, MN, PN, PS, PT, SS, WN*	-	-
347.1717 (13.15 min)	AL, CN, HN, MN, PN, PS, PT, SS, WN*	-	-
361.1513 (13.62 min)	AL, CN, HN, MN, PN, PS, PT, SS, WN**	-	-
477.2346 (24.83 min)	AL, CN, HN, MN, PN, PS, PT, SS, WN*	-	-

PISTACHIO			
Molecular feature	**Significantly higher than**	**Significantly lower than**	**No significant differences**
179.0561 (2.37 min)	AL, CN, HN, MN, PN, SS*	-	PI, PS, WN
323.0284 (2.71 min)	AL, CN, MN, PI, PN, PS, SS, WN** / HN***	-	-
133.0143 (2.91 min)	AL, CN, HN, MN, PI, PN, PS, SS, WN**	-	-
191.0198 (2.92 min)	AL, HN, MN, PI, PN, PS, SS, WN**	-	CN
346.0557 (3.09 min)	AL, CN, HN, MN, PI, PN, PS, SS, WN**	-	-
153.0194 (10.44 min)	HN* / CN, SS** / AL, MN, PI, PN, PS, WN***	-	-
431.1557 (12.23 min)	AL, CN, HN, PI, PN, PS, SS, WN**	-	MN
451.1357 (12.40 min)	AL, CN, HN, MN, PI, PN, PS, SS, WN**	-	-
451.1357 (12.82 min)	PI* / AL, CN, HN, MN, PN, PS, SS, WN**	-	-
395.1558 (13.19 min)	AL, CN, HN, MN, PI, PN, PS, SS, WN**	-	-
395.1558 (13.74 min)	AL, CN, HN, MN, PI, PN, PS, SS, WN**	-	-
323.1347 (14.02 min)	AL, CN, HN, MN, PI, PN, PS, SS, WN***	-	-
323.1346 (14.51 min)	SS** / AL, CN, HN, MN, PI, PN, PS, WN***	-	-
463.0881 (23.03 min)	CN* / AL, HN, MN, PI, PN, PS, SS, WN**	-	-
507.2439 (24.26 min)	AL, CN, HN, MN, PI, PN, PS, SS, WN**	-	-
PUMPKIN SEED			
Molecular feature	**Significantly higher than**	**Significantly lower than**	**No significant differences**
503.1623 (2.52 min)	MN*	CN, SS** / PT***	AL, HN, PI, PN, WN
209.0670 (2.54 min)	SS* / AL, CN, HN, MN, PI, PN, WN***	PT**	-
337.0784 (3.06 min)	AL, CN, HN, MN, PI, SS**	-	PN, PT, WN
213.0885 (3.40 min)	AL, CN, HN, MN, PI, PN, PT, SS, WN***	-	-
282.0848 (3.49 min)	CN, PN** / AL, HN, MN, PI, WN***	SS**	PT
431.1566 (10.32 min)	MN* / AL, CN, HN, PI, PN, PT, SS, WN***	-	-
673.3090 (22.22 min)	AL, CN, HN, MN, PI, PN, PT, SS, WN***	-	-
SUNFLOWER SEED			
Molecular feature	**Significantly higher than**	**Significantly lower than**	**No significant differences**
209.0567 (3.41 min)	AL, CN, HN, MN, PI, PN, PS, PT, WN***	-	-
575.2607 (8.48 min)	AL, CN, HN, MN, PI, PN, PS, PT, WN**	-	-
575.2614 (9.32 min)	AL, CN, HN, MN, PI, PN, PS, PT, WN**	-	-
353.0875 (14.22 min)	AL, CN, HN, MN, PI, PN, PS, PT, WN**	-	-
353.0875 (15.79 min)	AL, CN, HN, MN, PI, PN, PS, PT, WN**	-	-

491.1192 (16.74 min)	AL, CN, HN, MN, PI, PN, PS, PT, WN**	-	-
641.1510 (18.47 min)	AL, CN, HN, MN, PI, PN, PS, PT, WN**	-	-
542.1299 (18.55 min)	AL, CN, HN, MN, PI, PN, PS, PT, WN**	-	-
491.1190 (18.64 min)	AL, CN, HN, MN, PI, PN, PS, PT, WN**	-	-
542.1300 (19.95 min)	AL, CN, HN, MN, PI, PN, PS, PT, WN**	-	-
135.0452 (20.49 min)	AL, CN, HN, MN, PI, PN, PS, PT, WN***	-	-
515.1190 (21.04 min)	AL, CN, HN, MN, PI, PN, PS, PT, WN**	-	-
563.1766 (21.43 min)	AL, CN, HN, MN, PI, PN, PS, PT, WN**	-	-
563.1765 (23.33 min)	AL, CN, HN, MN, PI, PN, PS, PT, WN**	-	-
515.1190 (23.53 min)	AL, CN, HN, MN, PI, PN, PS, PT, WN**	-	-
WALNUT			
Molecular feature	**Significantly higher than**	**Significantly lower than**	**No significant differences**
785.0854 (10.35 min)	AL, CN, HN, MN, PI, PN, PS, PT, SS**	-	-
633.0744 (11.82 min)	AL, CN, HN, MN, PI, PN, PS, PT, SS*	-	-
305.0783 (13.78 min)	AL, CN, HN, MN, PI, PN, PS, PT, SS**	-	-
259.1191 (16.11 min)	AL, CN, HN, MN, PI, PN, PS, PT, SS**	-	-
259.1191 (17.04 min)	AL, CN, HN, MN, PI, PN, PS, PT, SS**	-	-
289.0834 (17.30 min)	AL, CN, HN, MN, PI, PN, PS, PT, SS**	-	-
261.1346 (19.64 min)	AL, CN, HN, MN, PI, PN, PS, PT, SS**	-	-
433.0415 (22.20 min)	AL, CN, HN, MN, PI, PN, PS, PT, SS**	-	-
403.1613 (22.59 min)	AL, CN, HN, MN, PI, PN, PS, PT, SS**	-	-
592.2042 (23.05 min)	AL, CN, HN, MN, PI, PN, PS, PT, SS*	-	-
371.2076 (23.67 min)	AL, CN, HN, MN, PI, PN, PS, PT, SS**	-	-
300.9991 (23.81 min)	AL, CN, HN, MN, PI, PN, PS, PT, SS**	-	-
413.2183 (25.10 min)	AL, CN, HN, MN, PI, PN, PS, PT, SS**	-	-
413.2184 (25.67 min)	AL, CN, HN, MN, PI, PN, PS, PT, SS*	-	-

AL: almond; CN: cashew nut; HN: hazelnut; MN: macadamia nut; PI: pine nut; PN: peanut; PS: pumpkin seed; PT: pistachio; SS: sunflower seed; WN: walnut

* $p < 1 \cdot 10^{-2}$; ** $p < 1 \cdot 10^{-3}$; *** $p < 1 \cdot 10^{-5}$

Table S3. Summary of the significance of the differences in the discriminant molecular features content between nut types.

ALMOND			
Molecular feature	**Natural *vs*. Toasted**	**Natural *vs*. Fried**	**Fried *vs*. Toasted**
386.9390 (2.03 min)	-	-	-
165.0403 (2.35 min)	Higher in natural**	-	-
209.0300 (2.38 min)	-	-	-
341.1083 (2.48 min)	Higher in natural*	-	-
683.2243 (2.48 min)	Higher in natural*	-	-
191.0564 (2.51 min)	Higher in toasted*	-	Higher in toasted*
533.1718 (2.51 min)	-	-	-
133.0143 (2.91 min)	-	-	-
191.0196 (3.39 min)	-	-	-
413.1658 (3.46 min)	-	-	-
456.1506 (13.52 min)	-	-	-
511.2538 (25.10 min)	-	-	-
HAZELNUT			
Molecular feature	**Natural *vs*. Toasted**		
131.0462 (2.27 min)	-		
223.0459 (2.44 min)	Higher in natural*		
262.0565 (2.48 min)	Higher in natural*		
205.0351 (2.57 min)	Higher in toasted**		
203.0197 (3.49 min)	-		
368.0984 (7.31 min)	-		
368.0985 (8.39 min)	-		
443.1920 (12.39 min)	-		
679.1813 (12.62 min)	-		
443.1920 (13.04 min)	-		
679.1809 (13.25 min)	-		
514.1561 (13.99 min)	-		
679.1810 (14.22 min)	-		
541.1458 (20.91 min)	-		
540.1719 (21.05 min)	-		
259.1911 (34.26 min)	-		
PEANUT			
Molecular feature	**Fried *vs*. Toasted**		
149.0090 (2.54 min)	-		
295.0455 (12.50 min)	-		
431.1553 (12.95 min)	-		
293.1142 (13.60 min)	-		
295.0454 (13.63 min)	-		
325.0563 (14.42 min)	-		
250.0719 (14.90 min)	-		
400.0672 (15.55 min)	-		
295.0455 (16.22 min)	-		
575.1975 (17.70 min)	-		
501.1513 (17.83 min)	-		
473.1658 (18.09 min)	-		
579.1729 (18.89 min)	-		
163.0400 (18.92 min)	-		
245.0929 (19.31 min)	-		
279.0507 (19.33 min)	-		
248.0926 (19.46 min)	-		
501.1038 (21.97 min)	-		
441.0824 (22.19 min)	-		
471.0927 (22.27 min)	-		
803.3337 (22.54 min)	-		

501.1036 (22.62 min)	-
137.0244 (22.73 min)	-
PUMPKIN SEED	
Molecular feature	**Natural *vs.* Toasted**
503.1623 (2.52 min)	Higher in natural*
209.0670 (2.54 min)	Higher in natural***
337.0784 (3.06 min)	-
213.0885 (3.40 min)	-
282.0848 (3.49 min)	-
431.1566 (10.32 min)	-
673.3090 (22.22 min)	-

* $p < 1 \cdot 10^{-2}$; ** $p < 1 \cdot 10^{-3}$; *** $p < 1 \cdot 10^{-4}$

Table S4. Summary of the significance of the differences in the discriminant molecular features content between the different thermal treatments of almond, hazelnut, peanut, and pumpkin seed samples.

3.2.6. PUBLICATION IX

FIA–HRMS fingerprinting subjected to chemometrics as a valuable tool to address food classification and authentication: Application to red wine, paprika, and vegetable oil samples.

Campmajó, G.; Saurina, J.; Núñez, O.

Food Chemistry. 2022, 373, 131491.

Food Chemistry 373 (2022) 131491

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

FIA− RMS fingerprinting subjected to chemometrics as a valuable tool to address food classification and authentication: Application to red wine, paprika, and vegetable oil samples

Guillem Campmajó [a,b,*], Javier Saurina [a,b], Oscar Núñez [a,b]

[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

ARTICLE INFO

Keywords:
FIA− RMS
Chemometrics
Data fusion
Red wine
Paprika
Olive oil

ABSTRACT

The rise of food fraud practices, affecting a wide variety of goods and their specific characteristics (e.g., quality or geographical origin), demands rapid high-throughput analytical approaches to ensure consumers protection. In this context, this study assesses flow injection analysis coupled to high-resolution mass spectrometry (FIA− RMS), using a fingerprinting approach and combined with chemometrics, to address four food authentication issues: (i) the geographical origin of three Spanish red wines, (ii) the geographical origin of three European paprikas, (iii) the distinction of olive oil from other vegetable oils and (iv) the assessment of its quality category. In each case, negative and positive ionisation FIA− RMS fingerprints, and two different data fusion strategies, were evaluated. After external validation, excellent classification accuracies were reached. Moreover, high-resolution mass spectrometry (RMS) allowed sample matrices characterisation by the putative identification of the most common ions.

1. Introduction

Globalisation has notoriously expanded international trade, increasing the number of participants between the production and consumption in the food chain. In this context, food fraud, which encompasses a sort of intentional manipulation practices in food products (i.e., adulteration, mislabelling, grey market, and counterfeit) aiming an economic gain (Morin & Lees, 2018), has become of great concern among consumers, food businesses, the scientific community, and government administrations. Because of the economic purpose behind food fraud, its likelihood is generally estimated using supply-and-demand and financial indicators influenced by macroeconomic trends and directly affected by unexpected situations such as the Suez Canal blockage or the CO ID-19 pandemic (Points, Manning, & Group, 2020). Moreover, goods well-valued for specific labelled particularities (e.g., geographical origin or production system), which enhance their reputation and increase their price, are likely to be affected by fraudulent practices since the difference between authentic and non-authentic products is difficult to measure.

Chromatographic and related techniques —such as capillary electrophoresis (CE), gas chromatography (GC), and liquid chromatography (LC)— with spectroscopic detection or coupled to mass spectrometry (MS), and combined with chemometrics, have proven excellent capacity to address complex food authentication issues through fingerprinting strategies (Cuadros-Rodríguez, Ruiz-Samblás, alverde-Som, Pérez-Castaño, & González-Casado, 2016; Medina, Perestrelo, Silva, Pereira, & Câmara, 2019). owever, the need for more rapid high-throughput analytical approaches, minimising sample analysis time and even costs, has focused the attention on direct MS techniques (Ibáñez, Simó, García-Cañas, Acunha, & Cifuentes, 2015). In this line, both ambient mass spectrometry (AMS) and flow injection analysis coupled to mass spectrometry (FIA−MS) seem to be potential alternatives to non-targeted chromatographic methods.

AMS comprises several techniques, mainly spray- or plasma-based —such as desorption electrospray ionisation (DESI) (Takáts, iseman, Gologan, & Cooks, 2004) and direct analysis in real-time (DART) (Cody, Laramée, & Durst, 2005), respectively—, that provide direct desorption/ionisation of analytes from the native sample or with minimal sample

* Corresponding author at: Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, Martí i Franquès 1-11, E08028 Barcelona, Spain.
E-mail address: gcampmajo@ub.edu (G. Campmajó).

https://doi.org/10.1016/j.foodchem.2021.131491
Received 21 June 2021; Received in revised form 26 October 2021; Accepted 26 October 2021
Available online 28 October 2021

treatment. Despite the significant number of advances that have been made in this field in the last decade and its advantages (such as real-time, in situ, and in vivo analysis), AMS techniques still lack three crucial aspects in fingerprinting approaches: good reproducibility, sensitivity, and wide molecular coverage (Kuo, Dutkiewicz, Pei, & su, 2020). Instead, FIA–MS, which is based on injecting a small volume of a liquid sample or a sample extract into an organic phase continuous stream that carries the sample bolus up to the mass spectrometer ion source, provides satisfactory reproducibility because of modern automatic autosamplers and injectors precision and repeatability (Gachumi, Purves, opf, & El-Aneed, 2020) and allows better ionisation efficiency through any atmospheric pressure ionisation (API) source —mostly electrospray ionisation (ESI)— than AMS. Therefore, FIA–MS offers a balance between chromatographic-MS and AMS techniques regarding analytical capability and sample analysis throughput (Nanita & Kaldon, 2016).

Several analytical methods aiming to solve food authentication issues by FIA–MS fingerprinting, combined with chemometrics, have been developed in the last years. In this line, nominal mass fingerprints acquired by flow injection analysis coupled to low-resolution mass spectrometry (FIA–LRMS), mainly based on ion trap technology MS instruments —i.e., ion trap (IT) and linear ion trap (LIT)— due to their higher sensitivity in full-scan MS mode than quadrupole mass analysers, have proved their potential in particular applications, such as the organic and conventional sage sample differentiation (Gao, Lu, Sheng, Chen, & u, 2013) or cinnamon species classification (Chen, Sun, & Ford, 2014). In contrast, exact mass fingerprints are obtained by flow injection analysis coupled to high-resolution mass spectrometry (FIA– RMS), using time-of-flight- (TOF) or Orbitrap-based mass analysers, which present a maximum resolving power up to 50,000 and 500,000 full-width at half maximum (F M), respectively (Rubert, Zachariasova, & ajslova, 2015). Thus, FIA– RMS leads to richer fingerprints, where near-isobaric compound signals are well-resolved, enhancing selectivity and providing better molecular coverage than FIA–LRMS. For instance, it has been already successfully applied to address oregano (Gao et al., 2014) and lettuce (Sun et al., 2018) production system or milk adulteration detection (Du et al., 2018). Moreover, taking advantage of its accurate mass measurements and isotopic abundance ratios, molecular *formulae* of specific ions can be determined and compared to publicly accessible databases for putative compound identification.

This study aimed to demonstrate FIA– RMS suitability to address certain food authenticity issues through a fingerprinting approach and its combination to principal component analysis (PCA), partial least squares regression-discriminant analysis (PLS-DA), and soft independent modelling of class analogy (SIMCA). Thus, the geographical origin of three Spanish red wines (*Catalunya*, *La Rioja*, and *Castilla y León*) and three European paprikas (*La Vera*, *Murcia*, and the Czech Republic), as well as the distinction of olive oil from other vegetable oils and the assessment of its quality category, were evaluated.

2. Materials and methods

2.1. Reagents and solutions

For the sample treatment, purified water was obtained using an Elix® 3 coupled to a Milli-Q® system (Millipore Corporation, Bedford, MA, USA) and filtered through a 0.22-µm nylon membrane; hexane and formic acid (96%) were provided from Merck (Darmstadt, Germany); U PLC-supergradient acetonitrile was from Panreac (Castellar del allès, Spain); and ethanol was purchased from R International Eurolab S. L. (Barcelona, Spain). Instead, for the FIA– RMS, LC–MS grade water and acetonitrile were from Merck.

2.2. Instrumentation

FIA was performed using an ultra-high-performance liquid chromatography (U PLC) system equipped with an Accela 1250 quaternary pump and an Accela autosampler (Thermo Fisher Scientific, San Jose, CA, USA). The sample injection volume was 10 µL. The carrier consisted of a 50:50 (*v*/*v*) mix, composed of water acidified with 0.1% formic acid (*v*/*v*) and acetonitrile, and was pumped isocratically at 150 µL·min^{-1} for 1.5 min.

The U PLC system was coupled to a hybrid quadrupole-Orbitrap (Q-Orbitrap) mass spectrometer (Q-Exactive Orbitrap, Thermo Fisher Scientific) equipped with a heated electrospray ionisation source (-ESI II) operating in both negative and positive ionisation modes. The -ESI source was set in an off-axis position to prevent and minimise mass spectrometer contamination. Nitrogen with a purity of 99.98%, purchased from Linde (Barcelona, Spain), was used for the ESI sheath, sweep, and auxiliary gas at flow rates of 40, 0, and 12 a.u. (arbitrary units), respectively. Moreover, the vaporiser temperature was set at 250 °C, the capillary temperature at 350 °C, the spray voltage at± 3.0 k (depending on the ionisation mode), and the S-lens RF level at 50 . The Q-Orbitrap mass analyser worked in full-scan MS mode, with an *m*/*z* range from 100 to 1500, at a mass resolution of 70,000 F M at *m*/*z* 200. Besides, an automatic gain control (AGC) target of 3.0×10^6, which is the number of ions to fill the C-Trap, and a maximum injection time of 100 ms, were established. Simultaneously to the full-scan MS mode, data-dependent scan mode (ddMS^2) was also performed with an intensity threshold of 1.0×10^5, a fixed first *m*/*z* of 50 for the registered product ion scan range, a quadrupole isolation window of 0.5 *m*/*z*, and applying stepped normalised collision energies (NCE) of 17.5, 35.0, and 52.5 e for ion fragmentation. Besides, in this event acquisition mode, a mass resolution of 17,500 F M at *m*/*z* 200, an AGC target value of 5.0×10^5, and a maximum injection time of 100 ms were also set.

The Q-Orbitrap system was tuned and calibrated every three days, using commercially available calibration solutions for both negative and positive ion modes (Thermo Fisher Scientific). Moreover, the calibur software *v* 4.1 (Thermo Fisher Scientific) was used to control the LC–MS system and acquire and process data.

2.3. Samples and sample treatment

In this study, three different sample sets (red wine, paprika, and olive and other vegetable oils), detailed in the present Section, were under evaluation by the proposed FIA– RMS method. In all their corresponding sample sequences, in order to ensure the quality of the results avoiding and controlling systematic errors and cross-contamination, a quality control (QC) sample —constructed by pooling equal aliquots of each sample of the set— and an extracting solvent blank were injected at the beginning and after every ten sample injections. Besides, samples were also randomly injected to minimise the effect of instrumental drifts on the chemometric models.

2.3.1. Red wine

A set of 94 red wine samples from three Spanish areas —50 from *Catalunya*, 25 from *La Rioja*, and 19 from *Castilla y León*— encompassing 15 different Protected Designation of Origin (PDO) labels (*Bierzo*, *Catalunya*, *Conca de Barberà*, *Costers del Segre*, *Empordà*, *Montsant*, *Penedès*, *Pla de Bages*, *Priorat*, *Ribera del Duero*, *Rioja*, *Tarragona*, *Terra Alta*, *Tierra de Castilla*, and *Toro*) and ten production years (1996, 2002, 2006, 2007, 2009, 2010, 2011, 2012, 2013, and 2014), and made from various grape varieties, were analysed. Prior to FIA– RMS analysis, samples were filtered with a 0.22-µm nylon filter (Scharlab, Sentmenat, Spain).

2.3.2. Paprika

One hundred eleven paprika samples, including different geographical origins —72 with *La Vera* PDO (Spain), 24 with *Murcia* PDO (Spain), and 15 from the Czech Republic (their specific region was

not labelled)— and types (hot, bittersweet, and sweet), were directly purchased from their production companies or bought in Spanish or Czech commercial supermarkets.

Regarding the sample treatment, a previously developed procedure was followed (Cetó et al., 2018). Briefly, 0.3 g of the sample were subjected to solid–liquid extraction (SLE) with 3 mL of water:acetonitrile (20:80, *v/v*) mix. After stirring in a ortex (Stuart, Stone, United Kingdom) for 1 min, sonicating (5510 Branson ultrasonic bath, ampton, N , USA) for 15 min, and centrifuging (ROTANTA 460 RS Centrifuge, ettich, Germany) for 30 min at 4500 rpm, the resulting supernatant extract was filtered, using a 0.22-µm nylon filter and kept at 4 °C in a glass injection vial until its analysis.

2.3.3. Olive and other vegetable oil

In this study, a total of 85 vegetable oil samples —46 olive oils, 15 sunflower oils, 6 corn oils, 6 soy oils, and 12 oils produced from mixtures of seeds (6 sunflower/corn oils and 6 sunflower/soy oils)— from various trademarks and purchased from Barcelona markets were analysed. Moreover, among the olive oil sample set, 12 were refined olive oils (OO), 4 virgin olive oils (OO), and 30 extra-virgin olive oils (E OO).

The employed sample treatment was based on a previously described method (Gosetti, Bolfi, Manfredi, Calabrese, & Marengo, 2015) with slight modifications. First, liquid–liquid extraction with low-temperature partition (LLE-LTP), using ethanol:water (70:30, *v/v*) as the extracting solvent, was carried out. Thus, in a 15 mL-polytetrafluoroethylene (PTFE) tube (Serviquimia, Barcelona, Spain), 2.00 g of oil sample were extracted by stirring for 2 min in a ortex in 2 mL of the extracting solvent. After centrifugation for 5 min at 3500 rpm, the mixture was frozen for 24 h at_18 °C. Then, the resulting supernatant extract was transferred into another PTFE tube for a defatting step with 2 mL of hexane, also by stirring in a ortex followed by centrifugation for 5 min at 3500 rpm. Finally, the aqueous ethanolic sample extract was filtered with a 0.22-µm nylon filter and stored at −18 °C in a 2-mL glass injection vial until FIA– RMS analysis.

2.4. Data analysis

2.4.1. Data matri construction

Raw data results were submitted to exact mass detection, chromatogram builder, isotopic peak grouper, and join aligner, using the mzMine 2.53 software (Pluskal, Castillo, illar-Briones, & Orešič, 2010). First, the exact mass detection step generated mass lists for each scan acquired in a sample, considering a noise level of 1.0×10^5. Then, the chromatogram builder allowed the joining of exact mass signals found in contiguous scans in a sample, establishing a peak time range of 0.05 – 0.40 min, an *m/z* tolerance of 5 ppm, and an intensity threshold of 1.0×10^5. After this, isotopes were removed, considering that the most representative isotope was the most intense and setting an *m/z* tolerance of 5 ppm. Finally, the join aligner allowed matching of exact masses detected across samples, establishing a mass tolerance of 5 ppm, a peak time tolerance of 0.35 min (the whole time range under evaluation), 95% of weight for *m/z*, and a 0% of weight for time. At the end of this workflow, a data matrix was constructed containing FIA– RMS fingerprints of the studied samples: samples× variables, where variables consisted of ion signal intensity values as a function of *m/z*. Moreover, to reduce the matrix dimensions, molecular features were filtered and only were selected those with a relative standard deviation (RSD, %) lower than 20% in the signals of the QC samples injected during the sample sequence.

2.4.2. Chemometric analysis

The obtained FIA– RMS fingerprints were then subjected to PCA, PLS-DA, and SIMCA, which were performed using Solo 8.6 chemometrics software from Eigenvector Research (Manson, A, USA). Details of the theoretical background of these chemometric methods are addressed elsewhere (Massart et al., 1997; old, 1976).

PCA relies on the concentration of the dataset's relevant information, originally arranged in the -matrix containing sample FIA– RMS fingerprints, into a reduced number of principal components (PCs). In this study, it allowed an exploratory chemometric analysis to evaluate QC sample behaviour (i.e., QC samples forming a compact group in the PCA scores plot indicates the absence of systematic errors during the sample injection and validates the chemometric results) and sample trends and groups.

Instead, PLS-DA, which uses the same -matrix as PCA, assigns each given sample into a numerically encoded class in the -matrix, depending on predefined sample characteristics (e.g., geographical or botanical origin). In this case, a reduced number of latent variables (L s) contain the most relevant information that links both matrices. The most appropriate number of L s to build the PLS-DA models was established at the first significant minimum point of the enetian blinds cross-validation (C) error. Besides, considering the complexity of the studied issues, where various sample classes were assessed, the hierarchical model builder (MB) was used, segregating the complete classification in a consecutive combination of two-input class PLS-DA models (classification decision tree). To evaluate and validate the predictive ability of the whole classificatory chemometric model, 60% of samples were randomly stratified as the calibration set and the remaining 40% as the external validation set. In this line, the performance of the developed classificatory method was checked through each class sensitivity (capability to detect true positives, i.e., samples belonging to a given class that have been correctly assigned) and specificity (capability to detect true negatives, i.e., samples that do not belong to a given class correctly assigned as negative), as well as the overall accuracy of the model (well-classified and misclassified sample ratio).

Finally, SIMCA is based on the definition of a target class by a PCs subspace. In this study, one-class SIMCA was applied to sample authentication, and therefore, each SIMCA model was composed of a unique PCA submodel corresponding to a specific sample class. Then, since it consists of a distance-based method of class modelling, when a new sample is projected into the model, its class membership is assessed according to its distance from the PCA submodel —calculated from the reduced Q residuals and otelling T^2 valu (normalised to 95% confidence limit) and combined using $d_i = \sqrt{\quad}$ being i the $_{red,i}$ index of each given unknown sample to be classified— and a previously established decision threshold. The latter was optimised in each case, maximising the performance of SIMCA in the calibration step by reaching the minimum error. Moreover, same external validation as in the PLS-DA study was carried out.

It should be pointed out that for each sample set, four different -matrices were used as chemical descriptors in the PLS-DA and SIMCA models: FIA– RMS fingerprints obtained with negative ionisation, with positive ionisation, and using a low-level (LLDF) or a mid-level data fusion (MLDF) strategy (Borràs et al., 2015). The LLDF -matrix concatenated both negative and positive ionisation FIA– RMS data. Instead, the MLDF only contained ten variables per each PLS-DA model involved in the classification decision tree. These variables corresponded to those presenting the highest selectivity ratio among the 50 with the highest variable importance in projection (IP) values obtained in the LLDF loadings of each PLS-DA model involved in the classification. In all cases, data was autoscaled to provide the same weight to each variable by suppressing differences in their magnitude and amplitude scales.

3. Results and discussion

3.1. FIA–HRMS fingerprint characterisation

Before the chemometric analysis, the obtained FIA– RMS fingerprints were visually inspected, and some of the most intense ions were

putatively identified to assess sample matrix characterisation, taking advantage of high-resolution mass spectrometry (RMS) capabilities. In this line, Table S1 summarises the putative identification of some of the most characteristic ions found in the food matrices under study, following Schymanski et al. RMS identification levels (Schymanski et al., 2014). Public databases, containing MS^2 data —mzCloud (igh-Chem LLC, Bratislava, Slovakia), Metlin (Smith et al., 2005), and The uman Metabolome Database (ishart et al., 2018)— and polyphenolic content in food, such as Phenol-Explorer (Rothwell et al., 2013), were consulted. Criteria followed in this process were established as follows: 5 ppm of exact mass tolerance, >90% of isotopic pattern fit, and MS^2 data agreement.

The negative and positive ionisation FIA− RMS spectra of a *Catalunya*, *La Rioja*, and *Castilla y León* red wine sample are shown in Fig. S1. A priori, no noticeable interregional differences could be highlighted since the most intense ions were commonly found in all samples without following a characteristic pattern due to geographical origin. Among the putatively identified compounds, negative ionisation FIA− RMS fingerprints contained certain molecules known to be found in wine, such as several organic acids (being tartaric acid the base peak) (Ivanova-Petropulos et al., 2018), hydroxybenzoic (e.g., gallic acid) and hydroxycinnamic acids (e.g., caffeic and caffeoyl tartaric acid) (Gutiérrez-Escobar, Aliaño-González, & Cantos- illar, 2021), and monosaccharide and sugar-related compounds. Instead, amino acids and choline and furan compounds were found in positive ionisation spectra. Several anthocyanins, which influence the wine colour (Garrido & Borges, 2013), were also found. Moreover, some coumarins, released from wood into the wine during the maturation stage (roboňová & Sádecká, 2020), were detected in their $[M+Na]^+$ form.

Regarding paprika samples, as an example, Fig. 1 depicts typical negative and positive ionisation FIA− RMS fingerprints for hot *La Vera*, *Murcia*, and the Czech Republic samples. At first glance, *La Vera* samples

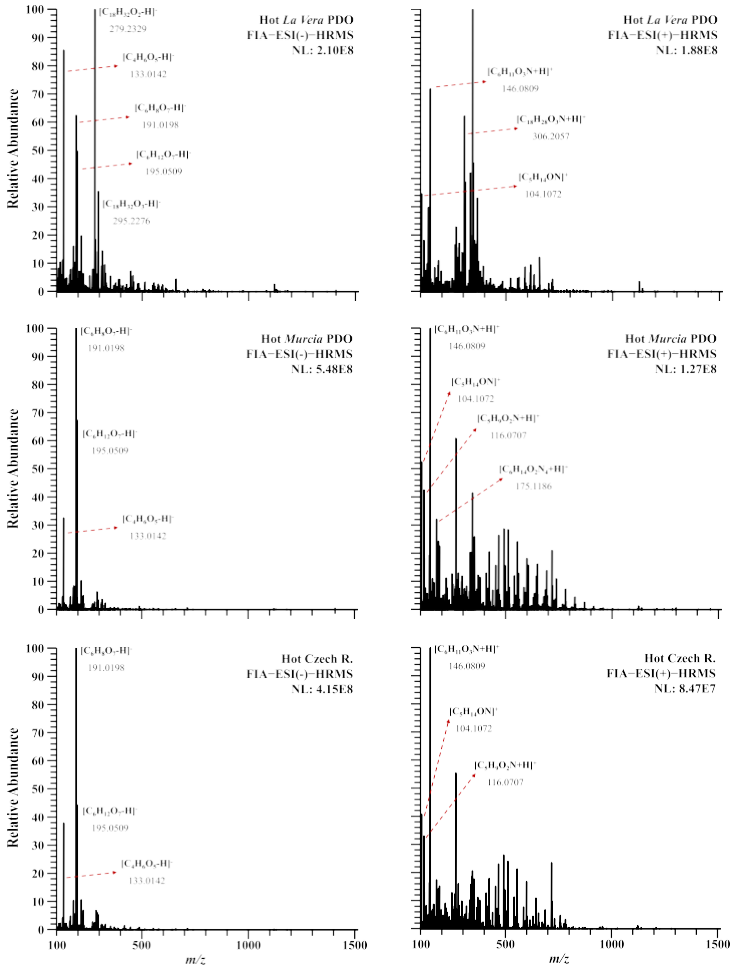


Fig. 1. Negative and positive ionisation FIA− RMS fingerprints obtained for a *La Vera*, *Murcia*, and the Czech Republic paprika sample.

presented distinctive fingerprints in the negative mode comparing to the remaining samples. For instance, the *m/z* 279.2329, corresponding to the deprotonated molecule of linoleic acid, one of the major fatty acids found in *Jaranda* and *Jariza Capsicum annuum* L. varieties (Pérez-Gálvez, Garrido-Fernández, Mínguez-Mosquera, Lozano-Ruiz, & Montero-de-Espinosa, 1999) that are used in *La Vera* paprika production, was particularly intense in their negative ionisation spectra. In addition, these paprika fingerprints, reproducible among samples belonging to the same geographical origin, contained common compounds such as organic acids or certain mono- and polyunsaturated fatty acids. *La Vera* samples particularity was also highlighted in positive ionisation FIA− RMS fingerprints that included many signals in the *m/z* range from 100 to 800. owever, in this case, spectra were slightly altered in *Murcia* and the Czech Republic samples according to the paprika type, while significantly modified in *La Vera* ones. In this line, compounds such as capsaicinoids may be related to these differences (Arrizabalaga-Larrañaga et al., 2021). Moreover, several amino acids and other compounds, including choline, 6-(hydroxymethyl)pyridin-3-ol, tropine, and 4-hydroxy-1-methyl-2-pyrrolidine carboxylic acid, were also detected in the positive ionisation FIA− RMS spectra as reported in Table S1.

Finally, both negative and positive ionisation olive oil FIA− RMS fingerprints were noticeably different from the remaining vegetable oil ones. Instead, as shown in Fig. S2, more similarities were found in olive oil spectra depending on their quality grade. Besides, while OO and E OO showed comparable fingerprints among samples belonging to the same group, more variability was observed in OO samples, which may be due to the different percentages of OO added to them for taste improvement. As shown in Table S1, among other compounds, several polyphenolic compounds well-known to be found in olive oil were identified in the analysed ethanolic sample extract. In this line, tyrosols predominated the negative ionisation spectra, although other polyphenols such as luteolin or dihydro-*p*-coumaric acid were also detected (Farré, Picó, & Barceló, 2019). In the positive ionisation mode, tyrosols were detected, forming an adduct with Na.

3.2. Red wine geographical origin classification and authentication

In this study, FIA− RMS fingerprints were proposed as chemical markers to address the geographical origin classification of three Spanish red wines: *Catalunya*, *La Rioja*, and *Castilla y León*. Thus, in a first attempt to evaluate their discriminating ability, an exploratory PCA was performed to both negative and positive ionisation data —104 × 440 and 104 × 972 (samples × variables) dimension data matrices, respectively—, aiming to observe QC sample behaviour as well as

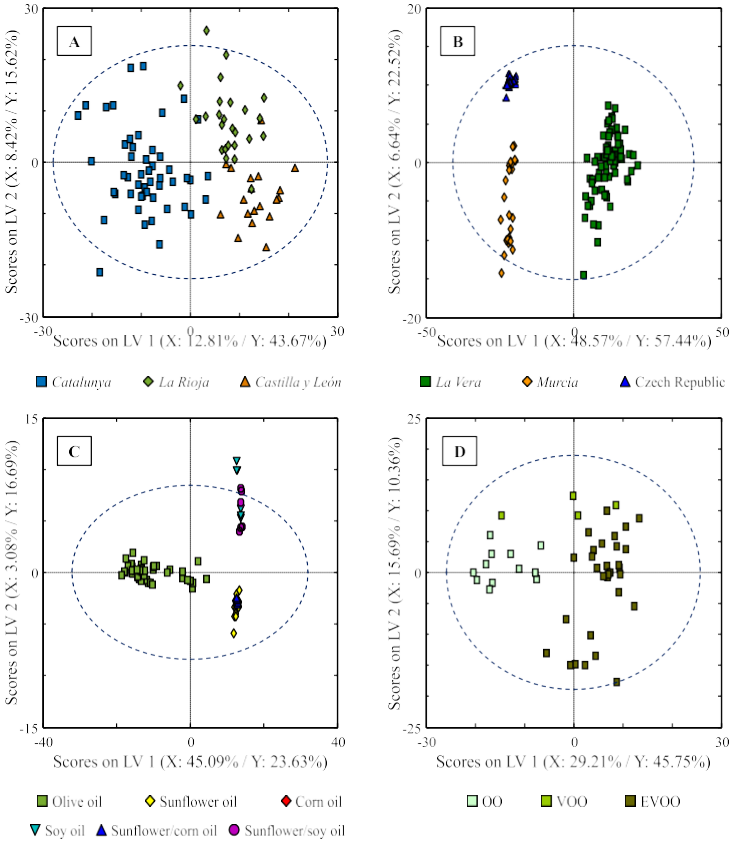


Fig. 2. PLS-DA scores plot of L 1 *vs.* L 2 obtained for: (A) the red wine samples analysed according to their geographical origin, using positive ionisation FIA− RMS fingerprints; (B) the paprika samples analysed according to their geographical origin, using negative ionisation FIA− RMS fingerprints; (C) the olive and vegetable oil samples analysed according to their botanical origin, using negative ionisation FIA− RMS fingerprints; and (D) the olive oil samples analysed according to their quality grade, using negative ionisation FIA− RMS fingerprints. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

sample groups and trends. Similar results were found in both cases. For instance, Fig. S3 shows the PCA scatter plot of scores on the PC2-PC1 (explaining 28.49% of the variance) obtained using positive ionisation data. hile QC samples were jointly located in the centre of the plot, indicating the lack of systematic errors, *Catalunya* samples were clearly distinguished at the bottom of the plot displaying negative PC2 values. Instead, *La Rioja* and *Castilla y León* samples shared more similarities since there was no evident discrimination between them (Fig. S3A). In this line, the representation of samples in the PCA scores plot according to non-*Tempranillo* and *Tempranillo*-based red wines (all *La Rioja* and most *Castilla y León* samples were mainly produced from *Tempranillo* grapes, while most *Catalunya* samples did not) allowed to prove its influence on sample distribution. A trend in the PC2 from non-*Tempranillo*, displaying negative values, to *Tempranillo*-based red wines, displaying positive values, was observed (Fig. S3B).

After the exploratory chemometric analysis, QC samples were excluded from -data matrices (resulting in 9✱ 440 and 94✱ 972 data matrices for negative and positive ionisation, respectively), which were then subjected to PLS-DA using the corresponding -data matrices, indicating sample geographical origin. As expected, the obtained PLS-DA scores plots improved non-supervised chemometric results for both negative and positive ionisation data. For instance, Fig. 2A depicts the scores plot of L 1 *vs*. L 2 obtained when using positive ionisation FIA− RMS data. In this case, six L s explaining the 87.98% -variance were required to build the PLS-DA model, allowing a good sample distribution according to their geographical origin.

In view of these results, a classification decision tree consisting of two consecutive rule nodes —1) *Catalunya vs*. Others and 2) *La Rioja vs*. *Castilla y León*— was proposed to address red wine geographical origin classification. As previously mentioned in Section 2.4.2, negative ionisation, positive ionisation, LLDF, and MLDF FIA− RMS data were tested. In this line, the data matrix dimensions and the number of L s used in each PLS-DA calibration model involved, as well as the resulting external validation classification parameters (class sensitivity, class specificity, and global accuracy), can be found in Table 1. In this context, LLDF FIA− RMS fingerprints provided the best external validation classificatory results with 86.8% accuracy. Contrarily, MLDF data, which contained much less sample information, only reached a 60.5% classification rate.

Instead, considering the suitability of class-modelling chemometric methods in the authentication field (Rodionova, Titova, & Pomerantsev, 2016), SIMCA was proposed to test the capacity of FIA− RMS to generate a characteristic fingerprint for each red wine class. Table 2 summarises the data matrix dimensions, the number of PCs established in each SIMCA model, and the assignation performance after the external validation. Satisfactory overall accuracy results were obtained for the four data matrices used (above 75.4%), although these values were generally slightly below those obtained in the classificatory study with PLS-DA. In fact, only in MLDF, SIMCA provided a better accuracy result than the obtained with the PLS-DA classification decision tree, mainly because of a substantial increase in *La Rioja* sensitivity and *Castilla y León* specificity.

Table 1
Calibration model parameters —data matrix dimensions (samples × variables) and number of L s— for each of the PLS-DA models built in the classificatory studies and corresponding obtained external validation classification results —class sensitivity (%), class specificity (%), and global accuracy (%)—.

RED WINE GE0GRAPHICAL 0RIGIN

	Calibration: model parameters				External validation: classification performance						
	***v*. 0thers**		***v*.**								Accuracy
	Data matrix	LVs	Data matrix	LVs	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	
FIA− RMS (-)	56 × 440	2	26 × 440	4	80.0	94.4	70.0	89.3	87.5	90.0	78.9
FIA− RMS (+)	56 × 972	5	26 × 972	1	94.4	100.0	90.0	92.9	62.5	100.0	84.2
LLDF	56 × 1412	3	26 × 1412	1	90.0	100.0	100.0	89.3	62.5	100.0	86.8
MLDF	56 × 20	2	26 × 20	3	75.0	100.0	20.0	82.1	75.0	76.7	60.5

PAPRIKA GE0GRAPHICAL 0RIGIN

	Calibration: model parameters				External validation: classification performance						
	***V r* *v*. 0thers**		***M rc* *v*. The Czech Republic**		***V r***		***M rc***		**The Czech Republic**		**Accuracy**
	Data matrix	LVs	Data matrix	LVs	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	
FIA− RMS (-)	66 × 553	1	23 × 553	2	100.0	100.0	100.0	100.0	100.0	100.0	100.0
FIA− RMS (+)	66 × 601	2	23 × 601	2	96.6	100.0	100.0	100.0	83.3	100.0	95.6
LLDF	66 × 1154	1	23 × 1154	2	100.0	100.0	100.0	100.0	100.0	100.0	100.0
MLDF	66 × 20	1	23 × 20	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0

0LIVE 0IL

Botanical origin

	Calibration: model parameters		External validation: classification performance		
	0live oil *v*. 0thers		**0live 0il**		**Accuracy**
	Data matrix	LVs	Sensitivity	Specificity	
FIA− RMS (-)	53 × 368	1	100.0	100.0	100.0
FIA− RMS (+)	53 × 739	1	100.0	100.0	100.0
LLDF	53 × 1107	1	100.0	100.0	100.0
MLDF	53 × 10	1	94.4	100.0	96.9

Quality

	Calibration: model parameters		External validation: classification performance				
	EV00 and V00 *v*. 00		**EV00 and V00**		**00**		**Accuracy**
	Data matrix	LVs	Sensitivity	Specificity	Sensitivity	Specificity	
FIA− RMS (-)	27 × 368	1	92.9	100.0	100.0	100.0	94.7
FIA− RMS (+)	27 × 739	3	85.7	100.0	100.0	92.9	89.5
LLDF	27 × 1107	1	92.9	100.0	100.0	92.9	94.7
MLDF	27 × 10	2	92.9	100.0	80.0	92.9	89.5

Table 2
Calibration model parameters —data matrix dimensions (samples × variables) and number of PCs— for each of the SIMCA models built in the class assignation studies and corresponding obtained external validation assignation results —class sensitivity (%), class specificity (%), and global accuracy (%)—.

RED WINE GE0GRAPHICAL 0RIGIN

	Calibration: model parameters				External validation: assignation performance						
	Data matrix	PCs									Accuracy
					Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	
FIA− RMS (-)	56 × 440	6	3	4	65.0	72.2	70.0	89.3	25.0	86.7	75.4
FIA− RMS (+)	56 × 972	4	3	3	75.0	61.1	80.0	92.9	50.0	80.0	77.2
LLDF	56 × 1412	6	3	4	55.0	72.2	70.0	96.4	25.0	90.0	76.3
MLDF	56 × 20	2	4	2	80.0	88.9	50.0	89.3	75.0	93.3	84.2

PAPRIKA GE0GRAPHICAL 0RIGIN

	Calibration: model parameters				External validation: assignation performance						
	Data matrix	PCs			*V r*		*M rc*		The Czech Republic		Accuracy
		V r	*M rc*	The Czech Republic	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	
FIA− RMS (-)	66 × 553	1	2	2	96.6	100.0	100.0	100.0	100.0	100.0	99.3
FIA− RMS (+)	66 × 601	1	2	1	93.1	100.0	70.0	100.0	66.7	100.0	94.8
LLDF	66 × 1154	1	2	1	96.6	100.0	80.0	100.0	83.3	100.0	97.0
MLDF	66 × 20	3	3	2	89.7	100.0	80.0	100.0	50.0	100.0	94.1

0LIVE 0IL

Botanical origin

	Calibration: model parameters		External validation: assignation performance		
	Data matrix	PCs	0live 0il		Accuracy
		0live 0il	Sensitivity	Specificity	
FIA− RMS (-)	53 × 368	2	83.3	100.0	90.6
FIA− RMS (+)	53 × 739	4	77.8	100.0	87.5
LLDF	53 × 1107	1	94.4	100.0	96.9
MLDF	53 × 10	1	94.4	100.0	96.9

Quality

	Calibration: model parameters		External validation: assignation performance		
	Data matrix	PCs	EV00 and V00		Accuracy
		EV00 and V00	Sensitivity	Specificity	
FIA− RMS (-)	27 × 368	3	85.7	100.0	89.5
FIA− RMS (+)	27 × 739	3	78.6	100.0	84.2
LLDF	27 × 1107	3	78.6	100.0	84.2
MLDF	27 × 10	1	85.7	100.0	89.5

3.3. Paprika geographical origin classification and authentication

The geographical origin authentication of paprika was also assessed through FIA− RMS fingerprinting. In this line, three different European paprika samples —*La Vera*, *Murcia*, and the Czech Republic— were analysed in negative and positive ESI modes, providing 123 × 553 and 123 × 601 (samples × variables) dimension data matrices, respectively. The obtained plots of scores of PC1 *vs*. PC2 are presented in Fig. S4. At first glance, QC samples appeared compactly in the centre of the plot, guaranteeing the validity of the obtained chemometric results. As expected after visual inspection of paprika FIA− RMS fingerprints in Section 3.1, *La Vera* samples were manifestly differentiated from the other samples, standing on the right side of both PCA plots. Moreover, in the negative ionisation data PCA scores plot (Fig. S4A), *Murcia* samples were separated according to their type (hot and sweet), being hot samples nearly located to the Czech Republic ones. Instead, *Murcia* and Czech samples slightly overlapped in the positive ionisation data PCA scores plot (Fig. S4B), independently of their type.

After QC exclusion, PLS-DA was applied to both matrices. Excellent geographical origin sample classification was achieved either in negative or positive ionisation modes. For example, the negative ionisation FIA− RMS data scores plot of L 1 *vs*. L 2 (three L s, describing 90.32% of -variance, were used to build the PLS-DA model) is depicted in Fig. 2B. To test and validate the classification ability of the acquired FIA− RMS fingerprints, the following nodes were proposed to build a classification decision tree: 1) *La Vera vs*. Others and 2) *Murcia vs*. the Czech Republic. In this line, the predictive capability of the built PLS-DA models was excellent, as shown in Table 1. The lowest classification rate was 95.6%, obtained with the positive ionisation data matrix, while negative ionisation and data fusion matrices allowed the complete correct classification of the test set samples. Moreover, as shown in Table 2, similar results were obtained when subjecting FIA− RMS fingerprints to SIMCA, proving the ability of the proposed model to authenticate the studied samples. Besides, in both PLS-DA and SIMCA studies, MLDF

provided excellent results as LLDF, indicating that a profiling approach focusing on the adequate specific markers could achieve similar results to a fingerprinting approach, which is in agreement with previous studies (Barbosa, Campmajó, Saurina, Puignou, & Núñez, 2020).

3.4. Olive oil botanical origin and quality classification and authentication

Finally, several vegetable oils described in Section 2.3.3, including olive oil, were analysed through the FIA− RMS fingerprinting method, aiming at both botanical origin and quality olive oil authenticity. Firstly, 94 ×368 and 94 ×739 (samples ×ariables) matrices corresponding to negative and positive ionisation data, respectively, were subjected to PCA. After checking QC sample correct behaviour, excellent discrimination of olive oil in front of the other vegetable oil samples was observed in both cases. Besides, a trend among olive oil samples according to their quality category was also found. For instance, Fig. S5 depicts the PCA scores plot of PC1 *vs*. PC2 (describing 57.24% of the variance) obtained with the negative ionisation FIA− RMS data matrix. In this case, olive oil samples were isolated from the other samples on the right side of the plot, displaying positive PC1 values. Moreover, PC1 also allowed a visual separation of E OO and OO samples, while most of the OO samples were jointly located to E OO.

After the exploratory analysis, supervised classificatory PLS-DA was performed to address olive oil authentication. On the one hand, focusing on the botanical origin authentication issue, negative and positive ionisation data matrices were subjected to PLS-DA, providing similar results. For instance, Fig. 2C contains the PLS-DA scatter plot of scores on the L 2-L 1 —corresponding to the three L s built model— obtained with the negative ionisation FIA− RMS data matrix, where olive oil samples were discriminated on the left side of the plot, displaying negative L 1 values. Besides, among the other vegetable oil samples, located on the right side of the plot, soy oil and its mix with sunflower oil samples displayed positive L 2 values, whereas the remaining samples presented negative L 2 values.

Since this study aimed to classify and authenticate olive oil in front of other vegetable oils, independently of their botanical origin, a two-input class PLS-DA single step —Olive oil *vs*. Others— was proposed. Table 1 shows the most relevant parameters of the PLS-DA calibration models and the classification performance after the external validation. As a result, except for the MLDF fingerprints that allowed a 96.9% classification accuracy, the constructed chemometric models provided the correct classification of all the test samples. Moreover, as shown in Table 2, excellent results were obtained when using SIMCA, particularly for LLDF and MLDF FIA− RMS fingerprints, which allowed an overall accuracy result of 96.9%.

Therefore, considering the great discrimination ability achieved with the MLDF FIA− RMS fingerprints, which suggested the suitability of a profiling strategy to address this food authentication issue, the *m*/*z* signals that formed the matrix were studied, and some of them were putatively identified. In this line, the *m*/*z* 123.0451, 137.0243, and 137.0606, found in negative ionisation FIA− RMS spectra, were assigned as the deprotonated molecule of 4-methylcatechol, hydroxybenzoic acid, and tyrosol, respectively; while in the positive ionisation, the *m*/*z* 415.1360 was identified as the $[M+Na]^+$ form of methyl oleuropein aglycone. Instead, the remaining discriminating ions (*m*/*z* 263.0534, 277.0329, 281.0644, 309.0595, and 735.4107 in the negative ionisation fingerprints, and 805.5800 in the positive ionisation ones) could not be identified. The fact that the identified compounds corresponded to substances well-known for their presence in olive oil proved the correct variable selection strategy, detailed in Section 2.4.2, through IP and selectivity ratio values (see Fig. S6). Moreover, when comparing the corresponding PCA scores plot with its loadings plot (Fig. S7), the selected ten variables were found on the right side of the plot, showing a direct correlation with olive oil samples.

On the other hand, olive oil quality authentication was also evaluated by subjecting the acquired FIA− RMS data —46 × 368 and 46 × 739 (samples × variables), negative and positive ionisation data matrices— to PLS-DA. In this context, Fig. 2D represents the scatter plot of scores of L 1 *vs.* L 2, describing 56.11% of -variance, attained using negative ionisation fingerprints. Similarly to the previous exploratory analysis results, good discrimination along the L 1 between E OO and OO samples was observed. Concerning OO samples, they seemed to be nearly positioned to E OO ones. Thus, considering the E OO and OO similarities found in both exploratory PCA and supervised PLS-DA and the scarcity of OO samples in the sample set, they were conjointly considered in the following classification and authentication study. Again, as performed in the botanical origin classification, a two-input class PLS-DA model, consisting of E OO and OO *vs*. OO, was proposed. As shown in Table 1, while positive ionisation and MLDF fingerprints reached an 89.5% classification accuracy, the corresponding negative ionisation and LLDF ones achieved a 94.7%. Regarding the SIMCA study, as shown in Table 2, negative ionisation and LLDF FIA− RMS fingerprints provided again the best results, reaching an accuracy of 89.5%.

4. Conclusions

The three representative cases under study (red wine, paprika, and olive oil) differed in the complexity of the samples as follows: (i) red wines presented similar FIA− RMS fingerprints without following a clear characteristic pattern due to geographical origin and lacking specific markers for the different classes. Besides, the interregional diversity due to varietal, climatic and geographical features made sample classification a complex issue. (ii) Instead, for paprika samples, distinctive FIA− RMS fingerprints were observed according to sample geographical origin. These differences could be related to the manufacturing processes and peculiarities of each origin. (iii) Finally, in the case of oils, a similar situation was faced since compositional FIA− RMS fingerprints of olive oils differed considerably from those of other vegetable sources (specific biomarkers could be encountered).

Therefore, FIA− RMS fingerprinting, combined with chemometrics, has proved to be a suitable high-throughput analytical approach to address the food classification and authentication issues under study since remarkable classification accuracies were obtained after external validation. Moreover, RMS conferred great selectivity and molecular coverage, leading to rich fingerprints, resulting in satisfactory results when using either negative ionisation, positive ionisation, or LLDF data. Furthermore, the successful application of the MLDF strategy to some of the studied food authentication cases also suggested the eligibility of targeted profiling approaches, focusing on specific compounds, to assess them.

Funding

The authors are grateful for the financial support received from the Spanish Ministry of Science, Innovation and Universities, through the grant PGC2018-095013-B-I00, and from the Agency for Administration of University and Research Grants (Generalitat de Catalunya), through the grants 2017SGR-171 and 2017SGR-310.

CRediT authorship contribution statement

Guillem Campmajó: Conceptualization, Methodology, alidation, Formal analysis, Investigation, riting – original draft, riting – review & editing, Supervision. **Javier Saurina:** Conceptualization, riting – review & editing, Supervision, Funding acquisition. **0scar Núñez:** Conceptualization, riting – original draft, riting – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ac no ledgements

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2021.131491.

References

Arrizabalaga-Larrañaga, A., Campmajó, G., Saurina, J., Núñez, O., Santos, F. J., & Moyano, E. (2021). Determination of capsaicinoids and carotenoids for the characterization and geographical origin authentication of paprika by U PLC−APCI− RMS. *Lwt, 139*, 110533. https://doi.org/10.1016/j.lwt.2020.110533

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, 8*(2), 591−602. https://doi.org/10.1021/acs.jafc.9b0605410.1021/acs.jafc.9b06054.s00110.1021/acs.jafc.9b06054.s002

Borràs, E., Ferré, J., Boqué, R., Mestres, M., Aceña, L., & Busto, O. (2015). Data fusion methodologies for food and beverage authentication and quality assessment - A review. *Analytica Chimica Acta, 891*, 1−14. https://doi.org/10.1016/j.aca.2015.04.042

Cetó, ., Serrano, N., Agató, M., Gámez, A., Esteban, M., Díaz-Cruz, J. M., & Núñez, O. (2018). Determination of PLC-U fingerprints of spanish paprika (Capsicum annuum L.) for its classification by linear discriminant analysis. *Sensors, 18*(12), 4479. https://doi.org/10.3390/s18124479

Chen, P., Sun, J., & Ford, P. (2014). Differentiation of the four major species of cinnamons (C. burmannii, C. verum, C. cassia, and C. loureiroi) using a flow injection mass spectrometric (FIMS) fingerprinting method. *Journal of Agricultural and Food Chemistry*, *2*(12), 2516−2521. https://doi.org/10.1021/jf405580c.

Cody, R. B., Laramée, J. A., & Durst, . D. (2005). ersatile new ion source for the analysis of materials in open air under ambient conditions. *Analytical Chemistry, 77*(8), 2297−2302. https://doi.org/10.1021/ac050162j10.1021/ac050162j.s00110.1021/ac050162j.s00210.1021/ac050162j.s003

Cuadros-Rodríguez, L., Ruiz-Samblás, C., alverde-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

Du, L., Lu, ., Cai, Z. (Julia), Bao, L., artmann, C., Gao, B., & u, L. (Lucy). (2018). Rapid detection of milk adulteration using intact protein flow injection mass spectrometric fingerprints combined with chemometrics. *Food Chemistry*, *240*, 573−578. https://doi.org/10.1016/j.foodchem.2017.07.107.

Farré, M., Picó, ., & Barceló, D. (2019). Direct analysis in real-time high-resolution mass spectrometry as a valuable tool for polyphenols profiling in olive oil. *Analytical Methods, 11*(4), 472−482. https://doi.org/10.1039/c8ay01865k

Gachumi, G., Purves, R. ., opf, C., & El-Aneed, A. (2020). Fast Quantification without Conventional Chromatography, the Growing Power of Mass Spectrometry. *Analytical Chemistry, 92*(13), 8628−8637. https://doi.org/10.1021/acs.analchem.0c00877

Gao, B., Lu, ., Sheng, ., Chen, P., & u, L. (2013). Differentiating organic and conventional sage by chromatographic and mass spectrometry flow injection fingerprints combined with principal component analysis. *Journal of Agricultural and Food Chemistry, 1*(12), 2957−2963. https://doi.org/10.1021/jf304994z

Gao, B., Qin, F., Ding, T., Chen, ., Lu, ., & u, L. (2014). Differentiating organically and conventionally grown oregano using ultraperformance liquid chromatography mass spectrometry (UPLC-MS), headspace gas chromatography with flame ionization detection (eadspace-GC-FID), and flow injection mass spectrum (FIMS). *Journal of Agricultural and Food Chemistry, 2*(32), 8075−8084. https://doi.org/10.1021/jf502419y

Garrido, J., & Borges, F. (2013). ine and grape polyphenols - A chemical perspective. *Food Research International, 4*(2), 1844−1858. https://doi.org/10.1016/j.foodres.2013.08.002

Gosetti, F., Bolfi, B., Manfredi, M., Calabrese, G., & Marengo, E. (2015). Determination of eight polyphenols and pantothenic acid in extra-virgin olive oil samples by a simple, fast, high-throughput and sensitive ultra high performance liquid chromatography with tandem mass spectrometry method. *Journal of Separation Science, 38*(18), 3130−3136. https://doi.org/10.1002/jssc.201500452

Gutiérrez-Escobar, R., Aliaño-González, M. J., & Cantos- illar, E. (2021). ine Polyphenol Content and Its Influence on ine Quality and Properties: A Review. *Molecules, 2* (3), 718. https://doi.org/10.3390/molecules26030718

roboňová, K., & Sádecká, J. (2020). Coumarins content in wine: Application of PLC, fluorescence spectrometry, and chemometric approach. *Journal of Food Science and Technology, 7*(1), 200−209. https://doi.org/10.1007/s13197-019-04048-2

Ibáñez, C., Simó, C., García-Cañas, ., Acunha, T., & Cifuentes, A. (2015). The role of direct high-resolution mass spectrometry in foodomics. *Analytical and Bioanalytical Chemistry, 407*(21), 6275−6287. https://doi.org/10.1007/s00216-015-8812-1

Ivanova-Petropulos, ., Naceva, Z., Sándor, ., Makszin, L., Deutsch-Nagy, L., Berkics, B., … Kilár, F. (2018). Fast determination of lactic, succinic, malic, tartaric, shikimic, and citric acids in red ranec wines by CZE-ESI-QTOF-MS. *Electrophoresis, 39*(13), 1597−1605. https://doi.org/10.1002/elps.201700492

Kuo, T. ., Dutkiewicz, E. P., Pei, J., & su, C. C. (2020). Ambient Ionization Mass Spectrometry Today and Tomorrow: Embracing Challenges and Opportunities. *Analytical Chemistry, 92*(3), 2353−2363. https://doi.org/10.1021/acs.analchem.9b05454

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

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, 8* , 163−176. https://doi.org/10.1016/j.tifs.2019.01.017

Morin, J.-F., & Lees, M. (2018). *FoodIntegrity handbook. A guide to food authenticity issues and analytical solutions* (J.-F. Morin & M. Lees, eds.). https://doi.org/10.32741/fihb.

Nanita, S. C., & Kaldon, L. G. (2016). Emerging flow injection mass spectrometry methods for high-throughput quantitative analysis. *Analytical and Bioanalytical Chemistry, 408*(1), 23−33. https://doi.org/10.1007/s00216-015-9193-1

Pérez-Gálvez, A., Garrido-Fernández, J., Mínguez-Mosquera, M. I., Lozano-Ruiz, M., & Montero-de-Espinosa, . (1999). Fatty Acid Composition of Two New Pepper arieties (Capsicum annuum L. cv. Jaranda and Jariza). Effect of Drying Process and Nutritional Aspects. *Journal of the American Oil Chemists' Society, 7* (2), 205−208. https://doi.org/10.1007/s11746-999-0219-8

Pluskal, T., Castillo, S., illar-Briones, A., & Orešič, M. (2010). MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics, 11*, 395. https://doi.org/10.1186/1471-2105-11-395

Points, J., Manning, L., & Group, A. (2020). Facing up to food fraud in a pandemic. *Food Science and Technology*, *34*(2), 16−20. https://doi.org/10.1002/fsat.3403_4.x.

Rodionova, O. ., Titova, A. ., & Pomerantsev, A. L. (2016). Discriminant analysis is an inappropriate method of authentication. *TrAC - Trends in Analytical Chemistry, 78*, 17−22. https://doi.org/10.1016/j.trac.2016.01.010

Rothwell, J. A., Perez-Jimenez, J., Neveu, ., Medina-Remón, A., M' iri, N., García-Lobato, P., ... Scalbert, A. (2013). Phenol-Explorer 3.0: a major update of the Phenol-Explorer database to incorporate data on the effects of food processing on polyphenol content. *Database*, *2013*. https://doi.org/10.1093/database/bat070.

Rubert, J., Zachariasova, M., & ajslova, J. (2015). Advances in high-resolution mass spectrometry based on metabolomics studies for food – a review. *Food Additives and Contaminants - Part A Chemistry, Analysis, Control, E posure and Risk Assessment, 32*(10), 1685−1708. https://doi.org/10.1080/19440049.2015.1084539

Schymanski, E. L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, . P., & ollender, J. (2014). Identifying small molecules via high resolution mass spectrometry: Communicating confidence. *Environmental Science and Technology, 48*(4), 2097−2098. https://doi.org/10.1021/es5002105

Smith, C. A., Maille, G. O ., ant, E. J., Qin, C., Trauger, S. A., Brandon, T. R., … Siuzdak, G. (2005). METLIN: A metabolite mass spectral database. *Therapeutic Drug Monitoring, 27*(6), 747−751. https://doi.org/10.1097/01.ftd.0000179845.53213.39

Sun, J., Zhang, M., Kubzdela, N., Luo, ., arnly, J. M., & Chen, P. (2018). Determination of ariance of Secondary Metabolites in Lettuces Grown Under Different Light Sources by Flow Injection Mass Spectrometric (FIMS) Fingerprinting and ANO A−PCA. *Journal of Analysis and Testing, 2*(4), 312−321. https://doi.org/10.1007/s41664-018-0072-6

Takáts, Z., iseman, J. M., Gologan, B., & Cooks, R. G. (2004). Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. *Science, 30* (5695), 471−473. https://doi.org/10.1126/science.1104404

ishart, D. S., Feunang, . D., Marcu, A., Guo, A. C., Liang, K., ázquez-Fresno, R., … Scalbert, A. (2018). MDB 4.0: The human metabolome database for 2018. *Nucleic Acids Research*, *4* (D1), D608−D617. https://doi.org/10.1093/nar/gkx1089.

old, S. (1976). Pattern recognition by means of disjoint principal components models. *Pattern Recognition, 8*(3), 127−139. https://doi.org/10.1016/0031-3203(76)90014-5

Supplementary Material

FIA-HRMS fingerprinting subjected to chemometrics as a valuable tool to address food classification and authentication: application to red wine, paprika, and vegetable oil samples

Guillem Campmajó[a,b,*], Javier Saurina[a,b], Oscar Núñez[a,b]

[a] Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, Marti i Franques 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 (Gaudi), E08921 Santa Coloma de Gramenet, Spain.

* Corresponding author: Guillem Campmajó
(Tel.: 34-93-403-9116, e-mail: gcampmajo@ub.edu)

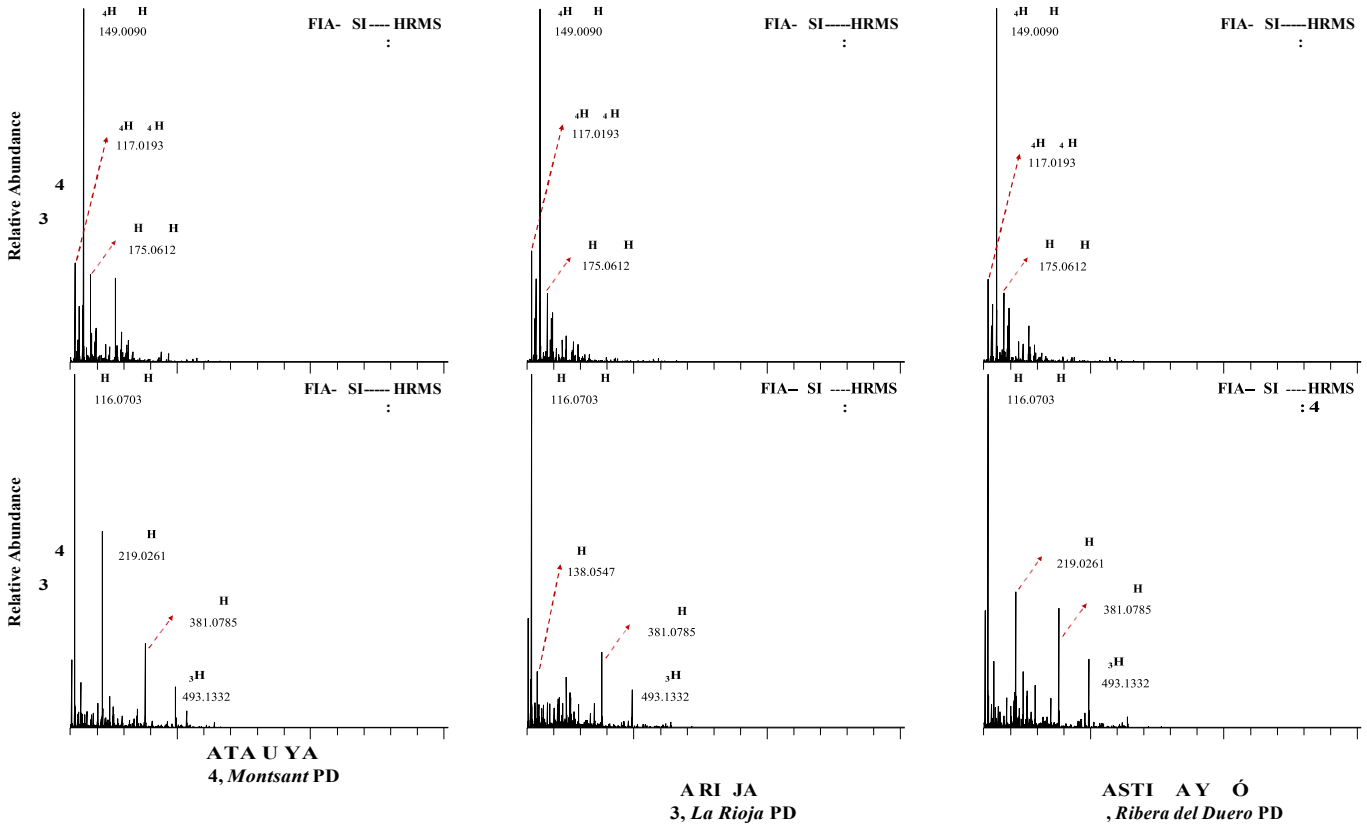


Fig S Negative and positive ionisation FIA-HRMS fingerprints obtained for a *Catalunya*, *La Rioja*, and *Castilla y León* red wine sample.

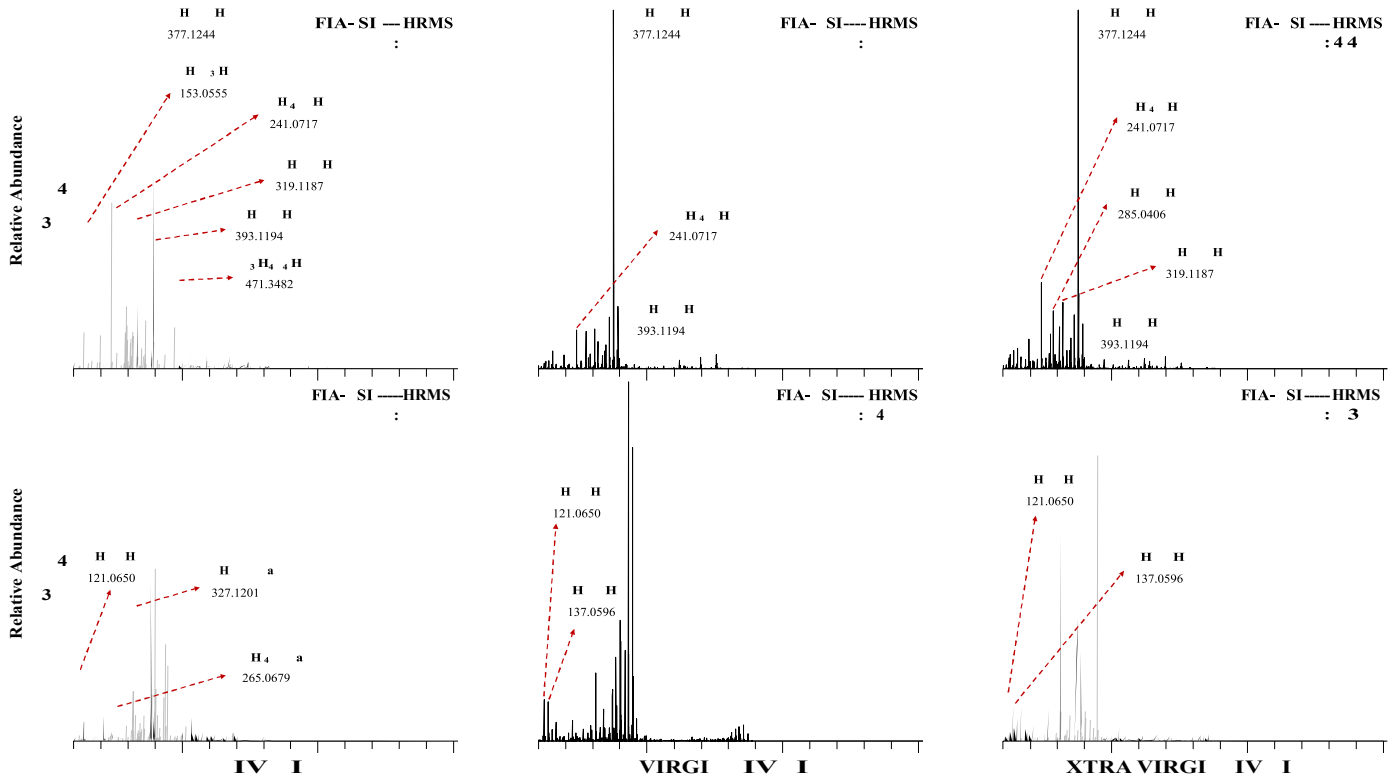


Fig S Negative and positive ionisation FIA-HRMS fingerprints obtained for an OO, VOO, and EVOO sample.

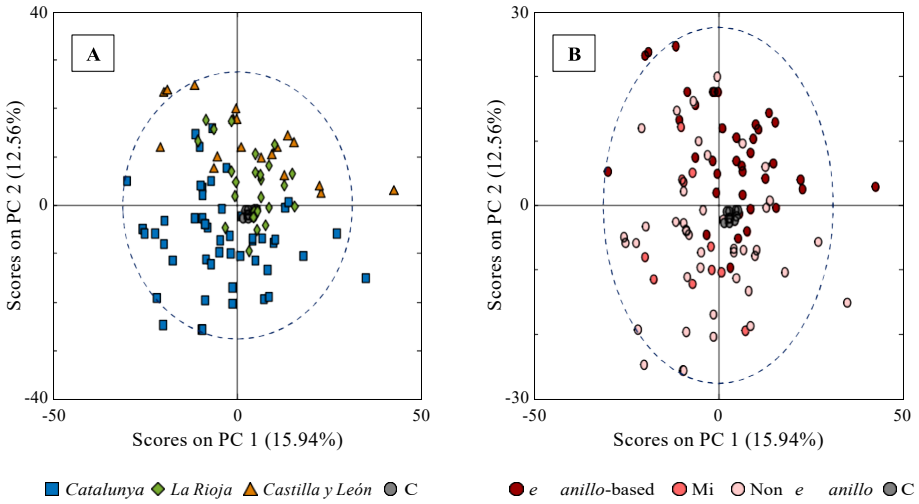


Fig S3 PCA scores plot of PC1 *vs*. PC2, using positive ionisation FIA-HRMS fingerprints, for the analysed red wine samples according to (A) their geographical origin and (B) grape variety.

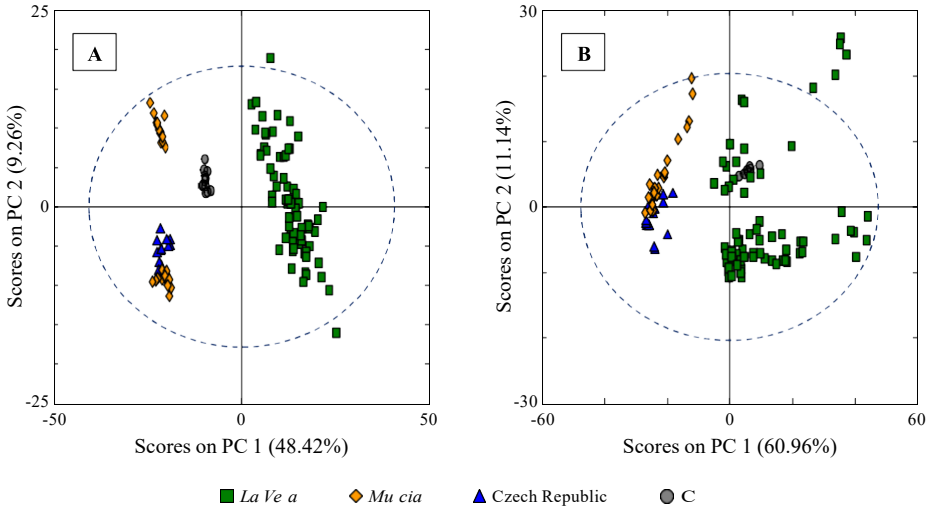


Fig S4 PCA scores plot of PC1 *vs*. PC2, using (A) negative and (B) positive ionisation FIA-HRMS fingerprints, for the analysed paprika samples according to their geographical origin.

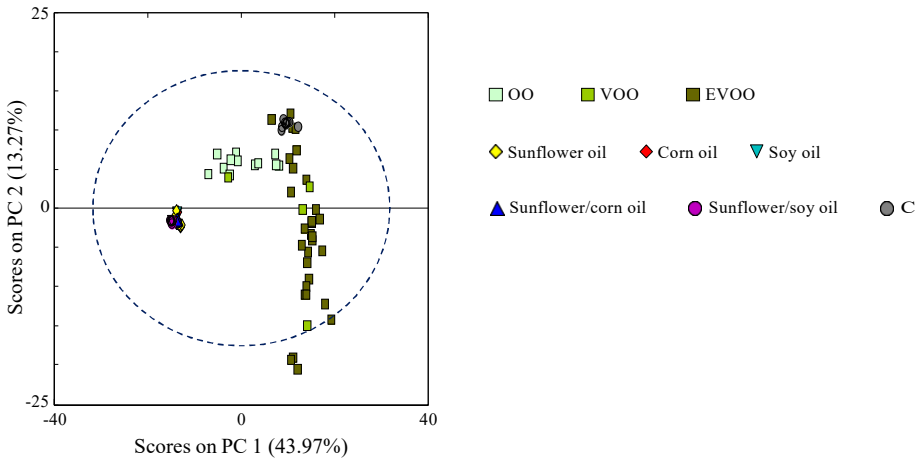


Fig S PCA scores plot of PC1 *vs*. PC2, using negative ionisation FIA-HRMS fingerprints, for the analysed olive and vegetable oil samples according to their botanical origin.

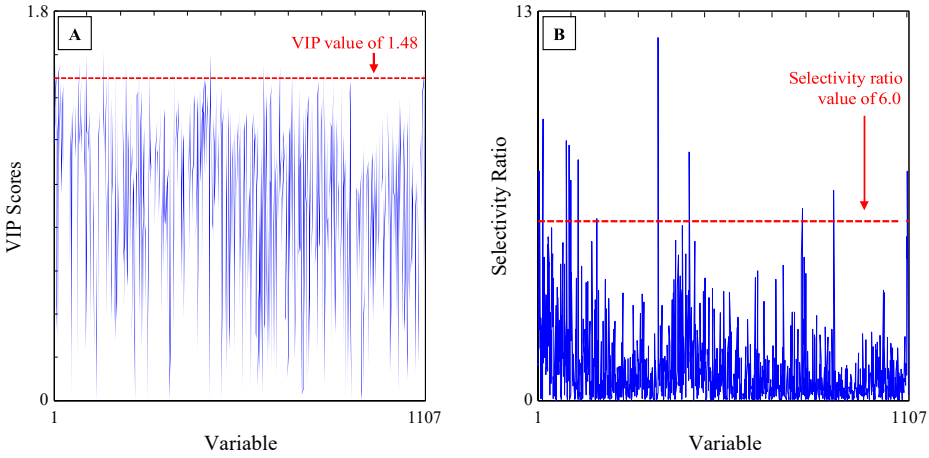


Fig S (A) VIP plot and (B) selectivity ratio plot obtained for the Olive oil *vs*. Others PLS-DA model, using LLDF data. The dashed line represents the threshold established after following the variable selection strategy described in Section 2.4.2 to construct the MLDF matri .

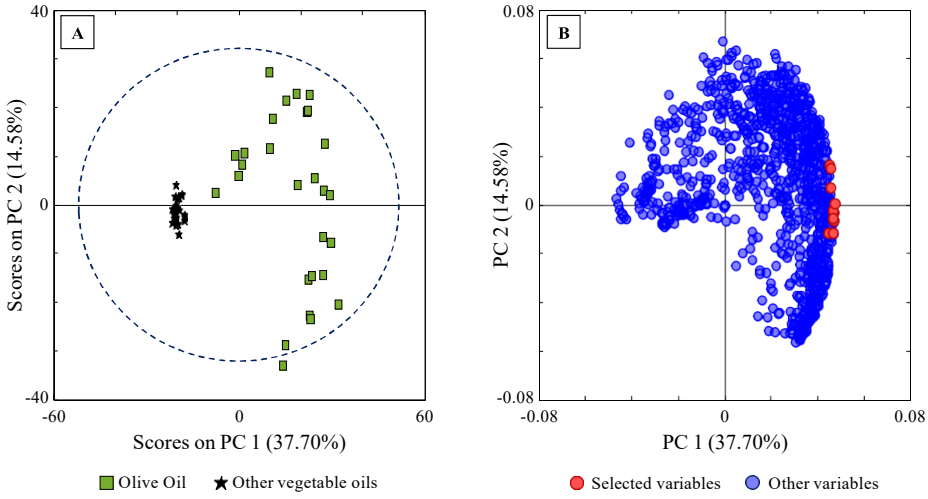


Fig S (A) PCA scores plot of PC1 *vs*. PC2 and (B) the corresponding loadings plot, using LLDF FIA-HRMS fingerprints, for the analysed vegetable oil samples.

Table S Putative identification by FIA-HRMS of some of the most characteristic ions found in the food matrices under study.

R D WI

Measured accurate mass	Ionisation mode	Ion assignment	Molecular formula	RDB e uivalent	Accurate mass error ppm	Isotopic pattern score %	MS HRMS	Putative identification	lass	Identification confidence level
117.0193	Negative	$[M-H]^-$	$C_4H_6O_4$	2.5	-0.017	99.98	73.0295; 82.3092; 92.8401; 99.0089; 106.2155	Succinic acid	Dicarbo ylic acid	II
129.0195	Negative	$[M-H]^-$	$C_5H_6O_4$	3.5	0.992	99.94	57.0345; 85.0296; 101.0245	Glutaconic acid	Dicarbo ylic acid	II
133.0143	Negative	$[M-H]^-$	$C_4H_6O_5$	2.5	0.627	99.81	71.0138; 72.9931; 87.0452; 89.0245; 115.0036	Malic acid	Dicarbo ylic acid	II
135.0299	Negative	$[M-H]^-$	$C_4H_8O_5$	1.5	0.247	99.69	59.0139; 71.0139; 75.0087; 89.0245; 117.0195	L-threonic acid	Carbo ylic acid (sugar acid)	II
145.0140	Negative	$[M-H]^-$	$C_5H_6O_5$	3.5	-1.769	99.39	57.0346; 73.0294; 83.0139; 99.0087; 101.0244; 121.2956; 127.0400	2-o oglutaric acid	Dicarbo ylic acid	II
147.0299	Negative	$[M-H]^-$	$C_5H_8O_5$	2.5	-0.181	99.61	57.0345; 85.0295; 87.0082; 101.0244; 103.0400; 129.0194	Citramalic acid	Dicarbo ylic acid	II
149.0090	Negative	$[M-H]^-$	$C_4H_6O_6$	2.5	-1.014	99.91	59.0139; 72.9930; 75.0087; 87.0087; 103.0037; 130.9986	Tartaric acid	Dicarbo ylic acid (sugar acid)	II
161.0455	Negative	$[M-H]^-$	$C_6H_{10}O_5$	2.5	-0.476	99.30	57.0346; 59.0139; 99.0452; 101.0244; 117.0558; 143.0348	3-hydro y-3-methylglutaric acid	Dicarbo ylic acid	II
169.0143	Negative	$[M-H]^-$	$C_7H_6O_5$	5.5	0.198	99.54	69.0346; 81.0346; 97.0296; 107.0139; 125.0244	Gallic acid	Polyphenol (phenolic acid - hydro ybenzoic acid)	II
173.0092	Negative	$[M-H]^-$	$C_6H_6O_6$	4.5	0.340	99.25	67.0190; 85.0295; 111.0088; 129.0196; 154.9986	*t ans*-aconitic acid	Tricarbo ylic acid	II
175.0612	Negative	$[M-H]^-$	$C_7H_{12}O_5$	2.5	0.133	99.93	85.0659; 113.0609; 131.0715; 157.0504	Isopropylmalic acid	Dicarbo ylic acid	III
177.0405	Negative	$[M-H]^-$	$C_6H_{10}O_6$	2.5	0.388	99.87	57.0346; 59.0139; 71.0139; 75.0088; 85.0295; 87.0088; 89.0245; 99.0088; 129.0193; 141.0194; 159.0297	Gluconolactone	Polyhydro y acid	II

179.0352	Negative	$[M-H]^-$	$C_9H_8O_4$	6.5	1.162	99.94	107.0504; 135.0453	Caffeic acid	Polyphenol (phenolic acid - hydro ycinnamic acid)	II
179.0563	Negative	$[M-H]^-$	$C_6H_{12}O_6$	1.5	0.774	99.26	59.0139; 71.0138; 75.0087; 85.0296; 89.0245; 95.0139; 101.0244; 113.0245; 135.0452; 161.0455		Monosaccharide	III
181.0720	Negative	$[M-H]^-$	$C_6H_{14}O_6$	0.5	1.097	99.58	59.0139; 71.0139; 73.0295; 83.0139; 85.0296; 89.0245; 101.0244; 113.0244; 115.0401; 119.0349; 131.0348; 163.0611		Sugar alcohol	III
191.0201	Negative	$[M-H]^-$	$C_6H_8O_7$	3.5	1.854	96.75	57.0346; 85.0296; 87.0088; 111.0088; 129.0195; 154.9985; 173.0093	Citric acid	Tricarbo ylic acid	II
193.0357	Negative	$[M-H]^-$	$C_6H_{10}O_7$	2.5	1.420	99.93	59.0139; 71.0139; 72.9931; 85.0296; 89.0245; 101.0245; 113.0245; 131.0352; 157.0141	Glucuronic and galacturonic acid	Carbo ylic acid (sugar acid)	II
195.0512	Negative	$[M-H]^-$	$C_6H_{12}O_7$	1.5	1.046	99.66	59.0139; 71.0138; 75.0087; 85.0296; 87.0088; 89.0243; 99.0090; 101.0244; 111.0089; 129.0196; 159.0297; 177.0399	Gluconic acid	Dicarbo ylic acid (sugar acid)	I
197.0458	Negative	$[M-H]^-$	$C_9H_{10}O_5$	5.5	1.438	99.07	95.0137; 123.0092; 138.0324; 153.0564; 166.9983; 182.0223	Syringic acid	Polyphenol (phenolic acid - hydro ybenzoic acid)	II
							124.0166; 169.0144	Ethyl gallate	Polyphenol (phenolic acid - hydro ybenzoic acid) derivative	II
209.0302	Negative	$[M-H]^-$	$C_6H_{10}O_8$	2.5	-0.241	99.58	71.0139; 85.0295; 111.0088; 129.0194; 133.0144; 147.0297; 191.0203	D-glucaric acid	Dicarbo ylic acid (sugar acid)	II
295.0462	Negative	$[M-H]^-$	$C_{13}H_{12}O_8$	8.5	0.947	99.12	149.0088; 163.0400	-coumaroyl tartaric acid	Polyphenol (phenolic acid - hydro ycinnamic acid)	II
311.0414	Negative	$[M-H]^-$	$C_{13}H_{12}O_9$	8.5	1.655	98.63	149.0088; 179.0352	Caffeoyl tartaric acid	Polyphenol (phenolic acid - hydro ycinnamic acid)	II
104.1068	Positive	$[M]^+$	$C_5H_{14}ON$	-0.5	-2.311	71.73	58.0654; 60.0811	Choline	Choline	II

116.0703	Positive	$[M+H]^+$	$C_5H_9O_2N$	1.5	-2.715	80.38	70.0651	Proline	Amino acid	II
118.0859	Positive	$[M+H]^+$	$C_5H_{11}O_2N$	0.5	-3.092	55.16	55.0546; 72.0808	Valine	Amino acid	II
							100.0757; 101.0596	5-aminovaleric acid	Amino acid	II
							59.0495; 70.0652; 72.0808; 87.0439	2-(methylamino)isobutyric acid	Amino acid	II
		$[M]^+$	$C_5H_{12}O_2N$				58.0654; 59.0733	Betaine	Amino acid	II
127.0387	Positive	$[M+H]^+$	$C_6H_6O_3$	3.5	-2.051	99.65	53.0389; 71.0492; 81.0335; 99.0441; 109.0285	Pyrogallol and phloroglucinol	Polyphenol (trihydro ybenzene)	III
							51.0232; 53.0389; 55.0182; 57.0339; 65.0384; 67.0178; 68.9975; 71.0128; 81.0335; 83.0126; 97.0284; 109.0285	Maltol	Pyranone	II
							51.0232; 53.0389; 55.0182; 57.0339; 69.0336; 81.0335; 97.0284; 109.0285	5-hydro ymethyl-2-furaldehyde	Furan	II
129.0544	Positive	$[M+H]^+$	$C_6H_8O_3$	2.5	-1.477	99.22	55.0182; 73.0284; 83.0604; 91.0390; 101.0233	Furaneol	Furan	II
132.1017	Positive	$[M+H]^+$	$C_6H_{13}O_2N$	0.5	-1.932	99.98	69.0700; 86.0964	Leucine and isoleucine	Amino acid	II
136.0616	Positive	$[M+H]^+$	$C_5H_5N_5$	5.5	-1.336	99.71	94.0401; 109.0508; 119.0353	Adenine	Nucleobase	II
138.0547	Positive	$[M]^+$	$C_7H_8O_2N$	4.5	-1.847	94.25	65.0388; 67.0542; 78.0339; 79.0416; 92.0494; 94.0652; 110.0600; 121.0649	Trigonelline	Alkaloid	II
146.1173	Positive	$[M]^+$	$C_7H_{16}O_2N$	0.5	-2.021	96.68	60.0811; 87.0441	Acetylcholine	Choline derivative	II
158.0814	Positive	$[M+H]^+$	$C_7H_{11}O_3N$	2.5	1.203	99.15	70.0652; 84.0444; 98.0601; 112.0757; 116.0706; 140.0705	Acetylproline	Amino acid	II
162.1121	Positive	$[M]^+$	$C_7H_{16}O_3N$	0.5	-2.035	84.37	60.0811; 85.0285; 102.0914; 103.0390	Carnitine	Amino acid	II
175.1186	Positive	$[M+H]^+$	$C_6H_{14}O_2N_4$	1.5	-1.840	90.52	60.0560; 70.0652; 112.0868; 116.0706; 130.0974; 157.1080; 158.0926	Arginine	Amino acid	II
185.0208	Positive	$[M+Na]^+$	$C_9H_6O_3$	6.5	-0.893	98.03		4-hydro ycoumarin	Coumarin	IV
201.0156	Positive	$[M+Na]^+$	$C_9H_6O_4$	6.5	-1.144	88.70		Esculetin	Coumarin	IV
215.0312	Positive	$[M+Na]^+$	$C_{10}H_8O_4$	6.5	-1.348	97.70		Scopoletin	Coumarin	IV
219.0261	Positive	$[M+K]^+$	$C_6H_{12}O_6$	0.5	-1.856	99.22			Monosaccharide	IV

229.1541	Positive	$[M+H]^+$	$C_{11}H_{20}O_3N_2$	2.5	-2.483	77.75	70.0652; 86.0964; 116.0706	Leucylproline	Dipeptide	II
258.1096	Positive	$[M]^+$	$C_8H_{21}O_6NP$	-0.5	1.977	99.92	60.0811; 86.0964; 104.1070; 124.9999; 184.0735	Glycerophosphocholine	Choline derivative	II
381.0785	Positive	$[M+K]^+$	$C_{12}H_{22}O_{11}$	1.5	-2.282	85.16			Disaccharide	IV
463.1228	Positive	$[M]^+$	$C_{22}H_{23}O_{11}$	11.5	-1.529	95.65	258.0521; 286.0472; 301.0705	Peonidin 3-O-glucoside	Polyphenol (flavonoid - anthocyanin)	II
479.1178	Positive	$[M]^+$	$C_{22}H_{23}O_{12}$	11.5	-1.216	99.06	274.0473; 302.0422; 317.0660	Petunidin 3-O-glucoside	Polyphenol (flavonoid - anthocyanin)	II
493.1332	Positive	$[M]^+$	$C_{23}H_{25}O_{12}$	11.5	-1.749	95.74	270.0522; 286.0465; 287.0544; 298.0466; 299.0546; 315.0493; 316.0575; 331.0804	Malvidin 3-O-glucoside	Polyphenol (flavonoid - anthocyanin)	II
535.1440	Positive	$[M]^+$	$C_{25}H_{27}O_{13}$	12.5	-1.116	96.49	242.0570; 270.0522; 287.0546; 299.0546; 315.0496; 331.0807	Malvidin 3-O-(6 -acetyl-glucoside)	Polyphenol (flavonoid - anthocyanin)	II
639.1699	Positive	$[M]^+$	$C_{32}H_{31}O_{14}$	17.5	-1.536	93.21	242.0572; 270.0519; 287.0548; 299.0552; 315.0499; 331.0811	Malvidin 3-O-coumaroylglucoside	Polyphenol (flavonoid - anthocyanin)	II

PAPRI A

Measured accurate mass	Ionisation mode	Ion assignment	Molecular formula	RDB e uivalent	Accurate mass error ppm	Isotopic pattern score %	MS HRMS	Tentative identification	lass	Identification confidence level
111.0087	Negative	$[M-H]^-$	$C_5H_4O_3$	4.5	-0.967	99.40	67.0190; 68.0223	2-furoic acid	Carbo ylic acid	II
117.0193	Negative	$[M-H]^-$	$C_4H_6O_4$	2.5	-0.188	99.98	73.0295; 74.0328; 99.0088	Succinic acid	Dicarbo ylic acid	II
133.0142	Negative	$[M-H]^-$	$C_4H_6O_5$	2.5	-0.651	99.90	71.0139; 72.9932; 87.0087; 89.0244; 115.0035	Malic acid	Dicarbo ylic acid	II
135.0298	Negative	$[M-H]^-$	$C_4H_8O_5$	1.5	-0.419	99.31	59.0139; 71.0139; 75.0088; 89.0245; 117.0081	L-threonic acid	Carbo ylic acid (sugar acid)	II
147.0299	Negative	$[M-H]^-$	$C_5H_8O_5$	2.5	-0.249	99.75	57.0346; 59.0139; 71.0139; 75.0088; 83.0138; 85.0295; 87.0088; 89.0244; 99.0088; 101.0244; 129.0194	8-ribono-1,4-lactone	Monosaccharide	II
							57.0346; 75.0088; 83.0138; 85.0295; 101.0244; 103.0400; 129.0194	D-a-hydro yglutaric acid	Dicarbo ylic acid	II
161.0455	Negative	$[M-H]^-$	$C_6H_{10}O_5$	2.5	-0.538	99.50	57.0346; 59.0139; 99.0451; 101.0243; 117.0344	3-hydro y-3-methylglutaric acid	Dicarbo ylic acid	II

173.0090	Negative	$[M-H]^-$	$C_6H_6O_6$	4.5	-1.047	99.87	67.0190; 85.0295; 111.0087; 129.0193; 154.9986	*t ans*-aconitic acid	Tricarbo ylic acid	II
175.0610	Negative	$[M-H]^-$	$C_7H_{12}O_5$	2.5	-0.895	99.12	59.0139; 85.0659; 113.0608; 115.0400; 131.0712; 157.0505	2-isopropylmalic acid	Dicarbo ylic acid	II
177.0403	Negative	$[M-H]^-$	$C_6H_{10}O_6$	2.5	-0.741	99.90	57.0346; 59.0139; 71.0139; 75.0088; 85.0295; 87.0088; 89.0244; 99.0088; 129.0194; 141.0194; 159.0299	Gluconolactone	Polyhydro y acid	II
179.0561	Negative	$[M-H]^-$	$C_6H_{12}O_6$	1.5	-0.287	99.94	59.0139; 71.0139; 75.0088; 85.0295; 89.0244; 94.0298; 131.0348; 135.0452; 161.0456		Monosaccharide	III
181.0717	Negative	$[M-H]^-$	$C_6H_{14}O_6$	0.5	-0.229	99.54	59.0139; 71.0139; 73.0295; 83.0138; 85.0295; 89.0244; 101.0243; 113.0243; 115.0397; 119.0349; 131.0348; 163.0611		Sugar alcohol	III
187.0975	Negative	$[M-H]^-$	$C_9H_{16}O_4$	2.5	-0.386	98.98	57.0346; 97.0659; 123.0816; 125.0971; 126.1006; 143.1077	Azelaic acid	Dicarbo ylic acid	II
191.0198	Negative	$[M-H]^-$	$C_6H_8O_7$	3.5	0.388	99.73	85.0295; 87.0087; 111.0086; 129.0193; 173.0091	Citric acid	Tricarbo ylic acid	II
191.0560	Negative	$[M-H]^-$	$C_7H_{12}O_6$	2.5	-0.687	99.67	85.0295; 87.0087; 93.0346; 111.0447; 127.0399; 173.0451	uinic acid	Carbo ylic acid	II
195.0509	Negative	$[M-H]^-$	$C_6H_{12}O_7$	1.5	-0.697	99.76	57.0346; 59.0139; 71.0139; 75.0088; 85.0295; 87.0088; 89.0244; 99.0088; 101.0244; 129.0194; 159.0296; 177.0405	Gluconic and galactonic acid	Dicarbo ylic acid (sugar acid)	II
205.0352	Negative	$[M-H]^-$	$C_7H_{10}O_7$	3.5	-0.711	99.65	67.0189; 69.0346; 71.0139; 72.9932; 75.0088; 81.0346; 87.0088; 99.0088; 111.0087; 125.0243; 131.0349; 143.0348; 161.0608; 164.0719; 177.0559; 190.0274		Tricarbo ylic acid	III
215.0328[a]	Negative	$[M+Cl]^-$	$C_6H_{12}O_6$	0.5	0.144	98.79	59.0139; 71.0139; 75.0088; 85.0295; 89.0245; 95.0135; 101.0244; 113.0245; 135.0452; 161.0456		Monosaccharide	III
279.2329	Negative	$[M-H]^-$	$C_{18}H_{32}O_2$	3.5	-0.299	99.79	59.0130; 261.2208	Linoleic acid	Polyunsaturated fatty acid	II

295.2276	Negative	$[M-H]^-$	$C_{18}H_{32}O_3$	3.5	-0.773	97.83	59.0138; 113.0971; 171.1026; 195.1388; 249.2227; 277.2171	Hydro yoctadecadienoic acid	Polyunsaturated fatty acid	II
313.2384	Negative	$[M-H]^-$	$C_{18}H_{34}O_4$	2.5	-0.137	97.76	99.0816; 129.0921; 183.1390; 201.1132; 251.2380; 277.2169; 295.2274	Dihydro yoctadecenoic acid	Monounsaturated fatty acid	II
329.2333	Negative	$[M-H]^-$	$C_{18}H_{34}O_5$	2.5	-0.174	70.47	59.0139; 99.0816; 127.1127; 139.1129; 171.1027; 201.1133; 211.1338; 229.1443; 293.2125; 311.2231	(15Z)-9,12,13-trihydro y-15-octadecenoic acid	Monounsaturated fatty acid	II
447.0931	Negative	$[M-H]^-$	$C_{21}H_{20}O_{11}$	12.5	-0.525	94.61	151.0035; 227.0340; 243.0299; 255.0297; 271.0246; 300.0274; 301.0351	uercitrin	Polyphenol (flavonoid - flavonol)	II
							151.0035; 227.0340; 243.0299; 255.0297; 271.0246; 284.0319; 285.0404; 300.0274; 301.0351; 327.0517	Astragalin	Polyphenol (flavonoid - flavonol)	II
659.3285	Negative	$[M-H]^-$	$C_{32}H_{52}O_{14}$	7.5	0.168	94.67	59.0140; 69.0347; 71.0139; 85.0295; 101.0245; 111.0085; 113.0243; 319.2266; 497.2770; 641.3340	Capsianoside I	Capsianoside	II
104.1072	Positive	$[M]^+$	$C_5H_{14}ON$	-0.5	2.011	99.91	58.0658; 60.0815	Choline	Choline	II
116.0707	Positive	$[M+H]^+$	$C_5H_9O_2N$	1.5	0.904	99.77	70.0657	Proline	Amino acid	II
118.0864	Positive	$[M+H]^+$	$C_5H_{11}O_2N$	0.5	0.888	99.82	55.0542; 72.0814	Valine	Amino acid	II
							100.0760; 101.0600	5-aminovaleric acid	Amino acid	II
							59.0499; 70.0657; 72.0814; 87.0445	2-(methylamino)isobutyric acid	Amino acid	II
		$[M]^+$	$C_5H_{12}O_2N$				58.0659; 59.0737	Betaine	Amino acid	II
126.0550	Positive	$[M+H]^+$	$C_6H_8O_2N$	3.5	0.039	99.98	68.0138; 80.0500; 108.0445	6-(hydro ymethyl)pyridin-3-ol	Pyridine	II
132.1019	Positive	$[M+H]^+$	$C_6H_{13}O_2N$	0.5	-0.267	99.07	69.0706; 86.0969	Leucine and isoleucine	Amino acid	II
142.1225	Positive	$[M+H]^+$	$C_8H_{14}ON$	1.5	-0.990	99.96	58.0659; 67.0548; 93.0704; 98.0968; 124.1121	Tropine	Alkaloid	II
144.1016	Positive	$[M]^+$	$C_7H_{14}O_2N$	1.5	-2.326	74.15	58.0659; 84.0811; 102.0554	Stachydrine	Amino acid	II

146.0809	Positive	$[M+H]^+$	$C_6H_{11}O_3N$	1.5	-1.573	98.14	82.0657; 100.0759; 128.0706	4-hydro y-1-methyl-2-pyrrolidine carbo ylic acid	Carbo ylic acid (pyrrolidine)	II
166.0860	Positive	$[M+H]^+$	$C_9H_{11}O_2N$	4.5	-1.536	98.86	103.0545; 120.0808; 131.0491; 149.0592	L-phenylalanine	Amino acid	II
175.1186	Positive	$[M+H]^+$	$C_6H_{14}O_2N_4$	1.5	-1.954	90.55	60.0564; 70.0658; 116.0707; 130.0974; 158.0922	Arginine	Amino acid	II
189.1594	Positive	$[M]^+$	$C_9H_{21}O_2N_2$	0.5	-1.821	95.86	60.0815; 84.0813; 130.0862	N^6.N^6.N^6-trimethyl-L-lysine	Amino acid	II
306.2057	Positive	$[M+H]^+$	$C_{18}H_{28}O_3N$	5.5	-2.156	90.02	122.0363; 137.0595; 153.1269; 170.1535; 182.1537	Capsaicin	Capsaicinoid	II

IV I

Measured accurate mass	Ionisation mode	Ion assignment	Molecular formula	RDB e	Accurate mass error ppm	Isotopic pattern score %	MS HRMS	Tentative identification	lass	Identification confidence level
121.0294	Negative	$[M-H]^-$	$C_7H_6O_2$	5.5	-0.518	99.87	92.0268; 93.0346	4-hydro ybenzaldehyde	Benzaldehyde	II
123.0451	Negative	$[M-H]^-$	$C_7H_8O_2$	4.5	-0.267	99.75	95.0501; 105.0346; 108.0218; 122.0372	4-methylcatechol	Polyphenol (alkylphenol)	II
153.0555	Negative	$[M-H]^-$	$C_8H_{10}O_3$	4.5	-1.290	98.79	93.0345; 95.0503; 105.0344; 108.0215; 122.0373; 123.0452	Hydro ytyrosol	Polyphenol (tyrosol)	II
165.0556	Negative	$[M-H]^-$	$C_9H_{10}O_3$	5.5	-0.712	99.56	59.0139; 93.0344; 121.0657	Dihydro- -coumaric acid	Polyphenol (phenolic acid - hydro yphenylpropanoic acid)	II
							93.0344; 105.0344; 121.0295; 123.0452	a-metho yphenylacetic acid	Carbo ylic acid (anisole)	II
							119.0504; 147.0449	3-phenyllactic acid	Carbo ylic acid	II
183.0663	Negative	$[M-H]^-$	$C_9H_{12}O_4$	4.5	0.043	99.26	69,0346; 95,0503	Dialdehydic decarbo ymethyl elenolic acid	Polyphenol (tyrosol)	II
195.0665	Negative	$[M-H]^-$	$C_{10}H_{12}O_4$	5.5	0.912	99.88	59.0139; 135.0450; 153.0555	Hydro ytyrosol acetate	Polyphenol (tyrosol)	II
241.0717	Negative	$[M-H]^-$	$C_{11}H_{14}O_6$	5.5	-0.462	99.02	139.0036; 196.7264	Elenolic acid	Polyphenol (tyrosol)	II

285.0406	Negative	$[M-H]^-$	$C_{15}H_{10}O_6$	11.5	0.452	99.59	107.0140; 133.0295; 151.0033; 175.0400; 199.0404; 217.0501; 241.0506; 257.0451	Luteolin	Polyphenol (flavonoid - flavone)	II
319.1187	Negative	$[M-H]^-$	$C_{17}H_{20}O_6$	8.5	0.089	99.43	59.0139; 69.0346; 85.0296; 95.0502; 111.0088; 121.0297; 155.0713; 183.0663; 195.0675; 199.0613	Oleacein	Polyphenol (tyrosol)	II
361.1295	Negative	$[M-H]^-$	$C_{19}H_{22}O_7$	9.5	0.730	98.15	69.0346; 95.0501; 101.0244; 111.0088; 127.0401; 259.0976; 291.0876	Ligstroside aglycone	Polyphenol (tyrosol)	II
377.1244	Negative	$[M-H]^-$	$C_{19}H_{22}O_8$	9.5	0.528	97.00	59.0139; 68.9982; 95.0503; 111.0088; 139.0037; 149.0242; 275.0913; 307.0826; 327.0875; 345.0980	Oleuropein aglycone	Polyphenol (tyrosol)	II
391.1402	Negative	$[M-H]^-$	$C_{20}H_{24}O_8$	9.5	0.816	99.48	59.0138; 68.9980; 95.0502; 111.0090; 139.0033; 149.0242; 275.0910; 321.0980; 340.5735; 349.3463	Methyl oleuropein aglycone	Polyphenol (tyrosol)	II
393.1194	Negative	$[M-H]^-$	$C_{19}H_{22}O_9$	9.5	0.851	98.51	59.0139; 68.9982; 95.0504; 109.0659; 137.0607; 181.0507; 289.1077; 317.1033; 361.0935	Hydro y oleuropein aglycone	Polyphenol (tyrosol)	II
471.3482	Negative	$[M-H]^-$	$C_{30}H_{48}O_4$	7.5	0.439	97.46	393.3153; 423.3277	Maslinic acid	Isoprenoid (triterpene)	II
121.0650	Positive	$[M+H]^+$	C_8H_8O	4.5	1.392	99.76	65.0395; 77.0394; 79.0549; 91.0550; 93.0706; 95.0498; 103.0548; 105.0453	Acetophenone	Ketone	II
137.0596	Positive	$[M+H]^+$	$C_8H_8O_2$	4.5	-0.628	99.68	63.0242; 65.0395; 91.0550; 119.0496	3-methylbenzoic acid	Carbo ylic acid	II
							67.0550; 79.0551; 81.0707; 94.0419; 109.0653	2-metho ybenzaldehyde	Benzaldehyde	II
165.0545	Positive	$[M+H]^+$	$C_9H_8O_3$	5.5	-1.034	99.91	133.0289; 137.0600; 147.0441	Dihydro yphenylpropenal	Phenylacetaldehyde	II
169.0857	Positive	$[M+H]^+$	$C_9H_{12}O_3$	3.5	-1.306	99.19	139.0755; 151.0756	Homovanillyl alcohol	Polyphenol (hydro yphenylalcohol)	II
195.0650	Positive	$[M+H]^+$	$C_{10}H_{10}O_4$	5.5	-0.796	98.64	125.0602; 153.0549; 167.0708; 177.0550	Ferulic acid	Polyphenol (phenolic acid - hydro ycinnamic acid)	II
265.0679	Positive	$[M+Na]^+$	$C_{11}H_{14}O_6$	4.5	-1.280	99.78		Elenolic acid	Polyphenol (tyrosol)	IV
327.1201	Positive	$[M+Na]^+$	$C_{17}H_{20}O_5$	7.5	-0.565	98.45		Oleocanthal	Polyphenol (tyrosol)	IV

343.1148	Positive	$[M+Na]^+$	$C_{17}H_{20}O_6$	7.5	-1.223	98.77	Oleacein	Polyphenol (tyrosol)	IV

3.2.7. PUBLICATION X

Differential mobility spectrometry coupled to mass spectrometry (DMS–MS) for the classification of Spanish PDO paprika.

Campmajó, G.; Saurina, J.; Núñez, O.; Sentellas, S.

Food Chemistry. 2022, 390, 133141.

Food Chemistry 390 (2022) 133141

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Differential mobility spectrometry coupled to mass spectrometry (DMS−MS) for the classification of Spanish PDO paprika

Guillem Campmajó [a,b], Javier Saurina [a,b], Oscar Núñez [a,b], Sonia Sentellas [a,b,c,·]

[a] Department of Chemical Engineering and Analytical ChemistryJ University of BarcelonaJ Martí i Franquès 1-11J E08028 BarcelonaJ Spain
[b] Research Institute in Food Nutrition and Food SafetyJ University of BarcelonaJ Av. Prat de la Riba 171J Edifici Recerca (Gaudí)J E08921 Santa Coloma de GramenetJ Spain
[c] Serra Húnter FellowJ Generalitat de CatalunyaJ Rambla de Catalunya 19-21J E08007 BarcelonaJ Spain

ARTICLEINFO

eywords
Ion mobility spectrometry
DMS−MS
Fingerprinting
Chemometrics
Food authentication
Paprika

ABSTRACT

Ion mobility spectrometry (IMS) has proved its huge potential in many research areas, especially when hyphenated with chromatographic techniques or mass spectrometry (MS). However, focusing on food analysis, and particularly in classification and authentication issues, very few applications have been reported. In this study, differential mobility spectrometry coupled to mass spectrometry (DMS−MS) is presented for the first time as an alternative and high-throughput technique for food classification and authentication purposes using a fingerprinting strategy. As a study case, 70 Spanish paprika samples (from *La Vera*, *Murcia*, and *Mallorca*) were analyzed by DMS−MS to address their classification —using partial least squares regression-discriminant analysis (PLS-DA)— and authentication —through soft independent modeling of class analogy (SIMCA). As a result, after external validation, complete sample classification according to their geographical origin and excellent *La Vera* and *Mallorca* sample authentication were reached.

1. Introduction

In the early 20th century, the first studies using ion mobility spectrometry (IMS) were conducted. However, it was not until the 1970s that Cohen and Karasek introduced it as an analytical tool (Cohen & Karasek, 1970; Karasek, 1974). Since then, and especially in the last two decades, this platform has attracted the interest of scientists as a powerful separation technique, owing to its capacity of separating strongly related compounds. In IMS, ions are separated in the gas phase based on their mobility, which depends on their charge, size, and shape (D'Atri et al., 2018; Dodds & Baker, 2019; Eiceman, Karpas, & Hill, 2013; Gabelica & Marklund, 2018; Kirk, 2019). However, the specific separation mechanism differs among the different platforms (manufacturer), being possible to establish three different categories. (i) Time-dispersive techniques, which encompass drift-time ion mobility spectrometry (DTIMS) and traveling wave ion mobility spectrometry (TWIMS), separate ions based on the time they require to go through the same pathway. (ii) Space-dispersive techniques, such as field asymmetric waveform ion mobility spectrometry (FAIMS), differential mobility spectrometry (DMS), and differential mobility analysis (DMA), rely on the different trajectories that ions describe based on their mobility. (iii) In ion-trapping with selective release techniques, such as trapped ion mobility spectrometry (TIMS), the ions are trapped in a pressurized region and are selectively ejected based on their mobilities (D'Atri et al., 2018).

In the last years, the interest in the hyphenation of IMS with other techniques has spectacularly grown. In this line, ion mobility spectrometry coupled to mass spectrometry (IMS−MS) combines the separation capacity based on the mobility of ions with the structural information provided by mass spectrometry. Beyond that, the addition of a third separation dimension, such as liquid chromatography (LC), opens excellent possibilities for analyzing complex samples. Indeed, IMS, as a standalone technique as well as coupled to LC and MS, has been extensively used in a wide range of research areas, from biomedical or pharmaceutical applications to environmental and security fields (Armenta, Esteve-Turrillas, & Alcalà, 2020; Cossoul et al., 2015; Hernández-Mesa, Escourrou, Monteau, Le Bizec, & Dervilly-Pinel, 2017; Odenkirk & Baker, 2084; To, Ben-Jaber, & Parkin, 2020).

In food analysis, where chromatographic techniques (often hyphenated to MS) are still the gold standard for determining a wide range

· Corresponding author at: Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, Martí i Franquès 1-11, E08028 Barcelona, Spain.
E-mail address sonia.sentellas@ub.edu (S. Sentellas).

https://doi.org/10.1016/j.foodchem.2022.133141
Received 5 January 2022; Received in revised form 4 April 2022; Accepted 1 May 2022
Available online 6 May 2022

of compounds, from natural components to additives or contaminants, IMS begins to be seen as an alternative. Although it can be used as a standalone technique, its combination with LC, gas chromatography (GC), or MS is usually preferred (Alikord, Mohammadi, Kamankesh, & Shariatifar, 2021; Domínguez, Frenich, & Romero-González, 2020; Hernández-Mesa et al., 2019). Until now, IMS has been mainly focused on the determination of specific compounds (targeted analysis) (Rue, Glinski, Glinski, & van Breemen, 2020; Wang, Harrington, Chang, Wu, & Chen, 2020; Will, Behrens, Macke, Quarles, & Karst, 2021), and fewer applications have been reported following a fingerprinting approach (non-targeted analysis) (Burnum-Johnson et al., 2019; Freire et al., 2021; Paglia, Smith, & Astarita, 2021). In this regard, the -omics strategy has been used basically for authentication and food integrity assessment, being DTIMS the preferred mode of IMS (Martín-Gómez & Arce, 2021). However, other separation mechanisms, such as space-dispersive methods, can also be exploited for fingerprinting analysis (Piñero et al., 2020).

DMS separates the ions based on their differential mobility under the influence of a high asymmetric radiofrequency. Under these conditions, only those ions with the proper mobility can describe the correct trajectory to traverse the DMS cell while interferences are deviated into the cell walls. A compensation voltage (CoV) is then applied to correct the trajectory of the ions letting only target ions enter the mass spectrometer. The main application of this technique deals with improving the method sensitivity by reducing background noise and removing isobaric interferences (Bravo-Veyrat & Hopfgartner, 2018; Dempsey, Moeller, & Poklis, 2018; Su et al., 2021). However, to our knowledge, DMS, or more specifically differential mobility spectrometry coupled to mass spectrometry (DMS–MS), has not been previously used for non-targeted analysis in food research. The good reproducibility, speed, and high separation capacity of this technique offer an attractive alternative not only to those well-established chromatographic methods but also to direct MS techniques, such as flow injection analysis coupled to high-resolution mass spectrometry (FIA–HRMS) or ambient mass spectrometry (AMS) (Campmajó, Saurina, & Núñez, 2021; Ibáñez, Simó, García-Cañas, Acunha, & Cifuentes, 2015).

This manuscript aimed at evaluating the applicability of direct infusion DMS–MS fingerprinting for food classification and authentication purposes, using Spanish paprika samples as a case study. In this regard, paprika is a red powdered spice, obtained from red pepper fruits of the genus *Capsicum* (Solanaceae family), widely used because of its characteristic organoleptic properties. Currently, only three paprika products are registered in Spain with the protected designation of origin (PDO) status: *Pimentón de La Vera*, *Pimentón de Murcia*, and *Pebre bord de Mallorca*. Although it ensures high-quality products, it also leads to higher prices, making them vulnerable to food fraud practices. To date, several fingerprinting methods based on LC, with spectroscopic detection or coupled to HRMS, have been developed to address paprika classification (Barbosa, Saurina, Puignou, & Núñez, 2020; Campmajó, Rodríguez-Javier, Saurina, & Núñez, 2021). In this study, the DMS–MS fingerprints of 70 paprika samples from *La Vera*, *Murcia*, and *Mallorca* PDOs, were used for the first time to classify and authenticate them through partial least squares regression-discriminant analysis (PLS-DA) and soft independent modeling of class analogy (SIMCA), respectively.

2. Materials and methods

2.1. Reagents and solutions

Regarding the sample treatment, water was purified using an Elix® 3 coupled to a Milli-Q® system (Millipore Corporation, Bedford, MA, USA) and filtered through a 0.22-µm nylon membrane, while UHPLC-supergradient acetonitrile was purchased from Panreac (Castellar del Vallès, Spain). For DMS optimization, technical grade acetone and UHPLC-supergradient acetonitrile and methanol obtained from Panreac and 2-propanol obtained from Merck (Darmstadt, Germany) were used. Phenolic compounds used in the DMS optimization were purchased from different suppliers: quercetin dihydrate from Riedel-de-Haën (Seelze, Germany); chlorogenic acid from HWI Analytik GMBH (Ruelzheim, Germany); gallic, homogentisic, and ferulic acids, and vanillin from Fluka (Steinheim, Germany); and D-(–)-quinic, caffeic, homovanillic, *p*-coumaric, sinapic, and betulinic acids, syringaldehyde, protocatechuic aldehyde, and rutin from Merck.

2.2. Instrumentation

A 5500 Qtrap (AB Sciex, Framingham, MA, USA) mass spectrometer with an electrospray ion source (ESI) and the SelexION differential mobility separation device (DMS, AB Sciex, Framingham, MA, USA), installed between the ionization source and the vacuum interface, was used for the analysis of samples. Paprika extracts were directly introduced for 1.6 min by infusion to the ionizations source at a rate of 5 $\mu L \cdot min^{-1}$ using the integrated syringe pump.

Regarding DMS conditions, the temperature was fixed at 225 °C (medium), the separation voltage (SV) and DMS offset (DMO) were set at 2500 V and 3 V, respectively, and the high DMS resolution enhancement option was selected. Under these conditions, a CoV ramp (from –10 to 7 V) was performed. MS detection in negative full-scan MS mode (Enhance MS, EMS) was used from 100 to 650 *m/z* at a scan rate of 1000 $Da \cdot s^{-1}$. Nitrogen, used as nebulizer and auxiliary gas, was set at 20, 15, and 0 arbitrary units for the curtain gas, the ion source gas 1, and the ion source gas 2, respectively. Besides, the ion spray voltage (IS) was set at – 4500 V without heating the ion source, and the declustering potential (DP) and the entrance potential (EP) were fixed at –100 V and –10 V, respectively. Analyst software (version 1.6.2) from AB Sciex was used for instrument control and data acquisition.

2.3. SamplesJ sample treatmentJ and sample analysis

In this study, 70 paprika samples belonging to the three Spanish regions with the PDO label were analyzed. In this line, 30 *La Vera* samples (10 of each paprika type: hot, sweet, and bittersweet) were directly purchased from paprika production companies, and 20 *Murcia* and 20 *Mallorca* samples (containing half hot and half sweet types, each one) were bought in Spanish commercial markets.

Samples were subjected to an ultrasound-assisted solid–liquid extraction (USLE) method, previously described by Cetó et al. (2018), using water:acetonitrile (20:80, *v/v*) as extracting solvent.

Samples were randomly analyzed to reduce the impact of any potential instrumental drift on the chemometric results. Moreover, a quality control (QC) sample, prepared by mixing 50 µL of each paprika sample extract, was also analyzed (at the beginning and every ten samples) to check for systematic errors along the analysis.

2.4. Data analysis

2.4.1. Data matrix construction

Raw data were converted to mzXML format using the MSConvertGUI software (Chambers et al., 2012). Then, aiming at constructing a data matrix containing DMS–MS fingerprints, data were processed using the mzMine 2.53 software (Pluskal, Castillo, Villar-Briones, & Orešič, 2010). Firstly, nominal mass detection centroided each mass spectrum acquired for a sample, through the "Wavelet transform" algorithm (establishing a noise level of 2.0×10^4, a scale level of 3, and a wavelet window size of 30%), particularly suitable for noisy low-resolution mass spectrometry (LRMS) data. Secondly, using the option of chromatogram builder, nominal mass signals found in at least 5 contiguous scans for a sample were connected, with a group intensity threshold of 3.0×10^4, a minimum highest intensity of 7.0×10^4, and an *m/z* tolerance of 800 ppm. Thirdly, each ionogram was then deconvoluted into individual peaks, using the "Baseline cut-off" algorithm that recognized peaks with a CoV duration range between 0.4 and 2.0 V and fulfilling the peak intensity

conditions established in the previous step. Fourthly, isotopes were removed, considering the most intense isotope as the most representative and setting an *m/z* tolerance of 800 ppm. Finally, the random sample consensus (RANSAC) aligner matched *m/z* signals detected across samples, establishing a mass tolerance of 1000 ppm, peak CoV tolerances of 1.5 and 0.5 V (before and after correction, respectively), and a minimum number of points of 80%. In the end, an X-data matrix containing DMS–MS fingerprints (ion peak area matrix in which each row corresponds to a sample (78 samples) and each column corresponds to a variable (203), being a variable a specific ion (specific *m/z*) migrating at a specific CoV) of the studied samples was obtained (Table S1).

2.4.2. Chemometric analysis

Principal component analysis (PCA), PLS-DA, and SIMCA were carried out using Solo 8.6 chemometrics software from Eigenvector Research (Manson, WA, USA). Details of the theoretical background of these chemometric methods are given elsewhere (Massart et al., 1997; Wold, 1976). In this paper, we will only make a brief description of the most relevant aspects related to our study.

Independently of the chemometric method employed, an X-data matrix was required, consisting of DMS–MS fingerprints. Moreover, data were autoscaled before the chemometric analysis to suppress differences in each variable's magnitude and amplitude scales. For such a purpose, data were mean-centered and subsequently divided by the standard deviation of the corresponding variable according to the following expression:

$$x_{iautoscaled} = \frac{x_i - \bar{x}}{s}$$

where $x_{i\ autoscaled}$ is the value for variable *i* after autoscaling, x_i is the original value for variable *i*, and $\bar{X}$ and *s* are the mean value and standard deviation, respectively.

A preliminary exploration of DMS-MS fingerprints by PCA assessed the behavior of samples and variables. PCA concentrates the relevant information of the X-matrix, contained in a large number of experimental variables, into a reduced number of mathematical variables called principal components (PCs). PCA relies on decomposing the experimental data matrix into two smaller submatrices of scores (T, with coordinates of the samples) and loadings (P^T, with the eigenvectors or coordinates of the variables). As a result, the scatter plot of scores on PC space depicts the sample layout which may reveal similarities and differences among sample characteristics such as origins and varieties. The loadings' plots may figure out the most descriptive variables and their correlations.

The supervised sample classification (according to geographical origin and type) was studied through PLS-DA and evaluated after external validation through sensitivity, specificity, and accuracy. The experimental X-matrix is correlated with the class matrix that encodes the sample membership to its class. The classification model is established to obtain the minimum error in the sample assignation to the corresponding classes. Here, the optimal number of latent variables (LVs) used in each PLS-DA model was established at the first significant minimum point of cross-validation (CV) error using the Venetian blinds method. Subsequently, the classification performance was assessed by external validation using independent test samples.

SIMCA was proposed for paprika sample authentication. SIMCA relies on a PCA model constructed using only samples belonging to a given class. Hence, a PCA model is obtained, for instance, with la Vera, Murcia, or Mallorca samples. Reduced Q residuals and Hotelling T^2 values, normalized to a 95% confidence limit, were used to calculate the distance between a new projected sample and the established PCA submodel. The number of PCs used in the PCA submodel, as well as the decision threshold, were optimized in each SIMCA model by maximizing the calibration step performance. Then, both the distance and the decision threshold assessed the sample class membership. Moreover, considering that SIMCA calibration models were built with less than 20 samples, the leave-one-out method was proposed for CV. Finally, as with PLS-DA, SIMCA models' performance was assessed by external validation.

3. Results and discussion

3.1. Selection of the DMS–MS conditions

A mixture of 15 phenolic compounds (namely chlorogenic, gallic, homogentisic, ferulic, D-(–)-quinic, caffeic, homovanillic, p-coumaric, sinapic, and betulinic acids, vanillin, quercetin, syringaldehyde, protocatechuic aldehyde, and rutin), previously identified as possible key components for the classification of paprika samples (Barbosa, Campmajó, Saurina, Puignou, & Núñez, 2020), was used to set DMS–MS conditions. As commented before, this work was not focused on optimizing the polyphenol separation but on the untargeted analysis of paprika samples to obtain characteristic sample fingerprints to discriminate samples according to geographical origin and type.

With this in mind, the polyphenol standard solution (10 $mg{\cdot}L^{-1}$ each compound) was introduced by infusion to the mass spectrometer through the SelexION differential mobility separation device to establish the DMS–MS parameters. Then, a negative full-scan using Q3 in ion trap mode (EMS) was recorded from 100 to 650 *m/z*. Total ion intensity was monitored to evaluate MS parameters such as DP and IS, obtaining the maximum signal intensity at 100 V and 4500 V, respectively. Following the same criterion, DMS temperature was fixed at 225 °C (medium).

The separation of the polyphenols included in the standard mixture was studied to choose the most appropriate SV. Hence, a ramp of the CoV was performed at different SV (from 1000 V to 4000 V, in steps of 500 V). As a result, the higher the SV, the higher degree of separation was observed. However, the signal intensity was strongly affected (Fig. S1), and therefore, as a compromise, an SV of 2500 V was selected to analyze paprika samples.

Additionally, several gas modifiers —namely methanol, acetonitrile, 2-propanol, and acetone— were evaluated. Volatile reagents introduced into the gas flow may interact differently with the ions to form clusters, thus modifying their mobility and affecting both separation and signal intensity. Hence, the effect of each modifier on the DMS separation of the selected polyphenols was studied at low (1.5%) and high (3.0%) modifier concentrations. However, no significant improvement was observed in any case. Therefore, and considering that this study aimed to use a simple method to generate discriminating sample DMS–MS fingerprints, the addition of a gas modifier was discarded.

Fig. 1 shows representative DMS–MS fingerprints for a hot *La Vera*, *Murcia*, and *Mallorca* PDO paprika sample. Several qualitative differences in the ionograms regarding compounds detected and their peak intensity can be observed. In this context, *Mallorca* paprika samples presented the most distinctive DMS–MS fingerprints. However, remarkable dissimilarities were encountered among *La Vera* and *Murcia* samples. Therefore, the DMS–MS fingerprints were proposed as chemical descriptors for further multivariate chemometric analysis.

3.2. Non-supervised and supervised chemometric analysis

After the visual inspection of the DMS–MS raw data, a 78 × 203 (samples × variables) data matrix was constructed following the procedure described in Section 2.4.1. Then, PCA was used for an exploratory assessment of the behavior of samples and QCs. Fig. S2 presents the PCA scatter plot of scores on the PC2-PC1 (describing 28.06% of the variance), where QC samples were in a compact cluster, discarding the presence of a systematic error (e.g., a shift in the analytical system) and, thus validating the subsequent chemometric results. *Mallorca* paprika samples were located on the top of the plot, displaying positive PC2

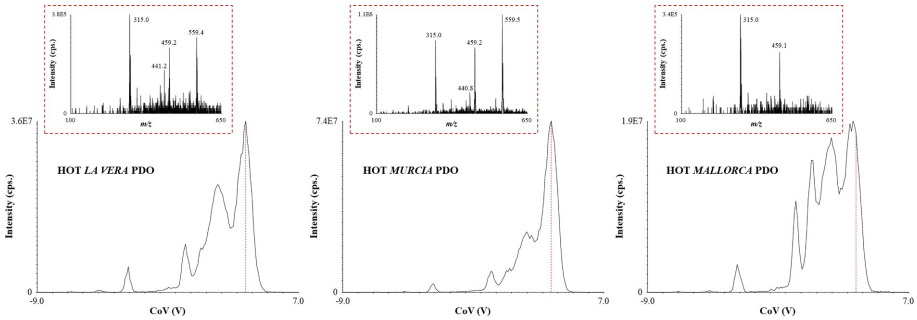


Fig. 1. DMS−MS fingerprints obtained for a hot La Vera, Murcia, and Mallorca PDO paprika sample at SV 2500 V, and MS full-scan spectra at CoV 3.58 V.

values, while *La Vera* and *Murcia* ones appeared jointly (Fig. S2A). Moreover, the PCA score plot noticed no sample distribution according to paprika type (Fig. S2B).

Once demonstrated the good performance of the analysis, QC samples were removed from the DMS−MS data matrix for the supervised chemometric classification carried out through PLS-DA, resulting in a dimension of 70×203 (samples ×variables). Firstly, sample classification according to the geographical origin (*La Vera*, *Murcia*, and *Mallorca*) was studied. In this context, a PLS-DA model built with two LVs (explaining a Y-variance of 58.51%) allowed an apparent distinction between the three Spanish regions under study, thus improving the non-supervised chemometric results given above. Hence, in the obtained scores plot of LV1 *vs*. LV2, LV2 values allowed the isolation of *Murcia* samples, located at the bottom of the diagram, while the separation of *La Vera* and *Mallorca* samples (placed on the top of the diagram, displaying positive LV2 values) was mainly attributed to LV1 values (Fig. 2A).

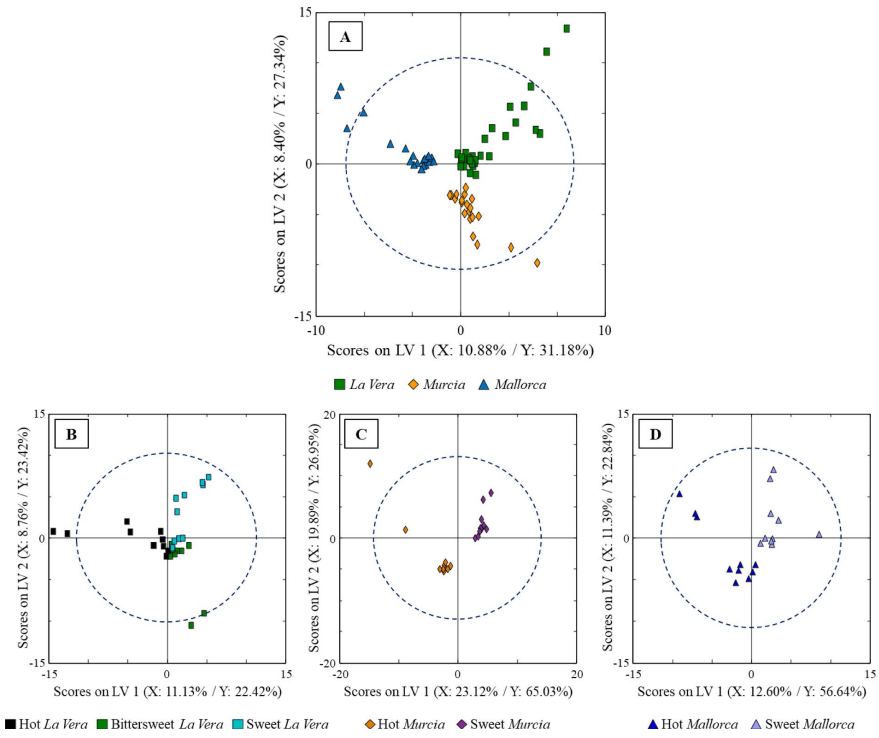


Fig. 2. PLS-DA scores plot of LV1 *vs*. LV2 obtained for the analyzed paprika samples according to their geographical origin, using DMS−MS fingerprints (A) and individual *La Vera* (B), *Murcia* (C), and *Mallorca* (D) PLS-DA models to classify samples according to their type.

External validation was carried out to evaluate the ability of the PLS-DA model established to classify paprika samples from the three Spanish regions distinguished with the PDO label. Thus, 60% of samples were randomly stratified and used as the calibration set, while the remaining 40% were used as the external validation set. As a result, a classification rate of 100% was achieved, being [12, 0, 0; 0, 8, 0; 0, 0, 8] the confusion matrix for the established PLS-DA model. Please note that rows in the confusion matrix correspond to *La Vera*, *Murcia*, and *Mallorca* classes, respectively, and columns are given in this same order. Hence, the 12 *La Vera*, the 8 Murcia, and the 8 *Mallorca* samples were correctly classified into their respective classes.

Additionally, sample classification regarding paprika's type (hot, bittersweet, or sweet) was also evaluated by PLS-DA. In this line, Fig. S3 contains the plot of scores of LV1 *vs*. LV2 obtained after assigning each sample to its class, considering both geographical origin and type. As can be seen, although *Murcia* and *Mallorca* paprika samples seemed to follow a trend related to the sample type (*La Vera* samples appeared jointly in the plot), the similarities due to the geographical origin prevailed. Therefore, to better analyze sample grouping depending on the product type, individual PLS-DA models were built for each region under study. As a result, as shown in Fig. 2B, DMS−MS fingerprints allowed *La Vera* samples separation according to their type. To our knowledge, this separation has only been previously achieved using an untargeted ultra-high-performance liquid chromatography coupled to high-resolution mass spectrometry (UHPLC−HRMS) method (Barbosa et al., 2020). However, the UHPLC-HRMS method required more expensive and complex instrumentation than the method proposed here. In addition, 30 min were needed for analyzing each sample, while only 1.6 min were required using the DMS-MS method. Both PLS-DA scatter plots of scores of *Murcia* (Fig. 2C) and *Mallorca* (Fig. 2D) samples also discriminate between their hot and sweet type. However, in this case, no external validation could be performed because of the scarcity of samples for each class.

Finally, considering the excellent classificatory results obtained with PLS-DA, SIMCA was proposed as a one-class modeling chemometric technique to assess the authenticity of the Spanish paprika samples according to their geographical origin based on DMS−MS fingerprints. Again, as in the PLS-DA study, DMS−MS data was split into the calibration set (42× 203, samples× variables) and the validation set (28 × 203, samples × variables). Table 1 shows the number of PCs and the decision threshold selected in each SIMCA model, as well as the authentication performance in terms of sensitivity, specificity, and accuracy after the external validation. The developed *La Vera* and *Mallorca* SIMCA models provided good accuracy results, although specificity and sensitivity were more limited. Instead, assignation results were poorer with the *Murcia* SIMCA model.

A variable selection strategy was applied to improve these results, given that the first PCs of the SIMCA model do not necessarily contain discriminant information. Thus, a new data matrix was constructed, containing only the 10 variables with the highest selectivity ratio among the 20 ones with the highest variable importance in projection (VIP) values obtained for each paprika geographical origin class in the previous PLS-DA model. As a result, considering that some variables were simultaneously relevant for two of the studied classes, 42 × 25 (samples × variables) and 28× 25 (samples× variables) calibration and external validation data matrices were built, respectively. As observed in Table 1, the applied variable selection strategy remarkably improved the assignation accuracy of the SIMCA models in all the cases. In this context, excellent results were obtained for *La Vera* (assignation rate of 92.9%) and *Mallorca* (assignation rate of 100.0%) authentication because of an enhancement in the specificity and sensitivity results, respectively, when using the reduced data matrix. Instead, although better assignation performance was achieved in the *Murcia* SIMCA model due to good specificity values, poor sensitivity values were obtained.

4. Conclusions

As commented before, LC or GC, often coupled to MS, is the preferred separation technique when dealing with classification or authentication in food analysis. The separation capacity of IMS to separate closely related compounds is well-known; however, not much research has been done using this technique for dealing with food classification or authentication issues. In fact, from our point of view, considering the separation potential of this technique, there is still a long way to go. Separations by IMS, and more specifically by DMS, offer faster and greener alternatives to the widely used LC counterpart. With this in mind, the applicability of DMS−MS was evaluated in this manuscript. It is worth highlighting the short analysis time required for sample analysis (1.6 min per sample). From the results obtained, we conclude that this technique was satisfactorily applied for the first time to a food classification and authentication issue through a fingerprinting approach combined with chemometrics. With such a purpose, 70 paprika samples from the three Spanish regions distinguished with the PDO label (*La Vera*, *Murcia*, and *Mallorca*) were used as a case study. Sample classification according to geographical origin and type was achieved by subjecting DMS−MS fingerprints to PLS-DA. In this context, in the first case, a classification accuracy of 100% was reached after external validation. Moreover, SIMCA results proved the ability of DMS−MS fingerprinting to authenticate *La Vera* and *Mallorca* paprika samples, especially after applying a previous variable selection strategy. Therefore, with this study, DMS−MS has been demonstrated to be a reliable high-throughput alternative to other currently applied techniques.

Funding

This research was funded by the "Agencia Estatal de Investigación", grant number PID2020-114401RB-C22.

CRediT authorship contribution statement

Guillem Campmajó: Methodology, Formal analysis, Investigation, Writing – original draft. **Javier Saurina:** Formal analysis, Writing – review & editing. **Oscar Núñez:** Conceptualization, Writing – review & editing. **Sonia Sentellas:** Conceptualization, Writing – original draft, Writing – review & editing.

Table 1
Calibration model parameters —number of PCs and established decision threshold— and external validation results —class sensitivity (%), class specificity (%), and global accuracy (%)— for each of the SIMCA models built.

	LA VERA					*MURCIA*					*MALL RCA*				
	Calibration model		External validation			Calibration model		External validation			Calibration model		External validation		
	PCs	Threshold	Sens.	Spec.	Accuracy	PCs	Threshold	Sens.	Spec.	Accuracy	PCs	Threshold	Sens.	Spec.	Accuracy
Non-reduced matrix	3	0.5	83.33	68.75	75.00	1	0.5	50.00	55.00	53.57	5	0.5	62.50	100.0	89.29
Reduced matrix	2	0.5	83.33	100.0	92.86	6	0.5	25.00	85.00	67.86	3	0.5	100.0	100.0	100.0

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors are grateful for the support received from the Agency for Administration of University and Research Grants (Generalitat de Catalunya), through the grant 2017SGR-171. G. Campmajó thanks the University of Barcelona for the PhD APIF fellowship.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2022.133141.

References

Alikord, M., Mohammadi, A., Kamankesh, M., & Shariatifar, N. (2021). Food safety and quality assessment: Comprehensive review and recent trends in the applications of ion mobility spectrometry (IMS). *Critical Reviews in Food Science and Nutrition*, 1–34. https://doi.org/10.1080/10408398.2021.1879003

Armenta, S., Esteve-Turrillas, F. A., & Alcalà, M. (2020). Analysis of hazardous chemicals by "stand alone" drift tube ion mobility spectrometry: A review. *Analytical MethodsJ 12*, 1163–1181. https://doi.org/10.1039/c9ay02268f

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 Agriculture and Food ChemistryJ 68*. https://doi.org/10.1021/acs.jafc.9b06054

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.J 9*. https://doi.org/10.3390/foods9040486

Bravo-Veyrat, S., & Hopfgartner, G. (2018). High-throughput liquid chromatography differential mobility spectrometry mass spectrometry for bioanalysis: Determination of reduced and oxidized form of glutathione in human blood. *Analytical and Bioanalytical ChemistryJ 410*, 7153–7161. https://doi.org/10.1007/s00216-018-1318-x

Burnum-Johnson, K. E., Zheng, X., Dodds, J. N., Ash, J., Fourches, D., Nicora, C. D., ... Baker, E. S. (2019). Ion mobility spectrometry and the omics: Distinguishing isomers, molecular classes and contaminant ions in complex samples. *TrAC - Trends in Analytical ChemistryJ 116*, 292–299. https://doi.org/10.1016/j.trac.2019.04.022

Campmajó, G., Rodríguez-Javier, L. R., Saurina, J., & Núñez, O. (2021). Assessment of paprika geographical origin fraud by high-performance liquid chromatography with fluorescence detection (HPLC-FLD) fingerprinting. *Food ChemistryJ 352*. https://doi.org/10.1016/j.foodchem.2021.129397

Campmajó, G., Saurina, J., & Núñez, O. (2021). FIA –HRMS fingerprinting subjected to chemometrics as a valuable tool to address food classification and authentication: Application to red wine, paprika, and vegetable oil samples. *Food ChemistryJ 373*, Article 131491. https://doi.org/10.1016/j.foodchem.2021.131491

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)J 18*. https://doi.org/10.3390/s18124479

Chambers, M. C., MacLean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., ... Mallick, P. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nature BiotechnologyJ 30*, 918–920. https://doi.org/10.1038/nbt.2377

Cohen, M. J., & Karasek, F. W. (1970). Plasma chromatography TM—new dimension for gas chromatography and mass spectrometry. *Journal of Chromatographic ScienceJ 8*, 330–337. https://doi.org/10.1093/chromsci/8.6.330

Cossoul, E., Hubert-Roux, M., Sebban, M., Churlaud, F., Oulyadi, H., & Afonso, C. (2015). Evaluation of atmospheric solid analysis probe ionization coupled to ion mobility mass spectrometry for characterization of poly(ether ether ketone) polymers. *Analytica Chimica ActaJ 856*, 46–53. https://doi.org/10.1016/j.aca.2014.12.013

D'Atri, V., Causon, T., Hernandez-Alba, O., Mutabazi, A., Veuthey, J.-L.-L., Cianferani, S., & Guillarme, D. (2018). Adding a new separation dimension to MS and LC-MS: What is the utility of ion mobility spectrometry? *Journal of Separation ScienceJ 41*, 20–67. https://doi.org/10.1002/jssc.201700919

Dempsey, S. K., Moeller, F. G., & Poklis, J. L. (2018). Rapid separation and quantitation of cocaine and its metabolites in human serum by differential mobility spectrometry-tandem mass spectrometry (DMS-MS-MS). *Journal of Analytical Toxicology.J 42*, 518–524. https://doi.org/10.1093/jat/bky055

Dodds, J. N., & Baker, E. S. (2019). Ion mobility spectrometry: Fundamental concepts, instrumentation, applications, and the road ahead. *Journal of the American Society for Mass SpectrometryJ 30*, 2185–2195. https://doi.org/10.1007/s13361-019-02288-2

Domínguez, I., Frenich, A. G., & Romero-González, R. (2020). Mass spectrometry approaches to ensure food safety. *Analytical MethodsJ 12*, 1148–1162. https://doi.org/10.1039/c9ay02681a

Eiceman, G. A., Karpas, Z., & Hill, H. H. (2013). *Ion Mobility Spectrometry* (Third Edition.). CRC Press.

Freire, R., Fernandez, L., Mallafré-Muro, C., Martín-Gómez, A., Madrid-Gambin, F., Oliveira, L., ... Marco, S. (2021). Full workflows for the analysis of gas chromatography—ion mobility spectrometry in foodomics: Application to the analysis of Iberian ham aroma. *Sensors.J 21*. https://doi.org/10.3390/s21186156

Gabelica, V., & Marklund, E. (2018). Fundamentals of ion mobility spectrometry. *Current pinion in Chemical BiologyJ 42*, 51–59. https://doi.org/10.1016/j.cbpa.2017.10.022

Hernández-Mesa, M., Escourrou, A., Monteau, F., Le Bizec, B., & Dervilly-Pinel, G. (2017). Current applications and perspectives of ion mobility spectrometry to answer chemical food safety issues. *TrAC - Trends in Analytical ChemistryJ 94*, 39–53. https://doi.org/10.1016/j.trac.2017.07.006

Hernández-Mesa, M., Ropartz, D., García-Campaña, A. M., Rogniaux, H., Dervilly-Pinel, G., & Le Bizec, B. (2019). Ion mobility spectrometry in food analysis: Principles, current applications and future trends. *MoleculesJ 24*, 1–29. https://doi.org/10.3390/molecules24152706

Ibáñez, C., Simó, C., García-Cañas, V., Acunha, T., & Cifuentes, A. (2015). The role of direct high-resolution mass spectrometry in foodomics. *Analytical and Bioanalytical ChemistryJ 407*, 6275–6287. https://doi.org/10.1007/s00216-015-8812-1

Karasek, F. W. (1974). Plasma chromatography. *Analytical ChemistryJ 46*. https://doi.org/10.1021/ac60344a724

Kirk, A. T. (2019). Ultra-high-resolution ion mobility spectrometry — current instrumentation, limitations, and future developments. *Analytical and Bioanalytical ChemistryJ 411*, 6229–6246.

Martín-Gómez, A., & Arce, L. (2021). Ion Mobility Spectrometry to Assess Food Authenticity and Integrity. In *Chromatogr. Relat. Sep. Tech. Food Integr. Authent.* (pp. 105–135). WORLD SCIENTIFIC (EUROPE). https://doi.org/10.1142/9781786349958_0005.

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* (first edition). Amsterdam, The Netherlands: Elsevier Science.

Odenkirk, M. T., & Baker, E. S. (2020). Utilizing Drift Tube Ion Mobility Spectrometry for the Evaluation of Metabolites and Xenobiotics. *Methods in Molecular Biology (New orkJ N J United States)J 2084*, 35–54. https://doi.org/10.1007/978-1-0716-0030-6_2

Paglia, G., Smith, A. J., & Astarita, G. (2021). Ion mobility mass spectrometry in the omics era: Challenges and opportunities for metabolomics and lipidomics. *Mass Spectrometry Reviews*. https://doi.org/10.1002/mas.21686

Piñero, M. Y., Amo-González, M., Ballesteros, R. D., Pérez, L. R., de la Mora, G. F., & Arce, L. (2020). Chemical fingerprinting of olive oils by electrospray ionization-differential mobility analysis-mass spectrometry: A new alternative to food authenticity testing. *Journal of the American Society for Mass SpectrometryJ 31*, 527–537. https://doi.org/10.1021/jasms.9b00006

Pluskal, T., Castillo, S., Villar-Briones, A., & Orešič, M. (2010). MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics.J 11*, 395. https://doi.org/10.1186/1471-2105-11-395

Rue, E. A., Glinski, J. A., Glinski, V. B., & van Breemen, R. B. (2020). Ion mobility-mass spectrometry for the separation and analysis of procyanidins. *Journal of Mass SpectrometryJ 55*. https://doi.org/10.1002/jms.4377

Su, B., Bettcher, L. F., Hsieh, W. Y., Hornburg, D., Pearson, M. J., Blomberg, N., ... Williams, K. J. (2021). A DMS Shotgun Lipidomics Workflow Application to Facilitate High-Throughput, Comprehensive Lipidomics. *Journal of the American Society for Mass SpectrometryJ 32*, 2655–2663. https://doi.org/10.1021/jasms.1c00203

To, K. C., Ben-Jaber, S., & Parkin, I. P. (2020). Recent developments in the field of explosive trace detection. *ACS NanoJ 14*, 10804–10833. https://doi.org/10.1021/acsnano.0c01579

Wang, Y., Harrington, P.de B., Chang, T., Wu, X., & Chen, P. (2020). Analysis of cranberry proanthocyanidins using UPLC–ion mobility–high-resolution mass spectrometry. *Analytical and Bioanalytical ChemistryJ 412*, 3653–3662. https://doi.org/10.1007/s00216-020-02601-z

Will, J. M., Behrens, A., Macke, M., Quarles, C. D., & Karst, U. (2021). Automated chiral analysis of amino acids based on chiral derivatization and trapped ion mobility-mass spectrometry. *Analytical ChemistryJ 93*, 878–885. https://doi.org/10.1021/acs.analchem.0c03481

Wold, S. (1976). Pattern recognition by means of disjoint principal components models. *Pattern RecognitionJ 8*, 127–139. https://doi.org/10.1016/0031-3203(76)90014-5

Supplementary Material

Differential mobility spectrometry coupled to mass spectrometry (DMS-MS) for the classification of Spanish PDO paprika

Guillem Campmajó[a,b], Javier Saurina[a,b], Oscar Núñez[a,b], Sonia Sentellas[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: Sonia Sentellas

(Tel.: 34-93-4035423, e-mail: sonia.sentellas@ub.edu)

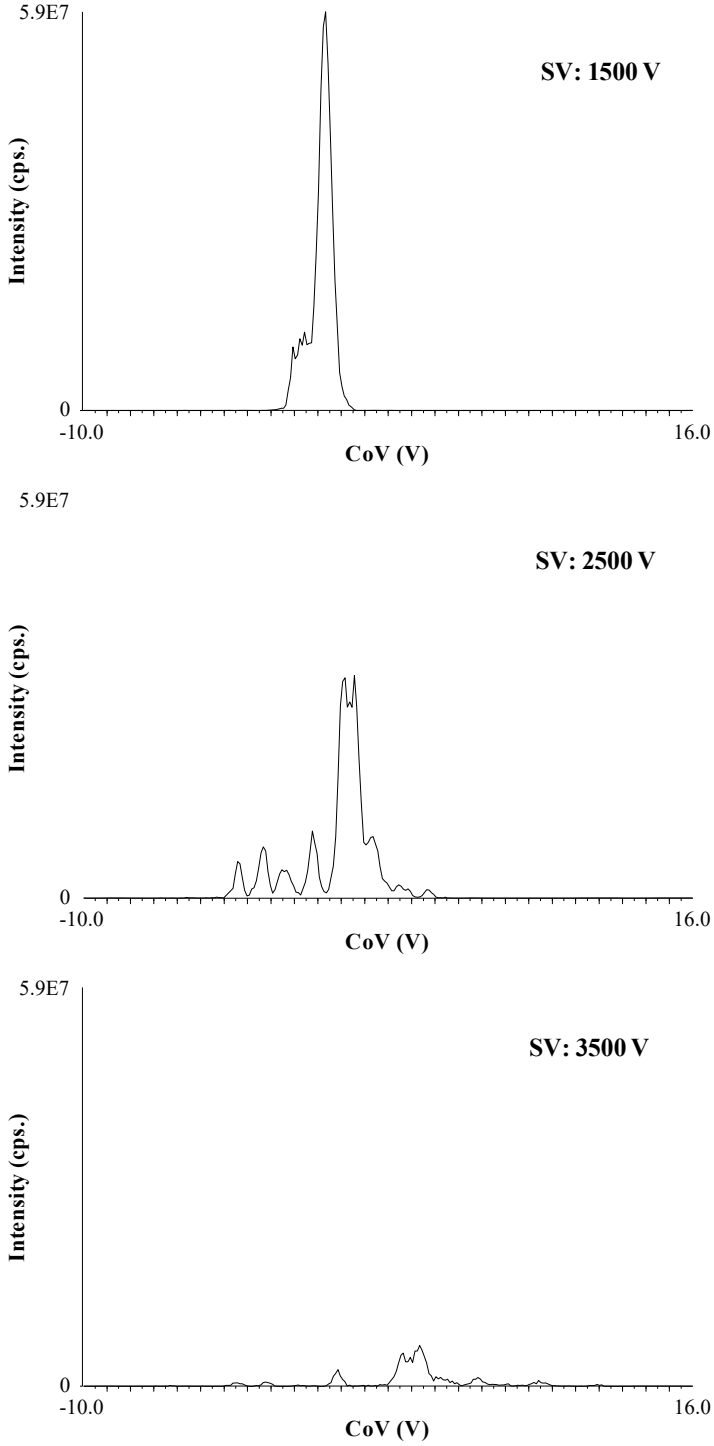


Fig. S1. Polyphenol standard mixture ionograms obtained at SV of 1500, 2500, and 3500 V.

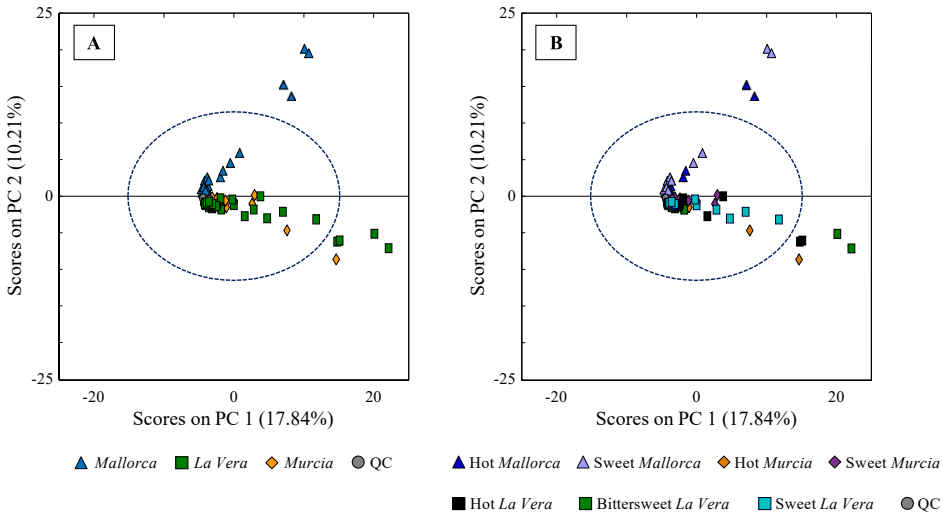


Fig. S2. PCA scores plot of PC1 *vs*. PC2, using DMS-MS fingerprints, for the analysed paprika samples according to (A) their geographical origin and (B) both geographical origin and type.

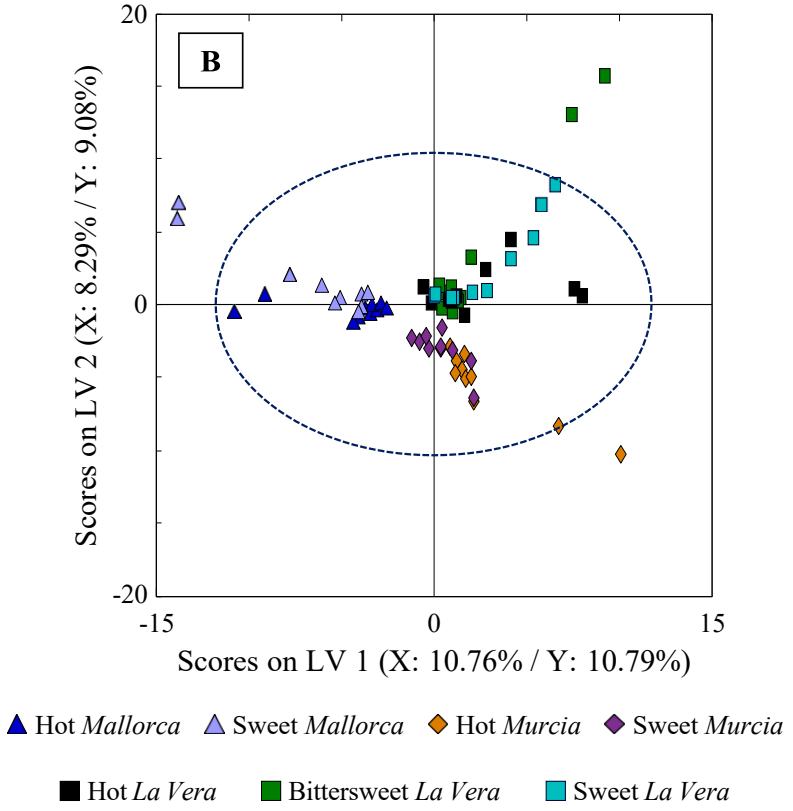


Fig. S3. PLS-DA scores plot of LV1 *vs*. LV2 obtained for the analysed paprika samples according to both their geographical origin and type, using DMS-MS fingerprints.

3.3. DISCUSSION

The publications contained in this chapter expand the application of metabolomic fingerprinting approaches to fraud cases involving nuts, hen eggs, paprika, red wine, and olive oil. As done in Chapter 2, this discussion section complements the already presented scientific results with a general comparison and analysis. In this line, the studies analysing the same food samples are first compared, followed by a discussion regarding the different fingerprinting methodologies used in this thesis.

Nevertheless, before discussing some details of the obtained results, it is remarkable to mention that since it is impossible to cover the entire metabolome and despite the non-targeted nature of the developed methodologies, most of the sample extraction and instrumental conditions were oriented to favour the detection of phenolic and polyphenolic compounds and their derivatives. The fact that, as previously described, these compounds are abundantly found in plant-based food products and have already been successfully proposed as potential markers in multiple fraud issues motivated this decision [27,28].

Therefore, for instance, Publications IV, VI, and VIII describe the analysis of nuts through non-targeted LC-UV, LC-FLD, and LC−HRMS methodologies, respectively. Thus, these studies aimed to achieve the nut type classification and, in some cases, the detection and quantitation of almond-based products adulteration with other nuts (*i.e.*, hazelnut or peanut). Thus, the applied sample treatment, consisting of an SLE extraction with acetone:water (70:30, *v/v*) followed by a defatting step with hexane, was optimised in Publication IV and subsequently implemented in the other studies. In particular, the selection of the extracting solvent was based on the measured total chromatographic peak area in the acquired LC-UV chromatograms at an absorption wavelength of 280 nm (corresponds to the maximum absorption wavelength of several phenolic classes such as hydroxybenzoic acids, flavanols, or flavanones). Hence, this parameter was established to estimate the extracted phytochemical content. Furthermore, the same chromatographic separation

conditions —C_{18} stationary phase column and acidified water with 0.1% formic acid (*v/v*) and methanol as the mobile phase components— were used in each case.

Regarding the chemometric analysis, comparison of the discriminant ability of each acquired chromatographic fingerprint is herein carried out through the results obtained after external validation of the built multi-class PLS-DA model (60% of samples for calibration and 40% for external validation), which includes the ten nut types analysed. These results were not previously provided in Publications IV and VI but were calculated for this discussion. Thus, 46.7%, 98.3%, and 100% classification accuracies were reached for LC-UV, LC-FLD, and LC−HRMS fingerprinting data, respectively. Besides, Table 3.2 provides each case's sensitivity and specificity values. While excellent classification was achieved for LC-FLD (only one walnut sample was misclassified as macadamia nut) and LC−HRMS data, poorer results were obtained in the case of LC-UV fingerprints.

Table 3.2. External validation results —sensitivity (%) and specificity (%)— obtained through the developed multi-class PLS-DA models (comprising the ten analysed nut types) using the LC-UV, LC-FLD, and LC−HRMS fingerprinting data.

	LC-UV		**LC-FLD**		**LC−HRMS**	
	Sensitivity	**Specificity**	**Sensitivity**	**Specificity**	**Sensitivity**	**Specificity**
Almond	8.3	81.3	100.0	100.0	100.0	100.0
Cashew nut	0.0	100.0	100.0	100.0	100.0	100.0
Hazelnut	25.0	98.1	100.0	100.0	100.0	100.0
Macadamia nut	100.0	96.4	100.0	98.2	100.0	100.0
Peanut	87.5	100.0	100.0	100.0	100.0	100.0
Pine nut	75.0	82.1	100.0	100.0	100.0	100.0
Pistachio	0.0	100.0	100.0	100.0	100.0	100.0
Pumpkin seed	37.5	80.8	100.0	100.0	100.0	100.0
Sunflower seed	100.0	100.0	100.0	100.0	100.0	100.0
Walnut	100.0	100.0	75.0	100.0	100.0	100.0

Therefore, focusing on LC-UV fingerprinting, the low classification accuracy could be attributed to the strong influence of walnut and sunflower seed classes in the multi-

class chemometric model. Therefore, in this discussion, to maximise the classification ability of these chromatographic fingerprints and considering that the PLS-DA performance improves when a few classes are studied, the classification decision tree shown in Figure 3.1 is proposed. As a result, an overall classification accuracy of 95% was reached (only 3 samples out of 60 were misclassified), proving that LC-UV fingerprints can also be excellent chemical markers for nut classification when following the proposed strategy.

Furthermore, aside from the global chemometric classification of nuts, Publications VI and VIII also dealt with the detection and quantitation of almond-based products (*i.e.*, almond flour and almond custard cream) adulterations with hazelnut or peanut. However, while LC-FLD fingerprinting was employed in Publication VI, targeted LC−HRMS profiles focusing on specific chemical markers were used instead of fingerprints in Publication VIII. In both cases, the PLS prediction of the level of adulteration provided errors (RMSEP values) below 10%. Furthermore, although the custard cream is a fatter food matrix than the flour, no significant differences between the corresponding prediction results were observed either through LC-FLD fingerprints or the LC−HRMS profiles.

In contrast, Publication V evaluated the ability of LC-UV fingerprinting, combined with chemometrics, to address the classification of hen eggs according to their production system (*i.e.*, eggs from cage hens, barn hens, and free range hens, and organic eggs). In this study, after a straightforward sample extraction with acetonitrile:water (80:20, *v/v*), a 35 min-chromatographic separation was carried out. In this line, previous optimisation of the chromatographic separation was done, aiming to achieve rich LC-UV fingerprints. Similarly, different absorption wavelengths were considered for the UV detection (*i.e.*, 250, 280, 310, 370, and 550 nm), 250 nm being the one providing the best conditions.

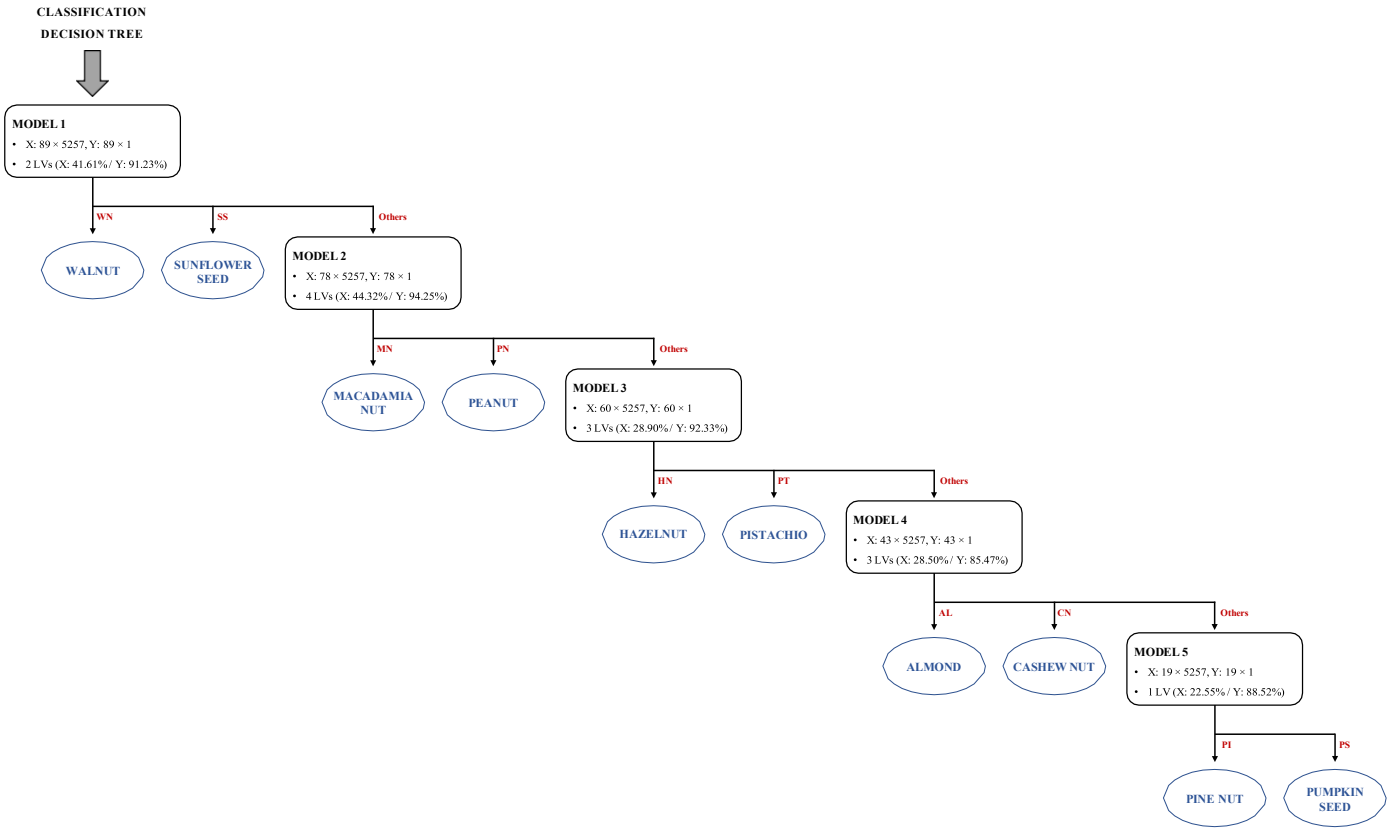


Figure 3.1. Flowchart of the designed decision tree for the classification of nuts through LC-UV fingerprinting, built using PLS-DA models as the rule node. Data matrices dimensions and LVs used to construct the calibration models are detailed.

Concerning the chemometric analysis, Publication V contained the external validation results obtained for three two-input PLS-DA models built using the LC-UV fingerprints in which higher-class eggs were evaluated in front of lower-class ones. Nevertheless, although classification accuracy values for each model were provided, no general classification results were given. In this context, to complement the results shown in Publication V, Table 3.3 presents the external validation values obtained after sample classification following the classification decision tree consisting of 1) organic hen eggs *vs*. others, 2) cage hen eggs *vs*. others, and 3) barn hen eggs *vs*. free range hen eggs. As a result, an overall classification accuracy of 84.3% was obtained, with complete distinction of organic hen eggs and satisfactory results for cage, barn, and free range hen eggs.

Table 3.3. External validation results after carrying out the proposed classification decision tree, based on PLS-DA models, for the assessment of hen eggs production system through LC-UV fingerprinting.

	Sensitivity	**Specificity**
Cage hen eggs	87.5	97.7
Barn hen eggs	78.9	93.8
Free range hen eggs	87.5	85.3
Organic hen eggs	100.0	100.0

Finally, Publications VII, IX, and X analysed paprika samples through non-targeted LC-FLD, FIA–HRMS, and DMS–MS, respectively. In each case, samples were extracted following the procedure previously employed in Publication I, which consisted of an SLE with water:acetonitrile (20:80, *v/v*) as the extracting solvent. Different sample sets were analysed in each publication due to sample availability; therefore, no chemometric comparison between them can be carried out. Nevertheless, regarding the paprika geographical origin, excellent classification results were obtained with each of the proposed fingerprinting methods.

Furthermore, focusing on *La Vera* paprika samples, their complete classification concerning their types —hot, bittersweet, and sweet— had only been reached before

through a non-targeted LC–HRMS method by Barbosa *et al*. [29]. Thus, in this thesis, while for instance low-level data fusion FIA–HRMS fingerprinting did not allow distinguishing between *La Vera* types (Figure 3.2.A), only DMS–MS fingerprinting discriminated them (Figure 3.2.A). Thus, according to the results, separation before MS (either by LC or IMS) is required to distinguish between the three *La Vera* flavour types. This fact may indicate that isomeric compounds could be responsible for these differences.

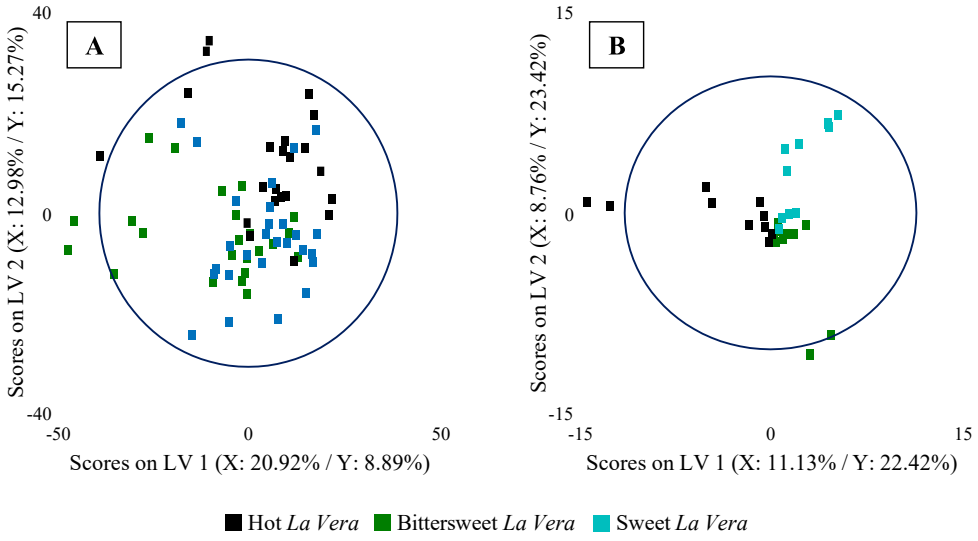


Figure 3.2. PLS-DA model of *La Vera* paprika samples according to their type, obtained through A) low-level data fusion FIA–HRMS fingerprinting or B) DMS–MS fingerprinting.

Therefore, the results presented in this chapter have demonstrated the huge potential of non-targeted chromatographic- and mass spectrometric-based methods to assess different food fraud issues. In particular, three different types of fingerprints have been proposed: 1) chromatographic fingerprints acquired with spectroscopic data, 2) non-targeted LC–HRMS data, and 3) direct MS-based fingerprinting. Each of them presents several advantages. Thus, for instance, chromatographic methods with spectroscopic detection systems are generally less expensive (*i.e*., both in terms of initial purchase and maintenance), more user-friendly, and easier data handling

procedures. Instead, despite its high costs, non-targeted LC–HRMS provides exceptional discriminant ability, mainly due to its high sensitivity and selectivity. Moreover, it can also allow the identification of markers that could be further determined through subsequent targeted analysis. Finally, direct MS-based fingerprinting methods' high throughput and speed are their most significant assets. Nevertheless, the instrument-dependence of non-targeted chromatographic and mass spectrometric methods (*e.g.*, variability inherent to the chromatographic separation and the detection system) is nowadays their main limitation, impeding their application in routine analysis. In this context, research seems to go in this direction in the following years [30,31].

3.4. REFERENCES

[1] L. Cuadros-Rodríguez, C. Ruiz-Samblás, L. Valverde-Som, E. Pérez-Castaño, A. González-Casado, Chromatographic fingerprinting: An innovative approach for food "identitation" and food authentication - A tutorial, Anal. Chim. Acta. 909 (2016) 9–23. https://doi.org/10.1016/j.aca.2015.12.042.

[2] C. Ibáñez, C. Simó, V. García-Cañas, T. Acunha, A. Cifuentes, The role of direct high-resolution mass spectrometry in foodomics, Anal. Bioanal. Chem. 407 (2015) 6275–6287. https://doi.org/10.1007/s00216-015-8812-1.

[3] M. Guerrero-Chanivet, F. Ortega-Gavilán, M.G. Bagur-González, M.J. Valcárcel-Muñoz, M. de V. García-Moreno, D.A. Guillén-Sánchez, Pattern recognition of GC-FID profiles of volatile compounds in Brandy de Jerez Using a chemometric approach based on their instrumental fingerprints, Food Bioprocess Technol. (2023). https://doi.org/10.1007/s11947-023-03039-8.

[4] D.N. Vera, A.M. Jiménez-Carvelo, L. Cuadros-Rodríguez, I. Ruisánchez, M.P. Callao, Authentication of the geographical origin of extra-virgin olive oil of the Arbequina cultivar by chromatographic fingerprinting and chemometrics, Talanta. 203 (2019) 194–202. https://doi.org/10.1016/j.talanta.2019.05.064.

[5] P. Morozzi, A. Zappi, F. Gottardi, M. Locatelli, D. Melucci, A quick and efficient non-targeted screening test for saffron authentication: Application of chemometrics to gas-chromatographic data, Molecules. 24 (2019) 2602. https://doi.org/10.3390/molecules24142602.

[6] S.P. Putri, M.M.M. Ikram, A. Sato, H.A. Dahlan, D. Rahmawati, Y. Ohto, E. Fukusaki, Application of gas chromatography-mass spectrometry-based metabolomics in food science and technology, J. Biosci. Bioeng. 133 (2022) 425–435. https://doi.org/10.1016/j.jbiosc.2022.01.011.

[7] T. Mikrou, M. Litsa, A. Papantoni, M. Kapsokefalou, C. Gardeli, A. Mallouchos, Effect of cultivar and geographical origin on the volatile composition of Greek monovarietal extra virgin olive oils, Chemosensors. 11 (2023) 80. https://doi.org/10.3390/chemosensors11020080.

[8] I. Escriche, A. Conchado, A.M. Peral, M. Juan-Borrás, Volatile markers as a reliable alternative for the correct classification of citrus monofloral honey, Food Res. Int. 168 (2023) 112699. https://doi.org/10.1016/j.foodres.2023.112699.

[9] S. Reale, A. Biancolillo, M. Foschi, F. Di Donato, E. Di Censo, A.A. D'Archivio, Geographical discrimination of Italian carrot (Daucus carota L.) varieties: A comparison between ATR FT-IR fingerprinting and HS-SPME/GC-MS volatile profiling, Food Control. 146 (2023) 109508. https://doi.org/10.1016/j.foodcont.2022.109508.

[10] Y. Yang, J. Li, J. Xing, W. Xing, C. Tang, Z. Rao, J. Zhang, Untargeted profiling and differentiation of volatiles in varieties of meat using GC Orbitrap MS, Foods. 11 (2022) 3997. https://doi.org/10.3390/foods11243997.

[11] A. Morabito, G. De Simone, M. Ferrario, F. Falcetta, R. Pastorelli, L. Brunelli, EASY-FIA: A readably usable standalone tool for high-resolution mass spectrometry metabolomics data pre-processing, Metabolites. 13 (2023) 13. https://doi.org/10.3390/metabo13010013.

[12] Z. Liu, M. Zhang, P. Chen, J.M. Harnly, J. Sun, Mass spectrometry-based nontargeted and targeted analytical approaches in fingerprinting and metabolomics of food and agricultural research, J. Agric. Food Chem. 70 (2022) 11138–11153. https://doi.org/10.1021/acs.jafc.2c01878.

[13] R. Gamboa-Becerra, J.M. Montero-Vargas, S. Martínez-Jarquín, E. Gálvez-Ponce, A. Moreno-Pedraza, R. Winkler, Rapid classification of coffee products by data mining models from direct electrospray and plasma-based mass spectrometry analyses, Food Anal. Methods. 10 (2017) 1359–1368. https://doi.org/10.1007/s12161-016-0696-y.

[14] M.J. Lerma-García, Y. Saucedo-Hernández, J.M. Herrero-Martínez, G. Ramis-Ramos, E. Jorge-Rodríguez, E.F. Simó-Alfonso, Statistical classification of pumpkin seed oils by direct infusion mass spectrometry: Correlation with GC-FID profiles, Eur. J. Lipid Sci. Technol. 117 (2015) 331–337. https://doi.org/10.1002/ejlt.201400250.

[15] P. Chen, J. Sun, P. Ford, Differentiation of the four major species of cinnamons (C. burmannii, C. verum, C. cassia, and C. loureiroi) using a flow injection mass spectrometric (FIMS) fingerprinting method, J. Agric. Food Chem. 62 (2014) 2516–2521. https://doi.org/10.1021/jf405580c.

[16] M. Vilà, À. Bedmar, J. Saurina, O. Núñez, S. Sentellas, High-throughput flow injection analysis–mass spectrometry (FIA-MS) fingerprinting for the authentication of tea application to the detection of teas adulterated with chicory, Foods. 11 (2022) 2153. https://doi.org/10.3390/foods11142153.

[17] V. Hrbek, M. Rektorisova, H. Chmelarova, J. Ovesna, J. Hajslova, Authenticity assessment of garlic using a metabolomic approach based on high resolution mass spectrometry, J. Food Compos. Anal. 67 (2018) 19–28. https://doi.org/10.1016/j.jfca.2017.12.020.

[18] T. Nikou, M. Witt, P. Stathopoulos, A. Barsch, M. Halabalaki, Olive oil quality and authenticity assessment aspects employing FIA-MRMS and LC-Orbitrap MS metabolomic approaches, Front. Public Heal. 8 (2020) 558226. https://doi.org/10.3389/fpubh.2020.558226.

[19] M. Hernández-Mesa, D. Ropartz, A.M. García-Campaña, H. Rogniaux, G. Dervilly-Pinel, B. Le Bizec, Ion mobility spectrometry in food analysis: Principles, current applications and future trends, Molecules. 24 (2019) 2706. https://doi.org/10.3390/molecules24152706.

[20] E. te Brinke, A. Arrizabalaga-Larrañaga, M.H. Blokland, Insights of ion mobility spectrometry and its application on food safety and authenticity: A review, Anal. Chim. Acta. 1222 (2022) 340039. https://doi.org/10.1016/j.aca.2022.340039.

[21] K. Masike, A. de Villiers, D. de Beer, E. Joubert, M.A. Stander, Application of direct injection-ion mobility spectrometry-mass spectrometry (DI-IMS-MS) for the analysis

of phenolics in honeybush and rooibos tea samples, J. Food Compos. Anal. 106 (2022) 104308. https://doi.org/10.1016/j.jfca.2021.104308.

[22] J. Sun, M. Zhang, N. Kubzdela, Y. Luo, J.M. Harnly, P. Chen, Determination of variance of secondary metabolites in lettuces grown under different light sources by flow injection mass spectrometric (FIMS) fingerprinting and ANOVA–PCA, J. Anal. Test. 2 (2018) 312–321. https://doi.org/10.1007/s41664-018-0072-6.

[23] W. Liu, Y. Zhao, J. Sun, G. Li, Y. Shan, P. Chen, Study the effects of drying processes on chemical compositions in daylily flowers using flow injection mass spectrometric fingerprinting method and chemometrics, Food Res. Int. 102 (2017) 493–503. https://doi.org/10.1016/j.foodres.2017.09.015.

[24] A.N. de Oliveira, S.R.F. Bolognini, L.C. Navarro, J. Delafiori, G.M. Sales, D.N. de Oliveira, R.R. Catharino, Tomato classification using mass spectrometry-machine learning technique: A food safety-enhancing platform, Food Chem. 398 (2023) 133870. https://doi.org/10.1016/j.foodchem.2022.133870.

[25] P. Geng, J. Sun, P. Chen, E. Brand, J. Frame, H. Meissner, J. Stewart, S. Gafner, S. Clark, J. Miller, J. Harnly, Characterization of maca (Lepidium meyenii/Lepidium peruvianum) using a mass spectral fingerprinting, metabolomic analysis, and genetic sequencing approach, Planta Med. 86 (2020) 674–685. https://doi.org/10.1055/a-1161-0372.

[26] W. Lu, Q. Jiang, H. Shi, Y. Niu, B. Gao, L. Yu, Partial least-squares-discriminant analysis differentiating chinese wolfberries by UPLC-MS and flow injection mass spectrometric (FIMS) fingerprints, J. Agric. Food Chem. 62 (2014) 9073–9080. https://doi.org/10.1021/jf502156n.

[27] C. Proestos, M. Pesic, Editorial: Polyphenols in food authenticity through the application of advanced omics technologies, Front. Nutr. 46 (2022) 10–11. https://doi.org/10.3389/fnut.2022.998635.

[28] P. Lucci, J. Saurina, O. Núñez, Trends in LC-MS and LC-HRMS analysis and characterization of polyphenols in food, TrAC - Trends Anal. Chem. 88 (2017) 1–24. https://doi.org/10.1016/j.trac.2016.12.006.

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

[30] L. Cuadros-Rodríguez, S. Martín-Torres, F. Ortega-Gavilán, A.M. Jiménez-Carvelo, R. López-Ruiz, A. Garrido-Frenich, M.G. Bagur-González, A. González-Casado, Standardization of chromatographic signals – Part II: Expanding instrument-agnostic fingerprints to reverse phase liquid chromatography, J. Chromatogr. A. 1641 (2021) 461973. https://doi.org/10.1016/j.chroma.2021.461973.

[31] M. Creydt, M. Fischer, Mass-spectrometry-based food metabolomics in routine applications: A basic standardization approach using housekeeping metabolites for the authentication of Asparagus, J. Agric. Food Chem. 68 (2020) 14343–14352. https://doi.org/10.1021/acs.jafc.0c01204.

CHAPTER 4

CONCLUSIONS

In the present thesis, several food fraud cases, including different food products (paprika, nuts and seeds, hen eggs, vegetable oils, and red wine) and issues (assessment of the botanical origin, geographical origin, production system, and product quality), have been evaluated through metabolomic profiling and fingerprinting approaches.

In this context, the following conclusions can be drawn from the experimental studies presented.

Regarding the metabolomic profiling approaches:

A total of 36 phenolic compounds belonging to different classes were determined by LC–MS/MS in paprika samples.

- Among the targeted compounds, 20 were detected and semi-quantified in the 111 paprika samples under study (16 being present in all of them).
- The phenolic profile allowed excellent sample classification according to the corresponding geographical origin through PLS-DA.

Paprika samples were analysed by LC−HRMS, using APCI as the ionisation source, to determine their capsaicinoid and carotenoid profile.

- After the semi-quantification of the targeted compounds, it was observed that the capsaicinoid content was strongly associated with the flavour paprika type. In contrast, carotenoid content could be related to the country of origin.
- A PLS-DA classification decision tree assessed the samples' geographical origin and flavour type, reaching an overall classification accuracy of 80.9% after external validation.

According to the obtained results, the phenolic profile proved to be a better chemical descriptor than the capsaicinoid and carotenoid one to address the paprika geographical origin.

Concerning the metabolomic fingerprinting approaches based on liquid chromatography with spectroscopic detection:

Nut and seed samples of ten different types were analysed through non-targeted LC-UV and LC-FLD methods.

- Although the LC-UV fingerprints (280 nm as the excitation wavelength) were richer in peak features than their LC-FLD counterparts (280 and 350 nm as the excitation and emission wavelengths), better descriptive performance was obtained in the latter case, which may be due to its more selective detection of highly relevant chemical descriptors.
- Using LC-UV fingerprinting, a classification decision tree was built to optimise sample PLS-DA classification. As a result, 95% of the samples were correctly classified after external validation.
- Instead, after external validation, a single multi-class PLS-DA model, built using non-targeted LC-FLD data, reached a classification accuracy of 98.3%.
- Moreover, LC-FLD fingerprinting was also successfully applied to detect and quantitate the adulteration of almond-based products with hazelnut or peanut by PLS.

The geographical origin of paprika samples was addressed using the acquired LC-FLD fingerprints.

- LC-FLD fingerprints were subjected to a PLS-DA classification decision tree for geographical origin classification. As a result, an accuracy of 97.9% was reached after external validation.
- Detection and quantitation of different paprika geographical blend scenarios by PLS provided external validation and prediction errors below 1.6% and 10.7%, respectively.

LC-UV fingerprinting was proposed to assess the classification of hen eggs according to their production system (*i.e.*, organic, free range, barn, and cage hen eggs).

- LC-UV fingerprints, acquired using 250 nm as the excitation wavelength, proved their excellent ability to effectively classify organic hen eggs from lower priced ones by PLS-DA completely. Moreover, an overall classification rate of 84.3% was obtained through a classification decision tree and after external validation.

About the metabolomic fingerprinting approaches based on LC–HRMS:

A non-targeted LC–HRMS method was used to classify nut and seed samples and tentatively identify the discriminant molecular features.

- The discriminant ability of the obtained LC–HRMS fingerprints was proved by achieving the complete sample classification through a multi-class PLS-DA model.
- A total of 136 molecular features —corresponding to the chemical differences between nut types according to the VIP scores— were tentatively identified and annotated.

A targeted LC–HRMS method, focusing on some of the previously found markers, was used to detect and quantitate almond-based products adulteration with hazelnut or peanut.

- PLS allowed the prediction of blend percentages with errors below 10%, proving the validity of the proposed chemical descriptors.

Regarding the metabolomic fingerprinting approaches based on direct MS-based techniques:

The authentication and classification of different food products were assessed by means of FIA–HRMS.

- A fast and high throughput method was proposed to analyse several food matrices, including red wine, paprika, and vegetable oils.
- Different data matrices —negative ionisation FIA–HRMS fingerprints, positive ionisation FIA–HRMS fingerprints, low-level FIA–HRMS fingerprints, and mid-level FIA–HRMS fingerprints— were evaluated.
- While no clear characteristic patterns due to geographical features were observed for red wine samples in the chemometric study, an excellent distinction was observed for the other food products analysed. Thus, compositional FIA–HRMS fingerprints of paprika differed considerably according to their geographical origin, while vegetable oils did it according to their botanical origin and, particularly, olive oils to their quality.

Non-targeted DMS–MS was evaluated as a high throughput method for assessing paprika's geographical origin.

- DMS–MS was applied satisfactorily for the first time to a food classification issue following a fingerprinting approach.
- The complete classification of paprika samples belonging to the three Spanish regions distinguished with the PDO label was reached through PLS-DA.
- DMS-MS fingerprints also distinguished the flavour types within a specific region.

Finally, overall, it can be concluded that both metabolomic profiling and fingerprinting approaches, through the analytical techniques herein proposed, have great potential to address a diverse set of food fraud issues (*e.g.*, mislabelling related to the species origin, the geographical origin, the production system, or the product quality). Nevertheless, further research about validating chemometric models and the inter-laboratory transference of fingerprinting methods is still required to establish standardised guidelines for their use in routine analysis.