



**HAL**  
open science

## Untargeted metabolomic approaches in food authenticity: a review that showcases biomarkers

Nicolas Mialon, Benoit Roig, Eric Capodanno, Axelle Cadiere

### ► To cite this version:

Nicolas Mialon, Benoit Roig, Eric Capodanno, Axelle Cadiere. Untargeted metabolomic approaches in food authenticity: a review that showcases biomarkers. *Food Chemistry*, 2023, pp.133856. 10.1016/j.foodchem.2022.133856 . hal-03749096

**HAL Id: hal-03749096**

**<https://hal.science/hal-03749096>**

Submitted on 27 Oct 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Review

### Untargeted metabolomic approaches in food authenticity: A review that showcases biomarkers

**Nicolas Mialon (a, b)**

**Benopit Roig (a)**

**Eric Capodanno (b)**

**Axelle Cadière (a)**

Univ. Nimes, UPR CHROME, Rue du Dr G. Salan, 30021 Nimes Cedex 1, France

Phytocontrol, 180 Rue Philippe Maupas, 30035 Nîmes Cedex 1, France

<https://doi.org/10.1016/j.foodchem.2022.133856>

## Highlights

Non-targeted analysis for [food authentication](#) and its evolution.

Biomarkers across matrixes and their potential overlaps.

Statistical methods and corresponding accuracies depending on the fraud.

Terroir definition to assess origin and quality.

Workflow for determining geographic origin without pre-determined outcomes.

## Abstract

The assessment of food authenticity is a topic that gained a lot of interest at the international level. This term includes misidentification of variety, origin, production system, processing but also adulteration. These frauds all have an analytical component, and research tends to offer new analytical solutions to manage them. One of them is non-targeted approaches, which get around the limitations of targeted analysis by detecting the unexpected. A wide range of products are studied such as wine, rice, olive oil, spices, and honey among the top five. Geographic origin is by far the fraud with the most attention. The main reason is probably the complexity to consider terroir effect and every other variable to determine an area of production. This review offers an overview of the potential of non-targeted analysis to assess food authenticity. These results also illustrate the capability to look for environmental terroir markers that could be cross-matrixes.

## Graphical abstract



## Abbreviations

ACO

Ant Colony Optimization

ANN

Artificial Neural Network

ANOVA

Analysis of Variance

ANRT

Association Nationale pour la Recherche Technologique

CEN

Comité Européen de Normalisation

CRM

Certified Raw Material

DD-SIMCA

Data Driven - Soft Independent Modelling by Class Analogy

DI-ESI

Direct Introduction - Electrospray

EMA

Economically Motivated Adulteration

e-nose

electronic-nose

e-tongue

electronic-tongue

EU

European Union

FTIR

Fourier-Transform Infrared

GC

Gas Chromatography

GI

Geographic Indication

HCA

Hierarchical Cluster Analysis

HRMS

High Resolution Mass Spectrometry

IOOC

International Olive Oil Council

k-NN  
k-Nearest Neighbors  
LC  
Liquid Chromatography  
LDA  
Linear Discriminant Analysis  
MIR  
Mid Infrared  
MLP  
Multi-Layer Perceptron  
MS  
Mass Spectrometry  
NGS  
Next Generation Sequencing  
NIR  
Near Infrared  
NMR  
Nuclear Magnetic Resonance  
OIV  
International Organization of Vine and Wine  
OPLS-DA  
Orthogonal Projection to Latent Structure - Discriminant Analysis  
PCA  
Principal Component Analysis  
PCR  
Polymerase Chain Reaction  
PDO  
Protection Designation of Origin  
PLSC-DA  
Projection to Latent Structure Correlation - Discriminant Analysis  
PLS-DA  
Projection to Latent Structure - Discriminant Analysis  
PTR  
Proton Transfer Reaction  
QC  
Quality Control  
QTOF  
Quadrupole - Time of Flight  
RF  
Random Forest  
USP  
United States Pharmacopeia

## **Keywords**

Food Fraud  
Non-targeted approaches  
Metabolomics  
Authenticity  
Adulteration

## 1. Introduction

Food authenticity is a concept at the essence of the contract between consumers and producers. It is closely related to food quality, and sometimes safety when it comes to adulteration. Still, acts of deceiving consumers are common since antiquity and measures were taken. Laws were pronounced against adulteration in both Rome and Greece. These laws became transnational in the thirteenth century ([Sumar & Ismail, 1995](#)).

However, new types of food fraud emerged with globalization. EU estimated the cost of food fraud at 30 to 40 billion euros every year, and number of alerts almost doubled between 2016 and 2019 ([European Commission. Directorate General for Health and Food Safety., 2020](#)). The CEN working agreement 17369:2019 defines food fraud as the act of “intentionally causing a mismatch between food product claims and food product characteristics” ([CWA 17369: Authentic and Fraud in the Feed and Food Chain- Concepts, Terms and Definitions, 2019](#)). That implies that fraud can be committed in two ways: either modifying food product characteristics (that would be product tampering), or food product claims (that would be record tampering or implicit claim violation).

Some of the records tampering can only be verified through traceability checks like batch numbers, ethical and sustainability practices, counterfeits, or selling products obtained from grey market. This part of food fraud is in the hand of regulation authorities and has no physicochemical components. However, product tampering and a few record tampering can be analytically verified ([Robson et al., 2021](#)) such as qualities, vintages, ripening times, farming practices, geographic origins, species identification, and some processes applied on the products (e.g., irradiation, “never frozen” for salmon or “not heated” for honey).

Both targeted and untargeted (or non-targeted) approaches are used for these purposes. Targeted techniques are used for decades and are very effective to identify and quantify known markers of a specific fraud such as triacyl [glycerides](#) (TAG) to assess oil variety ([Indelicato et al., 2017](#)). They are also the only methods available to detect the chemical origin of compounds with isotopic measurement (e.g., if they are issued from organic synthesis, biotechnology, or natural reaction). However, unknown, or unexpected markers are unavoidably missed because they are not part of databases. This limitation conducted to the [melamine](#) scandal in milks in 2008, which was not detected by targeted techniques ([Cavin et al., 2016](#)). Untargeted approaches in the food analytical field were developed from this observation ([Supplementary figure](#)). This field have been reviewed before with different interests: [Cubero-Leon et al., 2014](#) is the first one and retrieved mainly articles about adulteration where spectroscopic techniques were overrepresented. They also noticed that PCA was the most used algorithms. [Cuadros-Rodríguez et al., 2016](#) followed and specifically reviewed in chromatographic techniques on specific fractions or extracts. Finally, [Medina et al., 2019](#) is the first to talk about biomarkers to discriminate across food frauds. In each of these reviews, the term “fingerprint” is used for different purposes including a specific class of molecules, a specific range of a chromatogram or “molecular markers that represent a characteristic food state or condition” ([Medina et al., 2019](#)). This is a good illustration of how various terms in the literature are used to characterize untargeted workflows and how often there is confusion and inconsistency between them. Therefore, there are numerous keywords that were looked for, such as: fingerprint, profile, signature, untargeted, non-targeted, multi-

targeted, and chemistry-related omics workflows. Then, the article selection of this review was based on three criteria. First, our review is interested in differentiating between the different terms used, namely between profiling and non-targeting, where the term “profiling” is restricted to the study of one class of compounds (e.g., phenols, fatty acids, etc.) whereas the non-targeting strategy seems to be the most innovative. It is in this perspective of making the difference between the approaches and showing the need to develop the cross-targeted and non-targeted approaches that the review is written.

The purpose of these studies was to use untargeted methods to classify authentic samples or find promising authenticity markers. Many platforms both analytically and statistically were used to tackle specific authenticity issues, meaning geographic origin, specie identification, quality, farming practice or adulteration. In the end, models are evaluated regarding the assignation of samples to a specific group. Therefore, the identification of the markers is not mandatory or sometimes possible, depending on the technique. However, this identification and the clarification of biological roles of such molecules reinforce trust in the conclusion or even lead to a targeted approach. A good example is the identification of 16-O-Methylcafestol that demonstrated to be an adequate biomarker for distinguishing between Robusta and [Arabica coffee](#). ([Gunning et al., 2018](#)).

## 2. Untargeted metabolomics workflow for food authentication

Non-targeted approaches are far from being a standardized process. However, general guidelines were edited by USP. They state three steps that have been reviewed in each article: sampling, classification, and evaluation of the model. ([USP Guidance on Developing and Validating Non-Targeted Methods for Adulteration Detection, 2019](#)).

### 2.1. Experimental design and sampling

Untargeted experiments are mainly carried out in two ways: database filtering or global approaches ([Shao et al., 2019](#)). Database filtering often implies that components not in the database are not considered. Due to this bias, results are partly forecasted, and this approach should be considered as pseudo-targeted. Global workflows open the most possibilities and results are only based on statistical analysis; this is the only case considered as non-targeted analysis further in this review. The experimental plan needs to consider the variability needed to evaluate each modality. The number of samples for each modality (e.g., origin, species, quality...) should be enough to reflect natural variations like seasoning, terroirs, processes variations, etc. Recommendations for sample size range from 60 to 200 samples for each modality ([McGrath et al., 2018](#), [Uhlig et al., 2021](#)). Once this design step is completed, the first step in analysis can begin with sampling.

Collection and processing of the samples are crucial steps to obtain reliable results. As discussed by USP and Food Integrity group ([USP Guidance on Developing and Validating Non-Targeted Methods for Adulteration Detection, 2019](#), [McGrath et al., 2018](#)), adequate sampling is the first requirement to consider in the development process. To develop the model, a calibration set of typical samples should be obtained from reliable sources. For its evaluation, a validation set of comparable size, containing both typical and atypical samples, should be employed (Graphical Abstract). To confirm the authenticity of the sample, a recommended approach is to go through PDOs management organisms, applied research

institutes, federal control labs, or directly from producers. Acquiring and certifying all the metadata for each sample takes the utmost efforts, but it cannot be overlooked as the quality of the model highly depends on it. Regarding sample processing, it is recommended to keep it minimal ([Rubert et al., 2014](#)), as any sample manipulation can cause chemical changes, elimination of some compounds or contamination of the sample. This is a reason why techniques where samples can be directly subjected to analysis are especially interesting in untargeted experiments ([Beneito-Cambra et al., 2020](#)). However, some preparation is often needed prior to analysis which can include dilution, extraction, or [purification](#), depending on the analytical technique.

## 2.2. Techniques and data acquisition

The analytical platform that is used to perform the analyses is an important choice ([Supplementary table](#)). Three main systems can be distinguished: sensory based techniques like electronic noses and tongues (e-nose / e-tongue) ([Gliszczyńska-Świgło & Chmielewski, 2017](#)), spectroscopy including nuclear magnetic resonance (NMR) and Fourier-transform infrared (FTIR) with near and mid infrared (NIR and MIR) and finally, spectrometry hyphenated with gas or [liquid chromatography](#) (GC-MS/LC-MS). Some other detectors can be hyphenated with these separation methods as well. Each of these systems comes with its own benefits and drawbacks ([Ellis et al., 2012](#)). Chromatographic techniques offer high sensitivity and separation. GC-MS is the most reproducible and wide range of databases are available for identification, the downside is the limitation to [volatile compounds](#) and the need of [derivatization](#) to be expanded. LC-MS has better metabolite coverage, but preparation can be time consuming and induce robustness issues. Identification is also a challenging point. NMR and FTIR both offer detection of the whole sample with no separation and poor sensitivity. NMR gives reliable identification of compounds with a high analysis cost ([Shao et al., 2019](#)). On the other side, FTIR is an inexpensive alternative, but compound identification in a mixture is nearly impossible. Finally, e-sensors offer separation, sensitivity, and whole sample analysis, but identification cannot be carried out in a standalone method.

Analytical validation for each of them is technique-dependent, but guidelines for targeted methods can be applied to verify robustness, accuracy, reproducibility, etc. An interesting point that is shared by a lot of studies, is the use of pooled QC samples ([Mie et al., 2014](#), [Putri et al., 2019](#), [Zhao et al., 2017](#)) that provides information on the state of the system over the batches. They can also be used during the post processing to ensure the quality of the model.

## 2.3. Data treatment and result processing

As these results can be composed of thousands of features and hundreds of samples, data filtering is generally needed before applying [chemometric](#) analysis. For data filtering, main techniques are deconvolution or peak picking, peak or spectra alignment, denoising, normalization, bucketing, etc. Many workflows and tools are available such as Workflow4Metabolomics, XCMS, MetAlign, Mzmine 2, and other proprietary solutions ([MetAlign, n.d.](#), [Mzmine 2., n.d.](#), [Workflow4Metabolomics, n.d.](#), [XCMS, n.d.](#)).

Once the data is suitable for statistical analysis, [data compression](#) is often performed. There are two types of learning commonly used: supervised and unsupervised. The former optimizes the function based on the attended outcomes, whilst the latter is done without prior information and attempts to guess the data structure. Among these, one algorithm is the most frequent: principal component analysis (PCA) that is unsupervised and provides effective data



reduction. It was used in 83 % of the retrieved articles ([Table 2](#), [Table 3](#), [Table 4](#), [Table 5](#)). Linear discriminant analysis (LDA) is a supervised alternative but can also be used after PCA to perform classification ([Vestner, 2016](#)). Data reduction enables the selection of the most relevant features and observations in a large amount of data. These selected features are often verified using significance test or analysis of variance (ANOVA), and k-nearest-neighbors (k-NN) can provide this information for each feature and detect outliers ([Abbatangelo & Sberveglieri, 2019](#)). Another statistical approach is to use [artificial neural networks](#) (ANN) ([Gumus et al., 2018](#)) or hierarchical cluster analysis (HCA) to similarities between variables across different clusters. Finally, once selected data has been validated for significance and confidence, it can be used to build the model. Unknown samples can be compared to the model and classified as authentic or not in the case of univariate models, or as belonging to one category or none in the case of multivariate ones ([Alonso et al., 2015](#)). Algorithms to classify samples are mostly supervised and overfit quickly. As a result, cross validation is essential to ensure the model's adequacy ([McGrath et al., 2018](#)). Random forest (RF) is used for both classification and regression and can be helpful to create decision trees that can assign subsamples to each class. It is suitable for both one class and multiclass modelling. Recent improvements coupled RF with swarm intelligence such as ant colony optimization (ACO) to improve accuracy ([Kalogiouri et al., 2018](#)).

Others include soft independent modelling by class analogy (SIMCA) and projection to latent structure (PLS). SIMCA can be data-driven (DD) and is a one-class method. PLS-DA can be used for multi-class modelling as well as data compression. Variations of the method exist with orthogonal projection to latent structure (OPLS-DA) or projection to latent structure correlation (PLSC-DA). All these variations have been grouped as PLS-DA further in the review as they do not affect the predictivity of the model ([Trygg & Wold, 2002](#)).

It should be noted that due to the large amount of data generated and the diversity of untargeted approaches, machine learning is a promising tool to use. It is already commonly used in [metabolomics](#) ([Liebal et al., 2020](#)) but it stays a bit more confidential in food authenticity. As reviewed recently by [Liang et al., 2022](#), many types of neural network structures are used in the context of food authenticity. In most cases retrieved by Liang et al., these techniques improve results of classification. However, it is important to point out that hidden layers in neural networks have a significant disadvantage: determining the amount of discrimination for each molecule is challenging. As a result, either explainable AI must be used to rationalize the decision-making process, or other algorithms must be applied to gather this information.

Once the model is operable, the test sample set is used to evaluate its performance. USP mentioned a way to evaluate performances of chemometrics models with sensitivity and specificity ([USP Guidance on Developing and Validating Non-Targeted Methods for Adulteration Detection, 2019](#)). The sensitivity rate is “the number of correct atypical predictions from the method divided by the total number of atypical samples” while the specificity rate is “the number of correct typical predictions from the method divided by the total number of typical samples.” Despite this guideline, most studies use different indicators such as prediction accuracy and classification capability that are defined as the number of correct assignments (authentic or not), divided by the size of the sample ([Cajka et al., 2011](#); [Gumus et al., 2020](#)). The counterpart of this indicator, the error rate, is also found in literature. Once the model performances are evaluated and acceptable, new samples can be assigned using it. However, another step is often carried out: most important loadings (e.g. compounds) identification.



## 2.4. Biomarker identification

In non-targeted analysis, discriminatory molecules are referred as biomarkers. This denomination is relevant when they occurred naturally but might not be suitable when it comes to adulteration. Their identification is often the final step in the development of a new method and is technically optional. A long list of databases and tools can be used to help on compound identification ([Table 1](#)).

Table 1. Available databases for compound identification in [chromatography](#) and NMR.

Databases	Techniques	Spectral collection	Number of entries	Website
HMDB	NMR MS	Experimental + In Silico	> 115 000	<a href="https://hmdb.ca/">https://hmdb.ca/</a>
Kegg	–	–	> 15 000	<a href="https://www.genome.jp/kegg/">https://www.genome.jp/kegg/</a>
FoodB	–	–	> 70 000	<a href="https://foodb.ca/">https://foodb.ca/</a>
MetaCyc	–	–	> 17 000	<a href="https://metacyc.org/">https://metacyc.org/</a>
ChEBI	–	–	> 59 000	<a href="https://www.ebi.ac.uk/chebi/init.do">https://www.ebi.ac.uk/chebi/init.do</a>
Coconut	–	–	> 400 000	<a href="https://coconut.naturalproducts.net/">https://coconut.naturalproducts.net/</a>
MassBank	LC-MS GC-MS	Experimental	> 86 000	<a href="https://www.massbank.jp/">https://www.massbank.jp/</a>
Metfrag	LC-MS	In Silico	> 10 000 000	<a href="https://msbi.ipb-halle.de/MetFrag/">https://msbi.ipb-halle.de/MetFrag/</a>
MoNA	LC-MS	Experimental + In Silico	> 200 000	<a href="https://mona.fiehnlab.ucdavis.edu/">https://mona.fiehnlab.ucdavis.edu/</a>
GNPS	LC-MS	Experimental + In Silico	> 10 000	<a href="https://gnps.ucsd.edu">https://gnps.ucsd.edu</a>
Wiley	GC-MS	Experimental + In Silico	> 650 000	<a href="https://sciencesolutions.wiley.com">https://sciencesolutions.wiley.com</a>
NIST	GC-MS LC-MS	Experimental	> 100 000	<a href="https://chemdata.nist.gov/">https://chemdata.nist.gov/</a>
GMD	GC-MS	Experimental	> 25 000	<a href="https://gmd.mpimp-golm.mpg.de/">https://gmd.mpimp-golm.mpg.de/</a>
BMRB	NMR	Experimental	> 6 000	<a href="https://bigg.ucsd.edu/">https://bigg.ucsd.edu/</a>
BML NMR	NMR NMR	Experimental	208	<a href="https://www.bml_nmr.org/">https://www.bml_nmr.org/</a>
Metabolight	MS	Experimental	> 27 000	<a href="https://www.ebi.ac.uk/metabolights/">https://www.ebi.ac.uk/metabolights/</a>
LIPID MAPS	MS	Experimental + In Silico	> 45 000	<a href="https://www.lipidmaps.org/">https://www.lipidmaps.org/</a>
Metlin	MS	Experimental + In Silico	> 900 000	<a href="https://metlin.scripps.edu/">https://metlin.scripps.edu/</a>

Databases can be built using only chemical structures and names or include mass and [NMR spectra](#) as well. These data were mainly experimental, retrieved from publications, but the evolution of computational tools now allows *in silico* prediction of both new compounds or physicochemical properties that can increase the number of compounds available or the

trustworthiness. LC-HRMS workflows generally query HMDB, Kegg and Massbank databases because they are the most developed and easy to interface. For GC-MS, the NIST and Wiley are the oldest and most spread databases, the NIST is using experimental Kovats index, while the Wiley's are computed. Databases are not mandatory for NMR, but identification might be time consuming and challenging in the event of overlapping signals in complex mixtures. As mentioned above, identification of biomarkers is not mandatory to assign classes and authenticate foodstuffs, but it is relevant for official validation and method transfer into production.

## 2.5. Food authentication

The purpose of these analyses in the foodstuff field is to guarantee that the product is what it claims it is. Authentication of food requires specific aspects to be considered for each type of fraud. The UK [Food Crime Strategic Assessment, 2020](#) ([Food Crime Strategic Assessment, 2020](#)) stated that diversion of waste product, misrepresentation of premium status or provenance, adulteration and unlawful processing are the most important issues analytically verifiable. Products most susceptible to fraud are a constant top 10 including olive oil, wine, honey, meat, fruits, and vegetables. This report also shows that the most important fraud is mislabeling, counting for 47 % of the alerts, followed by adulteration (20 %) and unapproved treatment or process (16 %). Mislabeling includes products wrongfully labelled as organic, from premium quality (e.g., extra virgin olive oil instead of lampant or virgin), or from an incorrect origin. Adulteration is the act of adding, removing, substituting, or diluting a product. For instance, water can be added to fish, honey can be substituted by sugars and pollens, and colorants can be added to spices. Finally, unapproved treatments or processes are the use of unauthorized practices. For example, in the EU, it is forbidden to use ethylene oxide or radiation to sterilize product or [carbon monoxide](#) to enhance tuna's color.

This review was organized based on the type of fraud: misdescription was separated between organic or conventional farming distinction, premium quality usurpation, species identification and geographic origin. All forms of adulteration were considered together, but literature shows that the focus is on melamine-like adulteration; the use of exogen substances because they might be life-threatening. However, real-life scenario is closer to the dilution of a wine with a cheaper one instead of just adding one substance.

## 3. Food frauds and ways to identify them

Non-targeted analysis being a new area of analytical chemistry, >80 % of the retrieved articles came out within the last five years. Most represented techniques are MS-based with >75 % of the mentions. Many data were extracted from the research such as the statistical tools, the classification abilities, and the identified chemical family of biomarkers. The oldest food fraud studied by chemical methods was about adulteration, 200 years ago ([Accum, 1820](#)).

### 3.1. Adulteration

Adulteration is defined as the act of adding, substituting, diluting, or removing an ingredient from a product ([Jha, 2016](#)). It was firstly studied by targeted methods, until the melamine scandal showed that control plans were not foolproof. Since, untargeted methods started to be developed. Different kind of approaches exist with both spectroscopic and chromatographic methodologies ([Table 2](#)). The idea is to apply forensic techniques to fraud. Instead of keeping

a list of what fraudsters might use and check for them, the point is to be prepared and look for everything. If unexpected contaminants are found, further investigation is needed ([Primrose et al., 2010](#)). The main difficulty is the evaluation of the method. Because you must choose [adulterants](#) to create atypical samples, or screen enough commercial fraudulent samples to evaluate the accuracy of the method.

Table 2. Summary of screened matrixes for [food adulteration](#). Accuracy, threshold, and biomarkers are shown when available.

Matrix (Number of sample)	Analytical technique	Statistical tools	Accuracy	Adulteration threshold	Biomarker	Reference
TCM <sup>a</sup> (1800) <sup>b</sup>	FT-NIR	PLS-DA	93 %	2 to 5 %	–	( <a href="#">Fu et al., 2017</a> )
Palm oil (30) <sup>b</sup>	NMR	PCA	–	–	Triglycerides Phytosterols Fatty acids	( <a href="#">Perini et al., 2018</a> )
Cheese (84) <sup>b</sup>	UPLC- Orbitrap	PCA	87,5%	–	–	( <a href="#">Popping et al., 2017</a> )
Orange juice (120) <sup>b</sup>	DART- QTOF	PCA + PLS- DA	95 %	–	Tripeptides Amino acids Fatty acids	( <a href="#">Xu, Xu, Kelly, et al., 2020</a> )
Saffron (20) <sup>b</sup>	UPLC- QTOF	PCA + PLS- DA	100 %	1 %	Flavonoids Organic acids Carbohydrates	( <a href="#">Guijarro-Díez et al., 2015</a> )
Fruits juices (35) <sup>c</sup>	UPLC- QTOF	PCA + PLS- DA	–	1 %	Flavonoids Carbohydrates Amino acids	( <a href="#">Jandrić et al., 2014</a> )
Fruits extracts (86) <sup>c</sup>	HPLC-UV	PCA + PLS- DA	>97 %	2 %	–	( <a href="#">Pardo-Mates et al., 2017</a> )
Olive oil (9) <sup>c</sup>	DI-ESI-MS	PCA	–	–	–	( <a href="#">Goodacre et al., 2002</a> )
Fruits juices (96) <sup>c</sup>	UPLC- QTOF	PCA + PLS- DA	–	10 %	Flavonoids Organic acids Carbohydrates	( <a href="#">Zhang et al., 2018</a> )
Violet leaf extracts <sup>c</sup>	UPLC- QTOF	PCA	–	10 %	–	( <a href="#">Saint-Lary et al., 2016</a> )
Fruits juices (216) <sup>c</sup>	UPLC- QTOF	PLS- DA + DD- SIMCA	100 %	1 %	Amino acids Flavonoids Sphingolipids	( <a href="#">Xu, Xu, Strashnov, et al., 2020</a> )

a

Traditional Chinese Medicine.

b

Real matrixes.

c

Lab-made matrixes.

These studies show that fruit juices are the most studied matrix for adulteration. This is not surprising because adulteration is easier to perform on liquids. Adulteration is often referred as economically motivated adulteration (EMA), meaning that profit is the main motivation and therefore dependent of the product volumes. As a result, products with a large manufacturing volume or a high selling cost are the most targeted by fraudsters ([Shears, 2010](#)).

Accuracy, as shown in the table, is most likely not the greatest indicator to use in these cases. Instead, some articles put forward the detection threshold for typical adulterants. It is interesting because the economic motivation has a limit. If adulteration of extra-virgin oil is performed with 20 % of lampant olive oil, margins are increased by 19 %, making huge profits. If it is done with 1 % or 0.5 %, margins are so thin that adulteration is not economically worth anymore.

The reported adulteration thresholds in untargeted methods range from 1 to 10 %. ([Fu et al., 2017](#), [Guijarro-Díez et al., 2015](#), [Jandrić et al., 2014](#), [Pardo-Mates et al., 2017](#), [Xu et al., 2020](#)). Substitution, is often carried out by changing the genus identification of meat, fish or plants. It comes with different issues when these species are close to each other.

### 3.2. Botanic identification

Identification of botanical species used to be harder to establish since biological procedures were not as sophisticated as they are nowadays. With the improvement of DNA techniques such as polymerase chain reaction (PCR), [next generation sequencing](#) (NGS), and [metabarcoding](#), it is now easy to identify raw products. Despite the high performance of genetic analysis, the transformations between raw material and processed product can complicate the analysis. For example, the nature of honey causes some challenges with PCR inhibition. The preliminary filtering of the sample might also result in the loss of information. Furthermore, the number of reads collected cannot be used to calculate the contribution each species in a honey sample yet. Finally, this technique alone does not ensure the absence of fraud since exogenous DNA can be added. To be exhaustive, genetic analysis must be associated with other types of techniques to guarantee the origin (biochemical analysis and melissopalynology) ([Prosser & Hebert, 2017](#)).

Chemical analysis can be interesting because it is not affected by filtered and spiked honey for example or loss of DNA. Metabolite composition can be changed due to the process but there are more targets to consider and some of them may remain unaffected. This is the main value of untargeted analysis in these situations where DNA is corrupted or added after the process ([Table 3](#)). Application fields include challenging products such as olive oil or wine where there is an important diversity of species. For instance, the OIV lists 449 species of [Vitis vinifera](#) ([OIV - International Organisation of Vine and Wine](#)) while the IOOC estimates that 139 cultivars of olives account for 85 % of the production ([International Olive Oil Council, 2000](#)).

Table 3. Summary of screened matrixes for species identification. Accuracy and biomarkers are shown when available.

Matrix (Number of sample)	Analytical technique	Statistical tools	Accuracy	Number of varieties	Biomarker	Reference
TCM (65)	UPLC-QTOF	PCA + HCA + PLS-DA	100 %	5	Terpenoids	( <a href="#">Pan et al., 2020</a> )
Tea (88)	UPLC-Orbitrap	PCA + PLS-DA + ANOVA		3	Flavonoids	( <a href="#">Fraser et al., 2013</a> )
Olive oil (51)	UPLC-QTOF	PCA + ACO-RF	–	6	Flavonoids Oleocanthal	( <a href="#">Kalogiouri et al., 2018</a> )
TCM (37)	UPLC-QTOF	PCA + PLS-DA	–	4	Lipids Amino acids Organic acid Flavonoids	( <a href="#">Zhao et al., 2017</a> )
Legumes (3)	HPLC-Orbitrap	PCA + HCA + ANOVA	–	3	Flavonoids Polyphenols Carbohydrates	( <a href="#">Llorach et al., 2019</a> )
Honey (24)	UPLC-QTOF	PCA		1	Peptides	( <a href="#">Silva et al., 2021</a> )
Wine (79)	SPME-GC-MS	PCA + LDA	100 %	5	Esters Furfural Alcohols	( <a href="#">Ziótkowska et al., 2016</a> )
Triticum (60)	DART-Orbitrap	PCA + PLS-DA + RF	> 98 %	3	Lipids Polypeptides Diacylglycerols	( <a href="#">Miano et al., 2018</a> )
Shia, linseed, sesame (28)	SPME-GC-MS	PCA + RF		3	Organic acids	( <a href="#">Erban et al., 2019</a> )
Coffee (186)	HPLC-FLD	PCA + PLS-DA	100 %	2	Caffeic acids	( <a href="#">Núñez et al., 2021</a> )
Wine (343)	UPLC-QTOF DART-QTOF	PCA + PLS-DA	95 %	8	Flavonoids Polyphenols	( <a href="#">Rubert et al., 2014</a> )
Rice (165)	SPME-GC-MS	PCA + PLS-DA + DD-SIMCA	–	7	Esters Ketones Hydrocarbons Terpenoids	( <a href="#">Ch et al., 2021</a> )
Coffee (150)	HPLC-QTOF	PCA + PLS-DA	–	9	–	( <a href="#">Souard et al., 2018</a> )
Wine (43)	UPLC-QTOF	PCA + PLS-DA	92 %	3	–	( <a href="#">Uttl et al., 2019</a> )

Matrix (Number of sample)	Analytical technique	Statistical tools	Accuracy	Number of varieties	Biomarker	Reference
Wine (18)	HPLC-QTOF	PCA	–	2	Carbohydrates Amino acids Fatty acids	( <a href="#">Arbulu et al., 2015</a> )
Hazelnut (25)	UPLC-QTOF	PLS-DA + HCA	–	6	Polyphenols Sterols	( <a href="#">Ghisoni et al., 2020</a> )

Results obtained from these analyses are often unambiguous with an accuracy above 90 %. This goes up to 100 % when trying to separate cultivars from the same areas ([Pan et al., 2020](#)). Including products from various countries increases sample diversity, ensuring that observed variations are exclusively due to species differences. Multivariate models that account for both origins and species generally require more samples to achieve comparable accuracy ([Núñez et al., 2021](#), [Rubert et al., 2014](#)). Agricultural practices are sometimes difficult to assess with analytical tools. The organic farming labels are a good example, often assessed only by [pesticides residues](#).

### 3.3. Discrimination of organic farming

Organic farming is a fast-growing market since the 1990s and often associated with higher value. Inside the European Union, it is fully implemented since 2007 ([Council Regulation \(EC\) No 834/2007, 2007](#)). It includes general principles such as the interdiction of GMOs or ionizing rays and more specific regulations for fertilizing and repellent products. To be analytically confirmed, this involves the verification of compounds with very low residue limits and the general fingerprint of the sample. High resolution mass spectrometry (HRMS) is a suitable tool to perform both targeted regulatory analysis and untargeted fingerprinting ([Ballin & Laursen, 2019](#)). Five out of seven studies presented used HRMS platforms hyphenated with liquid chromatography ([Table 4](#)). The last ones used simple MS with direct introduction systems: [electrospray](#) (DI-ESI) or [proton transfer](#) (PTR-MS).

An interesting fact in these studies is that the mean number of samples is quite lower than in the other topics, same goes for accuracy. The studies with fewer than 40 samples achieve lower classification accuracy ranging from 73 % to 85 % ([Martínez Bueno et al., 2018](#), [Mie et al., 2014](#), [Xiao et al., 2018](#)). It is difficult to determine whether these results are due to a lower number of samples, or the quality of authentic references, or the quantified differences. Comparatively, the research on carrots with a higher number of samples showed a better and equivalent accuracy with studies on other frauds ([Cubero-Leon et al., 2018](#)). This would imply that this type of fraud needs more than 40 samples to achieve proper classification. Another observation is that organic farming assessment is mainly oriented into fruits and vegetables. Only two studies were interested in meat and starch ([Oliveira et al., 2015](#), [Xiao et al., 2018](#)). In general, meat is not the most studied product in untargeted approaches since fraud possibilities can often be resolved by targeted approaches. For species identification, DNA is almost always recoverable and NGS allows reliable confirmation ([Xing et al., 2019](#)). Other frauds such as geographic origin or farming practice can be fought with techniques like isotope or heavy metals measurement ([Liu et al., 2020](#), [Potočnik et al., 2020](#)). Organic products are generally associated with high quality items, which are also a target for fraud because of the potential profits.

### 3.4. Evaluation of premium qualities

High and premium quality products are associated with specific human activity (e.g., savoir-faire) and geographic location. The location influences geological composition, climate, water sources, and [phenology](#). All these aspects account for the so called “terroir effect” ([Seguin, 1986](#)). Terroir is an element mainly considered for wine cultures. However, it could be extended to any product in which at least one of these elements is essential to its quality. Quality evaluation includes different aspects: vintage, ageing or ripening time, expert classification, or specific quality appellations ([Table 4](#)). Origin can also be an interesting criterion, keeping in mind that it must be examined separately as it raises its own set of concerns.

Table 4. Summary of screened matrixes for premium qualities. Accuracy and biomarkers are shown when available.

Matrix (Number of sample)	Analytical technique	Statistical tools	Accuracy	Quality aspect	Biomarker	Reference
Tomatoes (25)	LC-Orbitrap	PCA	73 %	Farming practice	Flavonoids Amino acids Carotenoids	( <a href="#">Martínez Bueno et al., 2018</a> )
Pork (41)	ESI-MS/MS PTR-MS	PCA + SIMCA + ANOVA	100 %	Farming practice	FAME <sup>a</sup>	( <a href="#">Oliveira et al., 2015</a> )
Rice (40)	UPLC-QTOF	PCA + PLS-DA + HCA	85 %	Farming practice	Amino alcohol Anthocyanin Lignans	( <a href="#">Xiao et al., 2018</a> )
Cabbage (18)	UPLC-QTOF	PCA + PLS-DA	83 %	Farming practice	Amino acids Carbohydrates	( <a href="#">Mie et al., 2014</a> )
Tomatoes (18)	DI-ESI-MS	ANOVA + RF + HCA	–	Farming practice	–	( <a href="#">García-Casarrubias et al., 2019</a> )
Carrots (140)	UPLC-QTOF	PCA + PLS-DA	100 %	Farming practice	Carbohydrates Amino acids Terpenoids	( <a href="#">Cubero-Leon et al., 2018</a> )
Physalis (10)	UPLC-QTOF	PCA	–	Farming practice	Steroids Fatty acids	( <a href="#">Llano et al., 2018</a> )
Beer (265)	DART-QTOF	PLS-DA + LDA + ANN-MLP	100 %	Trappist beer	Carbohydrates Amino acids Organic acids	( <a href="#">Cajka et al., 2011</a> )



Matrix (Number of sample)	Analytical technique	Statistical tools	Accuracy	Quality aspect	Biomarker	Reference
Vegetable puree (18)	UPLC-QTOF NMR SPME-GC-MS	–	85 %	Industrial processing	Oxylipins Volatiles	( <a href="#">Lopez-Sanchez et al., 2015</a> )
Wine (315)	Raman	PCA	84 %	Maturation time	–	( <a href="#">Mandrile et al., 2016</a> )
Wine (26)	GC-MS	PCA	–	Vintage Maturation time	Carbohydrates Glycerol	( <a href="#">Schueuermann et al., 2016</a> )
Coffee (186)	HPLC-FLD	PCA + PLS-DA	100 %	Roasting degree	Caffeic acids	( <a href="#">Núñez et al., 2021</a> )
Ground coffee (47)	UPLC-QTOF SPME-GC-MS	PCA + PLS-DA + HCA	–	Overall quality	Methyl pentanoate 2-Furfurylthiol l-Homoserine Amino acids	( <a href="#">Rocchetti et al., 2020</a> )
Cheese (96)	HPLC-Orbitrap	PCA + PLS-DA	100 %	Ripening time	Organic acids Vitamins Carbohydrates	( <a href="#">Le Boucher et al., 2015</a> )
Angelica (20)	GC-TOF-MS	PCA	–	Overall quality	Amino acids Organic acids Amino acids	( <a href="#">Tianniam et al., 2008</a> )
Cheese (10)	GC-MS	PCA	–	Ripening time	Organic acids Fatty acids Carbohydrates	( <a href="#">Afshari et al., 2020</a> )
Olive oil (108)	GC-MS	PCA + PLS-DA	85 %	Overall quality	Volatiles	( <a href="#">Sales et al., 2019</a> )

a

Fatty acids [methyl ester](#).

Quality parameters are generally deciphered by experts of the fields. Many works are released every year, for example on vintages and appellations for wine with the Robert Parker Wine Advocate ([Robert Parker Wine Advocate, 2020](#)). Research works focused on the quality

aspects we deal with every day. Cheese, coffee, and alcoholic beverages are among the most consumed products daily ([Comprehensive European Food Consumption Database, 2021](#)). Overall quality being evaluated by sensory experts, it may be difficult to obtain identical results with analytical tools ([Rocchetti et al., 2020](#), [Sales et al., 2019](#), [Tianniam et al., 2008](#)). Maturation time for wine influences aroma, color, and phenolic composition. For example, Barolo and Barbaresco are both monovarietal wines that are aged over different times, at least 26 and 38 months respectively. They were compared with [Raman spectroscopy](#) with a prediction ability of 84 % ([Mandrile et al., 2016](#)). Wine being a complex product, a lot of variables may affect its ageing process: the age of the barrel, the temperature, the amount of oxygen and so on. For cheese, ripening time was studied either in model cheese or commercial denominations ([Afshari et al., 2020](#), [Le Boucher et al., 2015](#)). Differences were found, especially in the [microbial metabolism](#) that plays a key role in [cheese aroma](#). For the classification of beer, a [multilayer perceptron](#) (MLP) with ANN was used to predict quality with the global pattern of markers instead of individual consideration ([Cajka et al., 2011](#)). Another observation is the predominance of carbohydrates and amino acids as biomarkers in these studies. Sugars and amino acids play an important role in taste and might enhance the overall sensorial quality of the products ([Kochem, 2017](#)). This could explain the fact that six out of ten studies found them relevant biomarkers. Official signs identifying quality and origin are important to identify these high value products in Europe. While quality is frequently associated with provenance, origin can be an issue in and of itself.

### 3.5. Geographic origin

Geographic location is a key element for both prices and terroir recognition. Controlled appellations of origins such as geographic indication (GI) or protected designation of origin (PDO) are tools to protect high value products inside the European Union. Geographic origin is often a primary condition to be included in these labels. However, they often require additional criteria. These can be related to species, farming practices and savoir-faire ([Regulation \(EU\) No 1151/2012 of the European Parliament and of the Council of 21 November 2012 on Quality Schemes for Agricultural Products and Foodstuffs, 2012](#)). This fraud is probably the most spread and received the most attention from researchers. Many products from almost all categories were studied ([Table 5](#)).

Table 5. Summary of screened matrixes for geographic origin. Accuracy and biomarkers are shown when available.

Matrix (Number of sample)	Analytical technique	Statistical tools	Accuracy	Number of area	Biomarker	Reference
Serrano pepper (40)	NMR	PCA + PLS-DA	–	2	Amino acids Organic acids Carbohydrates	<a href="#">(Becerra-Martínez et al., 2017)</a>
Tea (88)	UPLC-Orbitrap	PCA + PLS-DA + ANOVA	–	2	Flavonoids	<a href="#">(Fraser et al., 2013)</a>
Paprika (122)	HPLC-FLD	PLS-DA	97,9%	5	Phenols	<a href="#">(Campmajó et al., 2017)</a>

Matrix (Number of sample)	Analytical technique	Statistical tools	Accuracy	Number of area	Biomarker	Reference
						<a href="#">(2021)</a>
Columbia coffee (41)	UPLC-QTOF	PCA + PLS-DA	95 %	5	Caffeic acids Organic acids Amino acids Carbohydrates	<a href="#">(Hoyos Ossa et al., 2018)</a>
Table olive (96)	UPLC-QTOF	PCA + PLS-DA	–	3	Phenols Flavonoids Fatty acids	<a href="#">(Kalogiouri et al., 2020)</a>
Margarine (35)	HPLC-DAD	PCA + PLS-DA + SIMCA	100 %	2	Fatty acids	<a href="#">(Bikrani et al., 2019)</a>
Lemon EO (16)	UPLC-QTOF NMR	PCA + PLS-DA	–	2	Flavonoids Fatty acids Sesquiterpenes	<a href="#">(Marti et al., 2014)</a>
Wine (315)	Raman	PCA	90 %	5	–	<a href="#">(Mandrile et al., 2016)</a>
Wine (79)	SPME-GC-MS	PCA + LDA	> 90 %	9	Esters Furfural Alcohols	<a href="#">(Ziółkowska et al., 2016)</a>
Oranges (137)	SPME-GC-MS e-nose	PCA + LDA + PLS-DA	97,8%	3	Terpenoids Heterocycles	<a href="#">(Centonze et al., 2019)</a>
Wine (60)	UPLC-QTOF	RF + GA	86,7%	3	Phenols	<a href="#">(Gil et al., 2020)</a>
Beef (40)	NMR	PCA + PLS-DA + ANOVA	–	4	Amino acids	<a href="#">(Jung et al., 2010)</a>
Hazelnut (236)	UPLC-QTOF	PCA + LDA + SIMCA + SVM	99,5%	5	Phospholipids Acylglycerol Secoiridoids Amino acids	<a href="#">(Klockmann et al., 2016)</a>
Indonesia coffee (80)	GC-MS	PCA	–	7	Organic acids Carbohydrates	<a href="#">(Putri et al., 2019)</a>
Wine (152)	UPLC-QTOF FT-ICR	PCA		5	–	<a href="#">(Roullier-Gall, Witting, et al., 2014)</a>

Matrix (Number of sample)	Analytical technique	Statistical tools	Accuracy	Number of area	Biomarker	Reference
Coffee (186)	HPLC-FLD	PCA + PLS-DA	100 %	7	Caffeic acids	<a href="#">(Núñez et al., 2021)</a>
Saffron (27)	UPLC-QTOF	HCA + PLS-DA	–	4	Flavonoids Phenols	<a href="#">(Senizza et al., 2019)</a>
Wheat (94)	UPLC-QTOF	PCA + PLS-DA	–	8	Fatty acids Lipids	<a href="#">(Jin et al., 2020)</a>
Beef (39)	UPLC-Orbitrap GC-MS	PCA + PLS-DA	> 90 %	3	Amino acids Phospholipids	<a href="#">(Man et al., 2021)</a>
Asparagus (317)	UPLC-QTOF	PCA + PLS-DA + ANOVA	> 82 %	5	Lipids	<a href="#">(Creydt &amp; Fischer, 2020)</a>
Olive oil (90)	UPLC-QTOF	PCA + PLS-DA	87 %	6	Fatty acids Carbohydrates Flavonoids	<a href="#">(Gil-Solsona et al., 2016)</a>
Saffron (20)	UPLC-QTOF	PCA + PLS-DA	100 %	2	Organic acids Carbohydrates Esters	<a href="#">(Guijarro-Díez et al., 2015)</a>
Rice (165)	SPME-GC-MS	PCA + PLS-DA + DD-SIMCA	100 %	3	Hydrocarbons Terpenoids	<a href="#">(Ch et al., 2021)</a>
Violet leaf extracts (28)	UPLC-QTOF	PCA	–	2	–	<a href="#">(Saint-Lary et al., 2016)</a>
Grape marc (57)	NMR	PCA + PLS-DA	100 %	2	Esters Alcohols Organic acids	<a href="#">(Fotakis &amp; Zervou, 2016)</a>
Milk (180)	NMR	PLS-DA	> 95 %	10	–	<a href="#">(Tenori et al., 2018)</a>
Angelica (6)	UPLC-QTOF	PCA + PLS-DA	–	3	Flavonoids Carbohydrates Phospholipids	<a href="#">(Tianniam et al., 2009)</a>
Rice (80)	HPLC-QTOF GC-MS	PCA + PLS-DA + RF	> 95 %	2	Carbohydrates Fatty acids	<a href="#">(Lim et al., 2018)</a>
Cheese	UPLC-	PLS-DA	–	2	Lipids	<a href="#">(Rocchetti,</a>

Matrix (Number of sample)	Analytical technique	Statistical tools	Accuracy	Number of area	Biomarker	Reference
(20)	QTOF				Amino acids Oligopeptides	<a href="#">Lucini, et al., 2018</a>
Wine (6)	UPLC-QTOF	PLS-DA + HCA	–	6	Polyphenols	<a href="#">(Rocchetti, Gatti, et al., 2018)</a>
Hazelnut (25)	UPLC-QTOF	PLS-DA + HCA	–	3	Polyphenols Sterols	<a href="#">(Ghisoni et al., 2020)</a>
Wine (110)	UPLC-QTOF	PCA	–	12	Flavonoids Amino acids	<a href="#">(Arapitsas et al., 2020)</a>

Out of 32 studies, wine was the matrix of interest in six of them. These studies are representative of the scales that can be considered. Four of them focused on discrimination at the regional or subregional level ([Arapitsas et al., 2020](#), [Gil et al., 2020](#), [Mandrile et al., 2016](#), [Roullier-Gall et al., 2014](#)). Overlaps in the PCA are observed in every case. It means that geographical proximity is associated with very slim differences between samples, even from different cultivars. For the last two studies, one discriminates samples from nine different countries with a classification rate over 90 % ([Ziółkowska et al., 2016](#)). The other one, only considered the chardonnays from three areas in Italy and three other countries: France, Israel, and Australia ([Rocchetti, Gatti, et al., 2018](#)). Origins were well separated apart from Aosta Valley in Italy and Burgundy in France with an almost perfect overlap. This discriminant analysis was performed based on the phenolic profile. Therefore, it suggests that phenols alone are not enough to discriminate wines from different origins. This could be explained by the fact they are produced by the process and not environmentally born or terroir specific.

For other matrixes, different approaches were considered, both in terms of analytical platforms and mathematical frameworks. NMR is an untargeted instrument that is often used for origin assessment ([Becerra-Martínez et al., 2017](#), [Fotakis and Zervou, 2016](#), [Jung et al., 2010](#), [Marti et al., 2014](#), [Tenori et al., 2018](#)). It is also the only method to be accredited in Europe for multivariate analysis (Deutsche Akkreditierungsstelle [GmbH, 2020](#)). An impressive work was done on milk by metabolomic fingerprinting: milk samples from ten farms were acquired and analyzed by NMR. Multivariate analysis achieved proper clustering for eight out of ten farms with very little overlaps, using only 19 metabolites ([Tenori et al., 2018](#)). It demonstrates the potential of NMR as an effective and powerful technique for non-targeted analysis.

The statistical method is almost as important as the analytical method. It is mainly carried out by PCA or PLS-DA, in 82 % and 65 % of the articles, respectively. However, two other algorithms were evaluated for origin analysis. A genetic algorithm (GA) inspired by natural selection was used for the selection of features. Better results were achieved with this approach compare to random forest on the same set of data ([Gil et al., 2020](#)). [Support vector machine](#) (SVM) is a supervised algorithm used to perform data clustering. A model with SVM and SIMCA was built to avoid false positive, samples were either properly assigned or not assigned with this combination ([Klockmann et al., 2016](#)). This demonstrates the analytical

capability to attribute a production area to an authentic product. It is important to notice that adulterated samples were not considered in these studies. Mixing different vintages, or different areas of production might make origin discrimination considerably harder, and this is certainly the most common scenario. If we combine these findings, we may describe terroir as the total of organic, biologic, and geologic environmental indicators, as well as processed born molecules from savoir faire and barrels, mills, or hives for example.

## 4. Discussion and perspectives

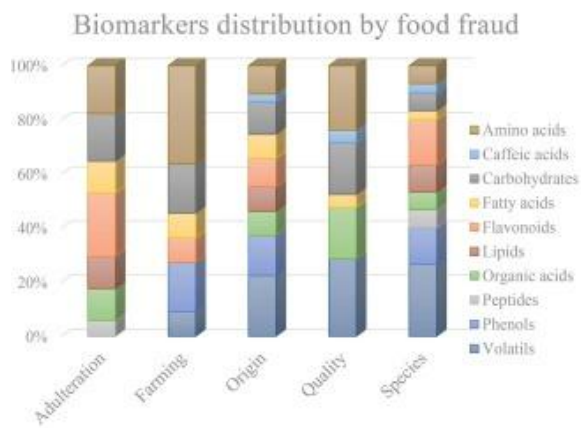
To summarize, recent advances in metabolomics coupled with the expansion of non-targeted and multivariate analysis allow promising results in the context of food fraud. Technologies including mass spectrometry and spectroscopy are relevant for authentication applications such as non-adulteration, validation of species, origin, farming practice or quality determination.

Some steps of the non-targeted workflow are reaching consensus such as the performance evaluation or the sampling criteria. Nonetheless, many of them still need to be standardized. In the coming years, concerns about the number of samples, the interlaboratory testing or the establishment of certified reference material (CRM) will certainly get attention.

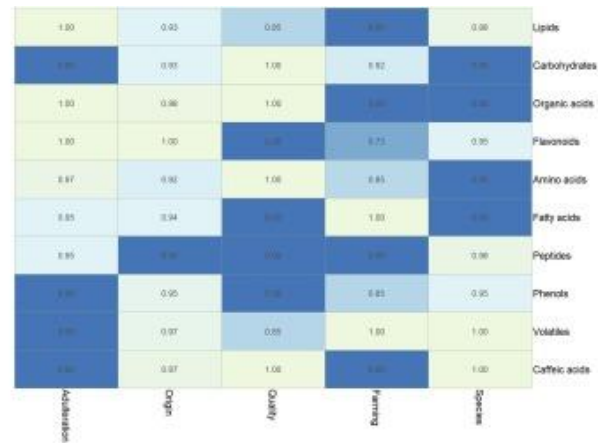
Overall, this literature review reveals that successful classification can be achieved in the case of authentic samples. Unexpected exogen adulteration is also well understood and already standardized for uses in food controls. However, no studies evaluated real-life scenario where these frauds can be mixed. For example, a wine sample adulterated with 20 % of a cheaper wine from a different origin: classification ability will certainly decrease in these cases. Future research might increase performance of multivariate modelling, but it seems difficult to achieve exhaustive models for each product.

There are different perspectives to tackle this issue. One of them is the increasing development of data fusion in the context of food authenticity ([Biancolillo et al., 2019](#)). For example, in determining the origin of an unknown sample, elemental or isotope analysis could narrow the search area. Non-targeted approaches can then be used to accomplish closer discrimination.

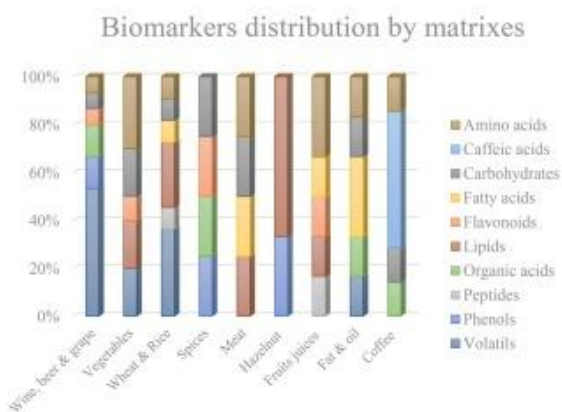
Another idea stands in the biomarkers identified in these studies. The distribution of the biomarkers regarding the type of fraud and their accuracy was analyzed [[Fig. 1](#)]. It appears clearly that some markers are more relevant than others: Amino acids, carbohydrates, [flavonoids](#), fatty acids, organic acids, and volatiles are found in almost every category. Accuracy was not available in all studies, but at least one of these biomarkers achieve a 100 % classification ability in each category. Lipids and phenols are also used in three of the five frauds considered and could be worth the attention. Because [proteomics](#) is a very new field in the context of authenticity ([Creydt & Fischer, 2018](#)), it can explain the underrepresentation of peptides that are certainly interesting compounds [[Fig. 1](#)].



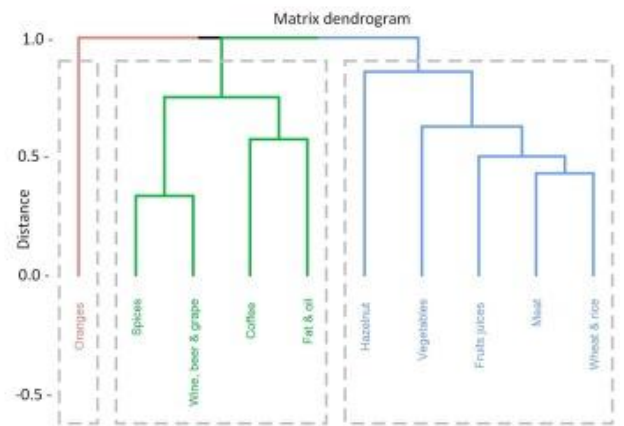
(a)



(b)



(c)



(d)

Fig. 1. (a) Distribution of biomarkers depending on the food fraud (b) Heatmap of mean accuracy depending on the biomarkers and the fraud considered. When 0 is displayed, it means no values were available. (c) Distribution of biomarkers depending on the matrix. matrixes with only one type of biomarkers were removed (d) Hierarchical cluster of matrixes depending on the biomarkers (distance: binary).

Based on these conclusions, the distribution and clustering of these biomarkers depending on the matrix was also analyzed [Fig. 1]. The first observation is that [caffeic acids](#) are matrix specific and cannot be considered as good terroir biomarkers. Next are the two most interesting markers in this distribution: amino acids and carbohydrates as they are covering seven of the ten groups of matrixes. [Volatile compounds](#), fatty acids and lipids can be added to these two markers to account for almost 70 % of the total, covering all groups with high accuracy.

The hierarchical cluster was built based on the biomarkers found in each matrix. It shows two groups: one is composed of spices, coffee, fat and oil, wine, beer, and grapes. The only relationship between those could be the organoleptic importance for each of them. The other group is made of vegetables, juices, meat, wheat, and rice. All four of them were studied with lipids and amino acids extensively.



To conclude, this could pave the way for a new strategy: instead of mapping all matrixes for all areas, there is a possibility that some biomarkers might be environmentally born and not matrix specific. These biomarkers could be used as terroir markers and be a real game changer in the perception of the terroir effect. This effect was studied in the case of wine ([Roullier-Gall, Lucio, et al., 2014](#)), where it was found that greater differences were observed within a given vintage. However, to our knowledge, the thesis of finding terroir biomarkers across various products has not been evaluated yet. It will surely be the framework of future research on new ways to ensure food authenticity at a global and reliable scale.

## 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

We would like to thank Dr. Damien Cabanes and Dr. Stephane Mussard for proofreading the manuscript before it was submitted and Ms Lea Sanchez for her help with the graphical abstract. This work was supported by the French Association Nationale de la Recherche et de la Technologie (ANRT) under the grant n°2020/1587.

## Appendix A. Supplementary data

[Download : Download Acrobat PDF file \(188KB\) : https://ars.els-cdn.com/content/image/1-s2.0-S0308814622018180-mmc1.pdf](https://ars.els-cdn.com/content/image/1-s2.0-S0308814622018180-mmc1.pdf)

Supplementary data 1.

[Download : Download Word document \(39KB\) : https://ars.els-cdn.com/content/image/1-s2.0-S0308814622018180-mmc2.docx](https://ars.els-cdn.com/content/image/1-s2.0-S0308814622018180-mmc2.docx)

Supplementary data 2.

## Data availability

Data will be made available on request.

## References

[Abbatangelo and Sberveglieri, 2019](#)

M. Abbatangelo, V. Sberveglieri

**Multidisciplinary approach to characterizing the fingerprint of Italian EVOO**

Molecules, 24 (8) (2019), p. 1457, [10.3390/molecules24081457](#)

[Accum, 1820](#)

F. Accum

**A treatise on adulterations of food**

And Culinary Poisons., 69 (1820)

[Afshari et al., 2020](#)

R. Afshari, C.J. Pillidge, D.A. Dias, A.M. Osborn, H. Gill  
**Microbiota and metabolite profiling combined with integrative analysis for differentiating cheeses of varying ripening ages**  
Frontiers in Microbiology, 11 (2020), Article 592060, [10.3389/fmicb.2020.592060](#)

[Alonso et al., 2015](#)

A. Alonso, S. Marsal, A. JuliÀ  
**Analytical methods in untargeted metabolomics: State of the art in 2015**  
Frontiers in Bioengineering and Biotechnology, 3 (2015), [10.3389/fbioe.2015.00023](#)

[Arapitsas et al., 2020](#)

P. Arapitsas, M. Ugliano, M. Marangon, P. Piombino, L. Rolle, V. Gerbi, ..., F. Mattivi  
**Use of untargeted liquid chromatography-mass spectrometry metabolome to discriminate italian monovarietal red wines, produced in their different terroirs**  
Journal of Agricultural and Food Chemistry, 68 (47) (2020), pp. 13353-13366,  
[10.1021/acs.jafc.0c00879](#)

[Arbulu et al., 2015](#)

M. Arbulu, M.C. Sampedro, A. Gómez-Caballero, M.A. Goicolea, R.J. Barrio  
**Untargeted metabolomic analysis using liquid chromatography quadrupole time-of-flight mass spectrometry for non-volatile profiling of wines**  
Analytica Chimica Acta, 858 (2015), pp. 32-41, [10.1016/j.aca.2014.12.028](#)

[Ballin and Laursen, 2019](#)

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

[Becerra-Martínez et al., 2017](#)

E. Becerra-Martínez, E. Florentino-Ramos, N. Pérez-Hernández, L. Gerardo Zepeda-Vallejo, N. Villa-Ruano, M. Velázquez-Ponce, ..., A.E. Bañuelos-Hernández  
**<sup>1</sup>H NMR-based metabolomic fingerprinting to determine metabolite levels in serrano peppers (Capsicum annum L.) grown in two different regions**  
Food Research International, 102 (2017), pp. 163-170, [10.1016/j.foodres.2017.10.005](#)

[Beneito-Cambra et al., 2020](#)

M. Beneito-Cambra, D. Moreno-González, J.F. García-Reyes, M. Bouza, B. Gilbert-López, A. Molina-Díaz  
**Direct analysis of olive oil and other vegetable oils by mass spectrometry: A review**  
Trends in Analytical Chemistry, 132 (2020), Article 116046,  
[10.1016/j.trac.2020.116046](#)

[Biancolillo et al., 2019](#)

Biancolillo, A., Boqué, R., Cocchi, M., & Marini, F. (2019). Data Fusion Strategies in Food Analysis. In *Data Handling in Science and Technology* (Vol. 31, pp. 271–310). Elsevier. <https://doi.org/10.1016/B978-0-444-63984-4.00010-7>.

[Bikrani et al., 2019](#)

S. Bikrani, A.M. Jiménez-Carvelo, M. Nechar, M.G. Bagur-González, B. Souhail, L. Cuadros-Rodríguez

**Authentication of the geographical origin of margarines and fat-spread products from liquid chromatographic UV-absorption fingerprints and chemometrics**  
Foods, 8 (11) (2019), p. 588, [10.3390/foods8110588](https://doi.org/10.3390/foods8110588)

[Cajka et al., 2011](#)

T. Cajka, K. Riddellova, M. Tomaniova, J. Hajslova

**Ambient mass spectrometry employing a DART ion source for metabolomic fingerprinting/profiling: A powerful tool for beer origin recognition**  
Metabolomics, 7 (4) (2011), pp. 500-508, [10.1007/s11306-010-0266-z](https://doi.org/10.1007/s11306-010-0266-z)

[Campmajó et al., 2021](#)

G. Campmajó, L.R. Rodríguez-Javier, J. Saurina, O. Núñez

**Assessment of paprika geographical origin fraud by high-performance liquid chromatography with fluorescence detection (HPLC-FLD) fingerprinting**  
Food Chemistry, 352 (2021), Article 129397, [10.1016/j.foodchem.2021.129397](https://doi.org/10.1016/j.foodchem.2021.129397)

[Cavin et al., 2016](#)

C. Cavin, G. Cottenet, C. Blancpain, T. Bessaire, N. Frank, P. Zbinden

**Food Adulteration: From vulnerability assessment to new analytical solutions**  
CHIMIA International Journal for Chemistry, 70 (5) (2016), pp. 329-333, [10.2533/chimia.2016.329](https://doi.org/10.2533/chimia.2016.329)

[CWA 17369: Authentic and Fraud in the Feed and Food Chain- Concepts, Terms and Definitions, 2019](#)

*CWA 17369: Authentic and fraud in the feed and food chain- Concepts, terms and definitions*, (2019) (testimony of CEN, European Committee for Standardization & CEN-CENELEC Management Centre (2019)).

[Centonze et al., 2019](#)

V. Centonze, V. Lippolis, S. Cervellieri, A. Damascelli, G. Casiello, M. Pascale, ..., F. Longobardi

**Discrimination of geographical origin of oranges (*Citrus sinensis* L. Osbeck) by mass spectrometry-based electronic nose and characterization of volatile compounds**

Food Chemistry, 277 (2019), pp. 25-30, [10.1016/j.foodchem.2018.10.105](https://doi.org/10.1016/j.foodchem.2018.10.105)

[Ch et al., 2021](#)

R. Ch, O. Chevallier, P. McCarron, T.F. McGrath, D. Wu, L. Nguyen Doan Duy, A.P. Kapil, M. McBride, C.T. Elliott

**Metabolomic fingerprinting of volatile organic compounds for the geographical discrimination of rice samples from China, Vietnam and India**

Food Chemistry, 334 (2021), Article 127553, [10.1016/j.foodchem.2020.127553](https://doi.org/10.1016/j.foodchem.2020.127553)

[Comprehensive European Food Consumption Database, 2021](#)

Comprehensive European Food Consumption Database. (2021).  
<https://www.efsa.europa.eu/en/microstrategy/foodex2-level-1>.

[Creydt and Fischer, 2018](#)

M. Creydt, M. Fischer

**Omics approaches for food authentication**

*ELECTROPHORESIS*, 39 (13) (2018), pp. 1569-1581, [10.1002/elps.201800004](https://doi.org/10.1002/elps.201800004)

[Creydt and Fischer, 2020](#)

M. Creydt, M. Fischer

**Mass-spectrometry-based food metabolomics in routine applications: A basic standardization approach using housekeeping metabolites for the authentication of asparagus**

*Journal of Agricultural and Food Chemistry*, 68 (49) (2020), pp. 14343-14352,  
[10.1021/acs.jafc.0c01204](https://doi.org/10.1021/acs.jafc.0c01204)

[Cuadros-Rodríguez et al., 2016](#)

L. Cuadros-Rodríguez, C. Ruiz-Samblás, L. Valverde-Som, E. Pérez-Castaño, A. González-Casado

**Chromatographic fingerprinting: An innovative approach for food “identification” and food authentication – A tutorial**

*Analytica Chimica Acta*, 909 (2016), pp. 9-23, [10.1016/j.aca.2015.12.042](https://doi.org/10.1016/j.aca.2015.12.042)

[Cubero-Leon et al., 2018](#)

E. Cubero-Leon, O. De Rudder, A. Maquet

**Metabolomics for organic food authentication: Results from a long-term field study in carrots**

*Food Chemistry*, 239 (2018), pp. 760-770, [10.1016/j.foodchem.2017.06.161](https://doi.org/10.1016/j.foodchem.2017.06.161)

[Cubero-Leon et al., 2014](#)

E. Cubero-Leon, R. Peñalver, A. Maquet

**Review on metabolomics for food authentication**

*Food Research International*, 60 (2014), pp. 95-107, [10.1016/j.foodres.2013.11.041](https://doi.org/10.1016/j.foodres.2013.11.041)

[GmbH, 2020](#)

Deutsche Akkreditierungsstelle GmbH. (2020). *Appendix to accreditation certificate D-PL-19229-01-00 according to DIN EN ISO/IEC 17025:2018*.

[Ellis et al., 2012](#)

D.I. Ellis, V.L. Brewster, W.B. Dunn, J.W. Allwood, A.P. Golovanov, R. Goodacre

**Fingerprinting food: Current technologies for the detection of food adulteration and contamination**

*Chemical Society Reviews*, 41 (17) (2012), p. 5706, [10.1039/c2cs35138b](https://doi.org/10.1039/c2cs35138b)

[Erban et al., 2019](#)

A. Erban, I. Fehrlé, F. Martinez-Seidel, F. Brigante, A.L. Más, V. Baroni, ..., J. Kopka

**Discovery of food identity markers by metabolomics and machine learning technology**

*Scientific Reports*, 9 (1) (2019), p. 9697, [10.1038/s41598-019-46113-y](https://doi.org/10.1038/s41598-019-46113-y)

[European Commission. Directorate General for Health and Food Safety, 2020](#)

European Commission. Directorate General for Health and Food Safety. (2020). *The EU food fraud network and the administrative assistance and cooperation system: 2019 annual report*. Publications Office. <https://data.europa.eu/doi/10.2875/326318>.

[Food Crime Strategic Assessment, 2020](#)

*Food Crime Strategic Assessment 2020* (p. 62). (2020). National Food Crime Unit and Scottish Food Crime and Incidents Unit.

[Fotakis and Zervou, 2016](#)

C. Fotakis, M. Zervou

**NMR metabolic fingerprinting and chemometrics driven authentication of Greek grape marc spirits**

Food Chemistry, 196 (2016), pp. 760-768, [10.1016/j.foodchem.2015.10.002](https://doi.org/10.1016/j.foodchem.2015.10.002)

[Fraser et al., 2013](#)

K. Fraser, G.A. Lane, D.E. Otter, Y. Hemar, S.-Y. Quek, S.J. Harrison, S. Rasmussen  
**Analysis of metabolic markers of tea origin by UHPLC and high resolution mass spectrometry**

Food Research International, 53 (2) (2013), pp. 827-835,

[10.1016/j.foodres.2012.10.015](https://doi.org/10.1016/j.foodres.2012.10.015)

[Fu et al., 2017](#)

H. Fu, Q. Yin, L. Xu, W. Wang, F. Chen, T. Yang

**A comprehensive quality evaluation method by FT-NIR spectroscopy and chemometric: Fine classification and untargeted authentication against multiple frauds for Chinese Ganoderma lucidum**

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 182 (2017), pp. 17-25, [10.1016/j.saa.2017.03.074](https://doi.org/10.1016/j.saa.2017.03.074)

[García-Casarrubias et al., 2019](#)

A. García-Casarrubias, R. Winkler, A. Tiessen

**Mass fingerprints of tomatoes fertilized with different nitrogen sources reveal potential biomarkers of organic farming**

Plant Foods for Human Nutrition, 74 (2) (2019), pp. 247-254, [10.1007/s11130-019-00726-w](https://doi.org/10.1007/s11130-019-00726-w)

[Ghisoni et al., 2020](#)

S. Ghisoni, L. Lucini, G. Rocchetti, G. Chiodelli, D. Farinelli, S. Tombesi, M. Trevisan

**Untargeted metabolomics with multivariate analysis to discriminate hazelnut (*Corylus avellana* L.) cultivars and their geographical origin**

Journal of the Science of Food and Agriculture, 100 (2) (2020), pp. 500-508,

[10.1002/jsfa.9998](https://doi.org/10.1002/jsfa.9998)

[Gil et al., 2020](#)

M. Gil, C. Reynes, G. Cazals, C. Enjalbal, R. Sabatier, C. Saucier

**Discrimination of rosé wines using shotgun metabolomics with a genetic algorithm and MS ion intensity ratios**

Scientific Reports, 10 (1) (2020), p. 1170, [10.1038/s41598-020-58193-2](https://doi.org/10.1038/s41598-020-58193-2)

[Gil-Solsona et al., 2016](#)

R. Gil-Solsona, M. Raro, C. Sales, L. Lacalle, R. Díaz, M. Ibáñez, ..., F.J. Hernández  
**Metabolomic approach for Extra virgin olive oil origin discrimination making use of ultra-high performance liquid chromatography – Quadrupole time-of-flight mass spectrometry**  
Food Control, 70 (2016), pp. 350-359, [10.1016/j.foodcont.2016.06.008](https://doi.org/10.1016/j.foodcont.2016.06.008)

[Gliszczyńska-Świgło and Chmielewski, 2017](#)

A. Gliszczyńska-Świgło, J. Chmielewski  
**Electronic nose as a tool for monitoring the authenticity of food. A Review**  
Food Analytical Methods, 10 (6) (2017), pp. 1800-1816, [10.1007/s12161-016-0739-4](https://doi.org/10.1007/s12161-016-0739-4)

[Goodacre et al., 2002](#)

R. Goodacre, S. Vaidyanathan, G. Bianchi, D.B. Kell  
**Metabolic profiling using direct infusion electrospray ionisation mass spectrometry for the characterisation of olive oils**  
The Analyst, 127 (11) (2002), pp. 1457-1462, [10.1039/b206037j](https://doi.org/10.1039/b206037j)

[Guijarro-Díez et al., 2015](#)

M. Guijarro-Díez, L. Nozal, M.L. Marina, A.L. Crego  
**Metabolomic fingerprinting of saffron by LC/MS: Novel authenticity markers**  
Analytical and Bioanalytical Chemistry, 407 (23) (2015), pp. 7197-7213,  
[10.1007/s00216-015-8882-0](https://doi.org/10.1007/s00216-015-8882-0)

[Gumus et al., 2020](#)

O. Gumus, Y. Erkan, G.Z. Pinar, E. Hasan  
**Comparison of different classification algorithms to identify geographic origins of olive oils**  
Journal of Food Science and Technology, 57 (2020), pp. 1535-1543, [10.1007/s13197-019-04189-4](https://doi.org/10.1007/s13197-019-04189-4)

[Gumus et al., 2018](#)

Z.P. Gumus, H. Ertas, E. Yasar, O. Gumus  
**Classification of olive oils using chromatography, principal component analysis and artificial neural network modelling**  
Journal of Food Measurement and Characterization, 12 (2) (2018), pp. 1325-1333,  
[10.1007/s11694-018-9746-z](https://doi.org/10.1007/s11694-018-9746-z)

[Gunning et al., 2018](#)

Y. Gunning, M. Defernez, A.D. Watson, N. Beadman, I.J. Colquhoun, G. Le Gall, ..., E.K. Kemsley  
**16-O-methylcafestol is present in ground roast Arabica coffees: Implications for authenticity testing**  
Food Chemistry, 248 (2018), pp. 52-60, [10.1016/j.foodchem.2017.12.034](https://doi.org/10.1016/j.foodchem.2017.12.034)

[Hoyos Ossa et al., 2018](#)

D.E. Hoyos Ossa, R. Gil-Solsona, G.A. Peñuela, J.V. Sancho, F.J. Hernández  
**Assessment of protected designation of origin for Colombian coffees based on HRMS-based metabolomics**  
Food Chemistry, 250 (2018), pp. 89-97, [10.1016/j.foodchem.2018.01.038](https://doi.org/10.1016/j.foodchem.2018.01.038)

[Indelicato et al., 2017](#)

S. Indelicato, D. Bongiorno, R. Pitonzo, V. Di Stefano, V. Calabrese, S. Indelicato, G. Avellone

**Triacylglycerols in edible oils: Determination, characterization, quantitation, chemometric approach and evaluation of adulterations**

Journal of Chromatography A, 1515 (2017), pp. 1-16, [10.1016/j.chroma.2017.08.002](https://doi.org/10.1016/j.chroma.2017.08.002)

[International Olive Oil Council, 2000](#)

International Olive Oil Council. (2000). *World Catalogue of Olive Varieties* (1st ed.).

[Jandrić et al., 2014](#)

Z. Jandrić, D. Roberts, M.N. Rathor, A. Abraham, M. Islam, A. Cannavan

**Assessment of fruit juice authenticity using UPLC–QToF MS: A metabolomics approach**

Food Chemistry, 148 (2014), pp. 7-17, [10.1016/j.foodchem.2013.10.014](https://doi.org/10.1016/j.foodchem.2013.10.014)

[Jha, 2016](#)

S.N. Jha

**Food safety and quality**

10.1016/B978-0-12-420084-5.00001-9

Rapid Detection of Food Adulterants and Contaminants, Elsevier (2016), pp. 1-24

[Jin et al., 2020](#)

M. Jin, W. Zheng, Y. Zhang, B. Gao, L. Yu, (Lucy)

**Lipid Compositions and Geographical Discrimination of 94 Geographically Authentic Wheat Samples Based on UPLC-MS with Non-Targeted Lipidomic Approach**

Foods, 10 (1) (2020), p. 10, [10.3390/foods10010010](https://doi.org/10.3390/foods10010010)

[Jung et al., 2010](#)

Y. Jung, J. Lee, J. Kwon, K.-S. Lee, D.H. Ryu, G.-S. Hwang

**Discrimination of the geographical origin of beef by <sup>1</sup>H NMR-based metabolomics**

Journal of Agricultural and Food Chemistry, 58 (19) (2010), pp. 10458-10466, [10.1021/jf102194t](https://doi.org/10.1021/jf102194t)

[Kalogiouri et al., 2020](#)

N.P. Kalogiouri, R. Aalizadeh, M.E. Dasenaki, N.S. Thomaidis

**Authentication of Greek PDO kalamata table olives: A novel non-target high resolution mass spectrometric approach**

Molecules, 25 (12) (2020), p. 2919, [10.3390/molecules25122919](https://doi.org/10.3390/molecules25122919)

[Kalogiouri et al., 2018](#)

N.P. Kalogiouri, R. Aalizadeh, N.S. Thomaidis

**Application of an advanced and wide scope non-target screening workflow with LC-ESI-QTOF-MS and chemometrics for the classification of the Greek olive oil varieties**

Food Chemistry, 256 (2018), pp. 53-61, [10.1016/j.foodchem.2018.02.101](https://doi.org/10.1016/j.foodchem.2018.02.101)



[USP Guidance on Developing and Validating Non-Targeted Methods for Adulteration Detection, 2019](#)

USP Guidance on Developing and Validating Non-Targeted Methods for Adulteration Detection, USP, FCC 1510 (2019).

[Klockmann et al., 2016](#)

S. Klockmann, E. Reiner, R. Bachmann, T. Hackl, M. Fischer

**Food fingerprinting: Metabolomic approaches for geographical origin discrimination of hazelnuts (*Corylus avellana*) by UPLC-QTOF-MS**

Journal of Agricultural and Food Chemistry, 64 (48) (2016), pp. 9253-9262, [10.1021/acs.jafc.6b04433](https://doi.org/10.1021/acs.jafc.6b04433)

[Kochem, 2017](#)

M. Kochem

**Type 1 taste receptors in taste and metabolism**

Annals of Nutrition and Metabolism, 70 (Suppl. 3) (2017), pp. 27-36, [10.1159/000478760](https://doi.org/10.1159/000478760)

[Le Boucher et al., 2015](#)

C. Le Boucher, F. Courant, A.-L. Royer, S. Jeanson, S. Lortal, G. Dervilly-Pinel, ..., B. Le Bizec

**LC-HRMS fingerprinting as an efficient approach to highlight fine differences in cheese metabolome during ripening**

Metabolomics, 11 (5) (2015), pp. 1117-1130, [10.1007/s11306-014-0769-0](https://doi.org/10.1007/s11306-014-0769-0)

[Liang et al., 2022](#)

N. Liang, S. Sun, C. Zhang, Y. He, Z. Qiu

**Advances in infrared spectroscopy combined with artificial neural network for the authentication and traceability of food**

Critical Reviews in Food Science and Nutrition, 62 (11) (2022), pp. 2963-2984, [10.1080/10408398.2020.1862045](https://doi.org/10.1080/10408398.2020.1862045)

[Liebal et al., 2020](#)

U.W. Liebal, A.N.T. Phan, M. Sudhakar, K. Raman, L.M. Blank

**Machine Learning Applications for Mass Spectrometry-Based Metabolomics**

Metabolites, 10 (6) (2020), p. 243, [10.3390/metabo10060243](https://doi.org/10.3390/metabo10060243)

[Lim et al., 2018](#)

D.K. Lim, C. Mo, J.H. Lee, N.P. Long, Z. Dong, J. Li, ..., S.W. Kwon

**The integration of multi-platform MS-based metabolomics and multivariate analysis for the geographical origin discrimination of *Oryza sativa* L**

Journal of Food and Drug Analysis, 26 (2) (2018), pp. 769-777, [10.1016/j.jfda.2017.09.004](https://doi.org/10.1016/j.jfda.2017.09.004)

[Liu et al., 2020](#)

H. Liu, Y. Zeng, X. Zhao, H. Tong

**Improved geographical origin discrimination for tea using ICP-MS and ICP-OES techniques in combination with chemometric approach**

Journal of the Science of Food and Agriculture, 100 (8) (2020), pp. 3507-3516, [10.1002/jsfa.10392](https://doi.org/10.1002/jsfa.10392)

[Llano et al., 2018](#)

S.M. Llano, A.M. Muñoz-Jiménez, C. Jiménez-Cartagena, J. Londoño-Londoño, S. Medina

**Untargeted metabolomics reveals specific withanolides and fatty acyl glycoside as tentative metabolites to differentiate organic and conventional Physalis peruviana fruits**

Food Chemistry, 244 (2018), pp. 120-127, [10.1016/j.foodchem.2017.10.026](https://doi.org/10.1016/j.foodchem.2017.10.026)

[Llorach et al., 2019](#)

R. Llorach, C. Favari, D. Alonso, M. Garcia-Aloy, C. Andres-Lacueva, M. Urpi-Sarda

**Comparative metabolite fingerprinting of legumes using LC-MS-based untargeted metabolomics**

Food Research International, 126 (2019), Article 108666,

[10.1016/j.foodres.2019.108666](https://doi.org/10.1016/j.foodres.2019.108666)

[Lopez-Sanchez et al., 2015](#)

P. Lopez-Sanchez, R.C.H. de Vos, H.H. Jonker, R. Mumm, R.D. Hall, L. Bialek, ..., J. van Duynhoven

**Comprehensive metabolomics to evaluate the impact of industrial processing on the phytochemical composition of vegetable purees**

Food Chemistry, 168 (2015), pp. 348-355, [10.1016/j.foodchem.2014.07.076](https://doi.org/10.1016/j.foodchem.2014.07.076)

[Man et al., 2021](#)

K.-Y. Man, C.-O. Chan, H.-H. Tang, N. Dong, F. Capozzi, K.-H. Wong, ..., D.-K.-W. Mok

**Mass spectrometry-based untargeted metabolomics approach for differentiation of beef of different geographic origins**

Food Chemistry, 338 (2021), Article 127847, [10.1016/j.foodchem.2020.127847](https://doi.org/10.1016/j.foodchem.2020.127847)

[Mandrile et al., 2016](#)

L. Mandrile, G. Zeppa, A.M. Giovannozzi, A.M. Rossi

**Controlling protected designation of origin of wine by Raman spectroscopy**

Food Chemistry, 211 (2016), pp. 260-267, [10.1016/j.foodchem.2016.05.011](https://doi.org/10.1016/j.foodchem.2016.05.011)

[Marti et al., 2014](#)

G. Marti, J. Boccard, F. Mehl, B. Debrus, L. Marcourt, P. Merle, ..., J.-L. Wolfender

**Comprehensive profiling and marker identification in non-volatile citrus oil residues by mass spectrometry and nuclear magnetic resonance**

Food Chemistry, 150 (2014), pp. 235-245, [10.1016/j.foodchem.2013.10.103](https://doi.org/10.1016/j.foodchem.2013.10.103)

[Martínez Bueno et al., 2018](#)

M.J. Martínez Bueno, F.J. Díaz-Galiano, Ł. Rajska, V. Cutillas, A.R. Fernández-Alba

**A non-targeted metabolomic approach to identify food markers to support discrimination between organic and conventional tomato crops**

Journal of Chromatography A, 1546 (2018), pp. 66-76, [10.1016/j.chroma.2018.03.002](https://doi.org/10.1016/j.chroma.2018.03.002)

[McGrath et al., 2018](#)

T.F. McGrath, S.A. Haughey, J. Patterson, C. Fauhl-Hassek, J. Donarski, M. Alewijn, ..., C.T. Elliott

**What are the scientific challenges in moving from targeted to non-targeted methods for food fraud testing and how can they be addressed? – Spectroscopy case study**

Trends in Food Science & Technology, 76 (2018), pp. 38-55,  
[10.1016/j.tifs.2018.04.001](https://doi.org/10.1016/j.tifs.2018.04.001)

[Medina et al., 2019](#)

S. Medina, J.A. Pereira, P. Silva, R. Perestrelo, J.S. Câmara  
**Food fingerprints – A valuable tool to monitor food authenticity and safety**  
Food Chemistry, 278 (2019), pp. 144-162, [10.1016/j.foodchem.2018.11.046](https://doi.org/10.1016/j.foodchem.2018.11.046)

[MetAlign. n.d.](#)

*MetAlign*. (n.d.). Retrieved June 18, 2021, from  
<https://www.wur.nl/en/show/MetAlign-1.htm>.

[Miano et al., 2018](#)

B. Miano, L. Righetti, R. Piro, Dall'Asta, C., Folloni, S., Galaverna, G., & Suman, M  
**Direct analysis real-time–high-resolution mass spectrometry for *Triticum* species authentication**  
Food Additives & Contaminants: Part A, 35 (12) (2018), pp. 2291-2297,  
[10.1080/19440049.2018.1520398](https://doi.org/10.1080/19440049.2018.1520398)

[Mie et al., 2014](#)

A. Mie, K.H. Laursen, K.M. Åberg, J. Forshed, A. Lindahl, K. Thorup-Kristensen, ..., S. Husted  
**Discrimination of conventional and organic white cabbage from a long-term field trial study using untargeted LC-MS-based metabolomics**  
Analytical and Bioanalytical Chemistry, 406 (12) (2014), pp. 2885-2897,  
[10.1007/s00216-014-7704-0](https://doi.org/10.1007/s00216-014-7704-0)

[MZmine 2., n.d.](#)

*MZmine 2.* (n.d.). Retrieved June 18, 2021, from <http://mzmine.github.io/>.

[Núñez et al., 2021](#)

N. Núñez, C. Martínez, J. Saurina, O. Núñez  
**High-performance liquid chromatography with fluorescence detection fingerprints as chemical descriptors to authenticate the origin, variety and roasting degree of coffee by multivariate chemometric methods**  
Journal of the Science of Food and Agriculture, 101 (1) (2021), pp. 65-73,  
[10.1002/jsfa.10615](https://doi.org/10.1002/jsfa.10615)

[Council Regulation \(EC\) No 834/2007, 2007](#)

*Council Regulation (EC) No 834/2007*, (2007) (testimony of Official Journal of the European Union). <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=celex%3A32007R0834>.

[Regulation \(EU\) No 1151/2012 of the European Parliament and of the Council of 21 November 2012 on Quality Schemes for Agricultural Products and Foodstuffs, 2012](#)

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

[Oiv, 2021](#)

OIV - International Organisation of Vine and Wine. (n.d.). Retrieved June 18, 2021 from <https://www.oiv.int/en/>.

[Oliveira et al., 2015](#)

G. Oliveira, M. Alewijn, R. Boerrigter-Eenling, S. van Ruth  
**Compositional signatures of conventional, free range, and organic pork meat using fingerprint techniques**  
Foods, 4 (4) (2015), pp. 359-375, [10.3390/foods4030359](https://doi.org/10.3390/foods4030359)

[Pan et al., 2020](#)

H. Pan, C. Yao, S. Yao, W. Yang, W. Wu, D. Guo  
**A metabolomics strategy for authentication of plant medicines with multiple botanical origins, a case study of Uncariae Rammulus Cum Uncis**  
Journal of Separation Science, 43 (6) (2020), pp. 1043-1050, [10.1002/jssc.201901064](https://doi.org/10.1002/jssc.201901064)

[Pardo-Mates et al., 2017](#)

N. Pardo-Mates, A. Vera, S. Barbosa, M. Hidalgo-Serrano, O. Núñez, J. Saurina, ..., L. Puignou  
**Characterization, classification and authentication of fruit-based extracts by means of HPLC-UV chromatographic fingerprints, polyphenolic profiles and chemometric methods**  
Food Chemistry, 221 (2017), pp. 29-38, [10.1016/j.foodchem.2016.10.033](https://doi.org/10.1016/j.foodchem.2016.10.033)

[Perini et al., 2018](#)

M. Perini, M. Paolini, F. Camin, G. Appendino, F. Vitulo, E. De Combarieu, ..., R. Pace  
**Combined use of isotopic fingerprint and metabolomics analysis for the authentication of saw palmetto (*Serenoa repens*) extracts**  
Fitoterapia, 127 (2018), pp. 15-19, [10.1016/j.fitote.2018.04.011](https://doi.org/10.1016/j.fitote.2018.04.011)

[Popping et al., 2017](#)

B. Popping, E. De Dominicis, M. Dante, M. Nocetti  
**Identification of the geographic origin of Parmigiano Reggiano (P.D.O.) cheeses deploying non-targeted mass spectrometry and chemometrics**  
Foods, 6 (2) (2017), p. 13, [10.3390/foods6020013](https://doi.org/10.3390/foods6020013)

[Potočnik et al., 2020](#)

D. Potočnik, M. Nečemer, I. Perišić, M. Jagodic, D. Mazej, F. Camin, ..., N. Ogrinc  
**Geographical verification of Slovenian milk using stable isotope ratio, multi-element and multivariate modelling approaches**  
Food Chemistry, 326 (2020), Article 126958, [10.1016/j.foodchem.2020.126958](https://doi.org/10.1016/j.foodchem.2020.126958)

[Primrose et al., 2010](#)

S. Primrose, M. Woolfe, S. Rollinson  
**Food forensics: Methods for determining the authenticity of foodstuffs**

Trends in Food Science & Technology, 21 (12) (2010), pp. 582-590,  
[10.1016/j.tifs.2010.09.006](https://doi.org/10.1016/j.tifs.2010.09.006)

[Prosser and Hebert, 2017](#)

S.W.J. Prosser, P.D.N. Hebert

**Rapid identification of the botanical and entomological sources of honey using DNA metabarcoding**

Food Chemistry, 214 (2017), pp. 183-191, [10.1016/j.foodchem.2016.07.077](https://doi.org/10.1016/j.foodchem.2016.07.077)

[Putri et al., 2019](#)

S.P. Putri, T. Irifune, Yusianto, & Fukusaki, E

**GC/MS based metabolite profiling of Indonesian specialty coffee from different species and geographical origin**

Metabolomics, 15 (10) (2019), p. 126, [10.1007/s11306-019-1591-5](https://doi.org/10.1007/s11306-019-1591-5)

[Robert Parker Wine, 2020](#)

*Robert Parker Wine Advocate*. (2020).

<https://www.robertparker.com/resources/vintage-chart>.

[Robson et al., 2021](#)

K. Robson, M. Dean, S. Haughey, C. Elliott

**A comprehensive review of food fraud terminologies and food fraud mitigation guides**

Food Control, 120 (2021), Article 107516, [10.1016/j.foodcont.2020.107516](https://doi.org/10.1016/j.foodcont.2020.107516)

[Rocchetti et al., 2020](#)

G. Rocchetti, G.P. Braceschi, L. Odello, T. Bertuzzi, M. Trevisan, L. Lucini

**Identification of markers of sensory quality in ground coffee: An untargeted metabolomics approach**

Metabolomics, 16 (12) (2020), p. 127, [10.1007/s11306-020-01751-6](https://doi.org/10.1007/s11306-020-01751-6)

[Rocchetti et al., 2018](#)

G. Rocchetti, M. Gatti, L. Bavaresco, L. Lucini

**Untargeted metabolomics to investigate the phenolic composition of Chardonnay wines from different origins**

Journal of Food Composition and Analysis, 71 (2018), pp. 87-93,

[10.1016/j.jfca.2018.05.010](https://doi.org/10.1016/j.jfca.2018.05.010)

[Rocchetti et al., 2018](#)

G. Rocchetti, L. Lucini, A. Gallo, F. Masoero, M. Trevisan, G. Giuberti

**Untargeted metabolomics reveals differences in chemical fingerprints between PDO and non-PDO Grana Padano cheeses**

Food Research International, 113 (2018), pp. 407-413, [10.1016/j.foodres.2018.07.029](https://doi.org/10.1016/j.foodres.2018.07.029)

[Roullier-Gall et al., 2014](#)

C. Roullier-Gall, M. Lucio, L. Noret, P. Schmitt-Kopplin, R.D. Gougeon

**How subtle Is the “Terroir” Effect? Chemistry-related signatures of two “Climats de Bourgogne”**

PLoS ONE, 9 (5) (2014), p. e97615

[Roullier-Gall et al., 2014](#)

C. Roullier-Gall, M. Witting, R.D. Gougeon, P. Schmitt-Kopplin  
**High precision mass measurements for wine metabolomics**  
Frontiers in Chemistry, 2 (102) (2014), pp. 1-9, [10.3389/fchem.2014.00102](#)

[Rubert et al., 2014](#)

J. Rubert, O. Lacina, C. Fauhl-Hassek, J. Hajslova  
**Metabolic fingerprinting based on high-resolution tandem mass spectrometry: A reliable tool for wine authentication?**  
Analytical and Bioanalytical Chemistry, 406 (27) (2014), pp. 6791-6803,  
[10.1007/s00216-014-7864-y](#)

[Saint-Lary et al., 2016](#)

L. Saint-Lary, C. Roy, J.-P. Paris, J.-F. Martin, O.P. Thomas, X. Fernandez  
**Metabolomics for the authentication of natural extracts used in flavors and fragrances: the case study of violet leaf absolutes from *Viola odorata***  
Chemistry & Biodiversity, 13 (6) (2016), pp. 737-747, [10.1002/cbdv.201500230](#)

[Sales et al., 2019](#)

C. Sales, T. Portolés, L.G. Johnsen, M. Danielsen, J. Beltran  
**Olive oil quality classification and measurement of its organoleptic attributes by untargeted GC-MS and multivariate statistical-based approach**  
Food Chemistry, 271 (2019), pp. 488-496, [10.1016/j.foodchem.2018.07.200](#)

[Schueuermann et al., 2016](#)

C. Schueuermann, B. Khakimov, S.B. Engelsen, P. Bremer, P. Silcock  
**GC-MS metabolite profiling of extreme southern pinot noir wines: Effects of vintage, barrel maturation, and fermentation dominate over vineyard site and clone selection**  
Journal of Agricultural and Food Chemistry, 64 (11) (2016), pp. 2342-2351,  
[10.1021/acs.jafc.5b05861](#)

[Seguin, 1986](#)

G. Seguin  
**'Terroirs' and pedology of wine growing**  
Experientia, 42 (8) (1986), pp. 861-873, [10.1007/BF01941763](#)

[Senizza et al., 2019](#)

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 Research International, 126 (2019), Article 108584,  
[10.1016/j.foodres.2019.108584](#)

[Shao et al., 2019](#)

B. Shao, H. Li, J. Shen, Y. Wu  
**Nontargeted detection methods for food safety and integrity**  
Annual Review of Food Science and Technology, 10 (1) (2019), pp. 429-455,  
[10.1146/annurev-food-032818-121233](#)

[Shears, 2010](#)

P. Shears

**Food fraud – a current issue but an old problem**

British Food Journal, 112 (2) (2010), pp. 198-213, [10.1108/00070701011018879](https://doi.org/10.1108/00070701011018879)

[Silva et al., 2021](#)

B. Silva, A.C.O. Costa, S.S. Tchewonpi, J. Bönick, G. Huschek, L.V. Gonzaga, ..., 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 Research International, 141 (2021), Article 109991, [10.1016/j.foodres.2020.109991](https://doi.org/10.1016/j.foodres.2020.109991)

[Souard et al., 2018](#)

F. Souard, C. Delporte, P. Stoffelen, E.A. Thévenot, N. Noret, B. Dauvergne, ..., C. Stévigny

**Metabolomics fingerprint of coffee species determined by untargeted-profiling study using LC-HRMS**

Food Chemistry, 245 (2018), pp. 603-612, [10.1016/j.foodchem.2017.10.022](https://doi.org/10.1016/j.foodchem.2017.10.022)

[Sumar and Ismail, 1995](#)

S. Sumar, H. Ismail

**Adulteration of foods – past and present**

Nutrition & Food Science, 95 (4) (1995), pp. 11-15, [10.1108/00346659510088663](https://doi.org/10.1108/00346659510088663)

[Tenori et al., 2018](#)

L. Tenori, C. Santucci, G. Meoni, V. Morrocchi, G. Matteucci, C. Luchinat

**NMR metabolomic fingerprinting distinguishes milk from different farms**

Food Research International, 113 (2018), pp. 131-139, [10.1016/j.foodres.2018.06.066](https://doi.org/10.1016/j.foodres.2018.06.066)

[Tianniam et al., 2009](#)

S. Tianniam, T. Bamba, E. Fukusaki

**Non-targeted metabolite fingerprinting of oriental folk medicine *Angelica acutiloba* roots by ultra performance liquid chromatography time-of-flight mass spectrometry**

Journal of Separation Science, 32 (13) (2009), pp. 2233-2244, [10.1002/jssc.200900121](https://doi.org/10.1002/jssc.200900121)

[Tianniam et al., 2008](#)

S. Tianniam, L. Tarachiwin, T. Bamba, A. Kobayashi, E. Fukusaki

**Metabolic profiling of *Angelica acutiloba* roots utilizing gas chromatography–time-of-flight–mass spectrometry for quality assessment based on cultivation area and cultivar via multivariate pattern recognition**

Journal of Bioscience and Bioengineering, 105 (6) (2008), pp. 655-659, [10.1263/jbb.105.655](https://doi.org/10.1263/jbb.105.655)

[Trygg and Wold, 2002](#)

J. Trygg, S. Wold



### **Orthogonal projections to latent structures (O-PLS)**

Journal of Chemometrics, 16 (3) (2002), pp. 119-128, [10.1002/cem.695](https://doi.org/10.1002/cem.695)

#### [Uhlig et al., 2021](#)

S. Uhlig, K. Nichani, M. Stoyke, P. Gowik

#### **Validation of binary non-targeted methods**

Mathematical framework and experimental designs., 15 (2021),

[10.1101/2021.01.19.427235](https://doi.org/10.1101/2021.01.19.427235)

#### [Uttl et al., 2019](#)

L. Uttl, K. Hurkova, V. Kocourek, J. Pulkrabova, M. Tomaniova, J. Hajslova

#### **Metabolomics-based authentication of wines according to grape variety**

Czech Journal of Food Sciences, 37 (4) (2019), pp. 239-245

[10.17221/82/2019-CJFS](https://doi.org/10.17221/82/2019-CJFS)

#### [Vestner, 2016](#)

J. Vestner

#### **New chemometric approaches to non-targeted GCMS fingerprinting analysis of wine volatiles**

Université de Bordeaux, Hochschule Geisenheim University (2016)

#### [Workflow4Metabolomics, n.d.](#)

*Workflow4Metabolomics*. (n.d.). Retrieved June 18, 2021, from <https://workflow4metabolomics.usegalaxy.fr/>.

#### [XCMS, n.d.](#)

*XCMS*. (n.d.). Retrieved June 18, 2021, from

[https://xcmsonline.scripps.edu/landing\\_page.php?pgcontent=mainPage](https://xcmsonline.scripps.edu/landing_page.php?pgcontent=mainPage).

#### [Xiao et al., 2018](#)

R. Xiao, Y. Ma, D. Zhang, L. Qian

#### **Discrimination of conventional and organic rice using untargeted LC-MS-based metabolomics**

Journal of Cereal Science, 82 (2018), pp. 73-81, [10.1016/j.jcs.2018.05.012](https://doi.org/10.1016/j.jcs.2018.05.012)

#### [Xing et al., 2019](#)

R.-R. Xing, N. Wang, R.-R. Hu, J.-K. Zhang, J.-X. Han, Y. Chen

#### **Application of next generation sequencing for species identification in meat and poultry products: A DNA metabarcoding approach**

Food Control, 101 (2019), pp. 173-179, [10.1016/j.foodcont.2019.02.034](https://doi.org/10.1016/j.foodcont.2019.02.034)

#### [Xu et al., 2020](#)

L. Xu, Z. Xu, S. Kelly, X. Liao

#### **Integrating untargeted metabolomics and targeted analysis for not from concentrate and from concentrate orange juices discrimination and authentication**

Food Chemistry, 329 (2020), Article 127130, [10.1016/j.foodchem.2020.127130](https://doi.org/10.1016/j.foodchem.2020.127130)

#### [Xu et al., 2020](#)

L. Xu, Z. Xu, I. Strashnov, X. Liao

**Use of information dependent acquisition mass spectra and sequential window acquisition of all theoretical fragment-ion mass spectra for fruit juices metabolomics and authentication**

Metabolomics, 16 (7) (2020), p. 81, [10.1007/s11306-020-01701-2](https://doi.org/10.1007/s11306-020-01701-2)

[Zhang et al., 2018](#)

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**

Journal of Agricultural and Food Chemistry, 66 (30) (2018), pp. 8199-8208,

[10.1021/acs.jafc.8b01682](https://doi.org/10.1021/acs.jafc.8b01682)

[Zhao et al., 2017](#)

S.-Y. Zhao, Z.-L. Liu, Y.-S. Shu, M.-L. Wang, D. He, Z.-Q. Song, ..., Y.-Y. Liu

**Chemotaxonomic classification applied to the identification of two closely-related citrus TCMs Using UPLC-Q-TOF-MS-based metabolomics**

Molecules, 22 (10) (2017), p. 1721, [10.3390/molecules22101721](https://doi.org/10.3390/molecules22101721)

[Ziółkowska et al., 2016](#)

A. Ziółkowska, E. Wąsowicz, H.H. Jeleń

**Differentiation of wines according to grape variety and geographical origin based on volatiles profiling using SPME-MS and SPME-GC/MS methods**

Food Chemistry, 213 (2016), pp. 714-720, [10.1016/j.foodchem.2016.06.120](https://doi.org/10.1016/j.foodchem.2016.06.120)