Food Control 145 (2023) 109497

Contents lists available at ScienceDirect

Food Control

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

Qualitative and quantitative food authentication of oregano using NGS and NMR with chemometrics

Friedemann Flügge [a,b], Tim Kerkow [b], Paulina Kowalski [b], Josephine Bornhöft [b], Eva Seemann [b], Marina Creydt [c], Burkhard Schütze [b], Ulrich L. Günther [a,*]

[a] University of Lübeck, Institute of Chemistry and Metabolomics, Ratzeburger Allee 160, 23562, Lübeck, Germany
[b] LADR GmbH Medizinisches Versorgungszentrum Dr. Kramer & Kollegen, Lauenburger Str. 67, 21502, Geesthacht, Germany
[c] Hamburg School of Food Science – Institute of Food Chemistry, University of Hamburg, Grindelallee 117, 20146, Hamburg, Germany

ARTICLE INFO

Keywords:
Food fraud
Metabolomics
NMR fingerprinting
Metabarcoding
Spices

ABSTRACT

Culinary herbs and spices represent increasingly popular food products, especially in an environment of increased consciousness of food quality. Health-conscious consumers expect clean and reliable labels for food products. For spices this implies significant challenges considering complex and long supply chains, often with unknown sources or intermediates. A recent report by the EU Joint Research Centre identified oregano as one of the most vulnerable spices for adulterations. The analysis of herbs is complex, traditionally using visual inspection, and more recently genetic and chemical analysis. Here, metabarcoding by next generation sequencing (NGS) has been used to identify the nature of oregano products and possible adulterations. In parallel, metabolite extracts are generated and analysed on a 400 MHz Nuclear Magnetic Resonance (NMR) spectrometer. NMR metabolic profiles in conjunction with multivariate statistics was used to identify oregano species and to identify and quantify the level of adulteration with other plant products.

1. Introduction

Food fraud has been defined as "any suspected intentional action by businesses or individuals for the purpose of deceiving purchasers and gaining undue advantage therefrom, in violation of the rules" referred to in Article 1 (2) of Regulation (EU) 2017/625 (the agri-food chain legislation) (Regulation (EU) 2017/625 of the European Parliament and of the Council, 2017). Currently, the EU imports, more than 300.000 tons of spices per annum, often with long and complicated supply chains, leaving many opportunities for fraudulent interaction. A report by the Joint Research Centre (JRC), a European Commission's science and knowledge service, has identified oregano as the most vulnerable of the six most frequently used spices (European Commission. Joint Research Centre., 2021). There are different types of adulteration, including the addition of additives, such as dyes or flavouring substances, other botanical plant products, or different parts of the same plant. On the market we find mainly three species of the genus *Origanum*. The most frequently grown species are *Origanum onites* (also called pot marjoram, mainly grown in Turkey and Greece) and *O. vulgare* (common oregano, mainly grown in Greece). The third, *O. majorana* (common name: sweet marjoram) represents a similar plant that must not be present in dried oregano according to the ISO standard (ISO 7925:1999, 1999). The European Spice Association (ESA) is a non-profit umbrella organization of the European spice industry which has also defined maximum levels of extraneous material (European Spice Association (ESA), 2018). For herbs such as Oregano the maximum level is 2% (w/w).

The JRC identified the most common adulterations of foreign botanical material in oregano as olive leaves in 27%, sweet marjoram in 4%, and myrtle leaves in 1% of tested oregano samples. Overall, potential adulterations were detected in 48% of the tested samples.

This situation indicates a clear need for new reliable methods to identify and quantify additions in spices. To identify the nature of adulteration by other botanical products, metabarcoding with the help

Abbreviations: ESA, European Spice Association; *J*-RES, *J*-resolved spectroscopy; LC-MS, liquid chromatography-mass spectrometry; NGS, next generation sequencing; NMR, nuclear magnetic resonance; *O. majorana*, *Origanum majorana*; *O. onites*, *Origanum onites*; *O. vulgare*, *Origanum vulgare*; PCA, principle component analysis; PLS, partial least squares; OPLS, orthogonal partial least squares; OPLS-DA, orthogonal partial least squares discriminant analysis; RMSE, root mean square error.

* Corresponding author.
E-mail address: ulrich.guenther@uni-luebeck.de (U.L. Günther).

https://doi.org/10.1016/j.foodcont.2022.109497
Received 3 August 2022; Received in revised form 6 October 2022; Accepted 4 November 2022
Available online 10 November 2022

of NGS is a powerful and increasingly used method (Haynes et al., 2019; Prosser & Hebert, 2017). NGS is extremely sensitive and there are now databases available which cover common types of plants with a very high specificity. However, additions of endogenous non-aromatic material, such as plant stems, or exogenous material not containing DNA cannot be detected. Furthermore, DNA read percentages can and should not be related to plant weight percentages, since the number of homologous chromosomes, gene expression levels as well as extraction efficiency may differ for different plant species. The ESA therefore advises in a recent white paper that NGS must be combined with an orthogonal analytical method for spice authentication (European Spice Association (ESA), 2021).

This can either be achieved by NMR or mass spectrometry. LC-MS has the advantage of covering a larger number of metabolites due to the high sensitivity and is commonly available in nutritional chemistry labs. The analysis is semi-quantitative and usually uses a form of fingerprinting without assigning spectral features. NMR has also been widely used for food quality control and is currently gaining popularity. Site-specific natural isotope fractionation nuclear magnetic resonance (SNIF-NMR) detecting D/H isotope ratios in alcohol has been a substantial commercial success (Martin et al., 1982). NMR metabolomics approaches are commercially available and have become increasingly popular, especially for quality control of fruit juices (Spraul et al., 2009). Other relevant applications of high resolution NMR on food systems comprise among others analytics of honeys, edible oils and wines (Godelmann et al., 2013; Guyader et al., 2018; Spiteri et al., 2015). These applications have in common that a 400 MHz spectrometer provides sufficient resolution and sensitivity, allowing a relatively low-cost option for food analysis. The advantage of the NMR approach is the high reproducibility of spectra with an error rate well below 1% and the ability to assign NMR signals to specific compounds, thus enabling a profiling approach. This is particularly important for certified tests which need to go beyond fingerprinting allowing the identification and quantification of specific markers as quality controls.

Here we have used a combined approach to identify and quantify adulterations in oregano products, mainly based on NGS and NMR. NGS was used to identify oregano species and the nature of added plant material, whereas NMR was used for initial fingerprinting to identify types of oregano, geographical origin, and the addition of other plants, such as olive leaves and marjoram. We used LC-MS to aid assignment of NMR signals, allowing us to assign approximately 80% of the ^{1}H NMR spectrum of oregano extracts.

2. Materials and methods

2.1. Chemicals

Deuterated solvents methanol-d_4 (99.8%) and D_2O (99.8%) were purchased from Cortecnet (Les Ulis, France). 3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid sodium salt (98.0%) was purchased from Eurisotop (Saint-Aubin Cedex, France). Sodium dihydrogen phosphate monohydrate (>98%) was purchased from Carl Roth (Karlsruhe, Germany). Disodium hydrogen phosphate (>99%), and formic acid (>98%) were purchased from Sigma-Aldrich (Merck KGaA, Darmstadt, Germany). Ultrapure water, methanol and acetonitrile (CH_3CN, LC-MS grade) were purchased from VWR (Radnor, Pennsylvania, USA).

2.2. Oregano, marjoram, and olive leaf samples

The samples were provided by or purchased from 33 different European suppliers as dried and rubbed or ground plant material. In total 92 oregano, 38 marjoram, and 2 olive leaf samples were included in this study. The declared regional origin of the individual samples is reported in Table 1 and Table 2.

Table 1
Genetic characterization and regional origin of the 92 oregano samples analysed in this work. Displayed is the respective number of sample (origins listed as provided by supplier).

Origanum species		Origin		Suspicious NGS result	
O. vulgare	27	Germany	3	–	–
		Chile	3	*O. majorana*	3
		Greece	11	-	–
		Italy	3	-	–
		unknown	7	*O. majorana*	3
O. onites	63	Greece	1	-	–
O. vulgare		Turkey	37	*O. majorana*	8
				Myrtus communis	4
				Olea europaea	1
				Salvia sp.	2
				Cistus creticus	2
				Arbutus andrachne	3
		Kosovo	1	*Salvia* sp.	1
				Cistus creticus	1
				Corylus avellana	1
		unknown	24	*O. majorana*	7
				Olea europaea	2
				Rhus sp.	1
O. onites	2	Turkey	2	–	–

Table 2
Sample number and regional origin of the marjoram, olive leaf, myrtle leaf, hazelnut leaf and pink rock-rose leaf samples analysed in this work.

Species	Origin		
O. majorana	38	Egypt	28
		Germany	6
		Netherlands	1
		unknown	3
Olea europaea	2	Turkey	1
		unknown	1
Myrtus communis	1	Turkey	1
Corylus avellana	1	Unknown	1
Cistus creticus	2	Turkey	1
		Unknown	1

2.3. Genetic characterization by NGS

Plant samples were prepared according to the Thermo Scientific NGS Food Authenticity Workflow. All materials and equipment were purchased from Thermo Fisher Scientific, Waltham, Massachusetts, USA. Total DNA was extracted from 1 g of plant material using GMO extraction kit (Ref: 4466336) according to the manual's instructions. The concentration of the extracted DNA was assessed with a Qubit™ 4 Fluorometer (Ref: Q33226) and dsDNA BRAssay Kit (Ref: Q32853) and adjusted to a final concentration of 2 ng/μL. DNA libraries were prepared using SGS™ All Species Plant DNA Analyser Kit I (Ref A38456) & II (Ref A38457) with the help of a SimpliAmp™ Thermal Cycler (Ref: A24811) and E-Base™ gel electrophoresis device (Ref: EBD03). Sequencing was carried out on an Ion GeneStudio™ S5 platform (Ref: A39513) using Ion 510™, Ion 520™ and Ion 530™ Food Protection Kit (Ref: A39476). Data were processed using the SGS™ All Species ID Software application by sequence comparison with the plant reference database covering 5,432 species entries.

2.4. Sample preparation for NMR metabolomics

After comparing different extraction methods and different choices of solvent, we decided to prepare samples in one step by extracting with deuterated solvents. As a mixture of MeOD and D_2O showed considerably more peaks than just D_2O or MeOD we chose a 1:1 mixture. Specifically, 50 mg (±1%) of the dried plant material was extracted with 1 mL of a 1:1 mixture of 100 mM sodium phosphate buffer (pH* 7.0) and MeOD containing the reference standard TMSP-2,2,3,3-d_4 (1 mM) using

a Precellys® 24 tissue homogenizer (Bertin Technologies SAS, France) with five Zirconia beads for 3 × 30 s at 6.500×*g*. Insoluble material was removed by centrifugation (10 min, 17.000×*g*) and filtration (Chromafil Xtra RC 0.2 μm syringe filters, Macherey-Nagel, Germany). 550 μL of the extract were transferred into 5 mm, 7" High Throughput NMR tubes (Wilmad, USA).

Commercially available reference compounds were purchased from Sigma-Aldrich (Merck KGaA, Darmstadt, Germany). Reference spectra were acquired with a concentration of 10 mM in pH adjusted solutions, respectively.

2.5. NMR experiments

Initially we compared 1D-^{1}H-NOESY and 1D projections of 2D *J*-resolved (*J*-RES) spectra and found considerable signal overlap causing problems with the assignment of the 1D-NOESYs. Of each extract a 1D ^{1}H NOESY and a 2 D *J*-RES was acquired at 298 K on a 400 MHz Bruker FoodScreener spectrometer (Bruker Biospin, Rheinstetten, Germany). Bruker standard pulse sequences noesygppr1d and jresgpprqf were used. The 1D NOESY experiment was recorded with 64 k data points, a sweep width of 8232.68 Hz, 32 scans and a relaxation delay of 4 s. The *J*-RES was recorded with 8k data points in the direct and 40 data points in the indirect dimension, a sweep width of 6684.49 Hz, 4 scans, and a relaxation delay of 2s. MeOD was used for the field-lock.

For identification of individual metabolites, we additionally used a set of 2D heteronuclear ^{13}C^{1}H experiments including HSQC (hsqcetgpsisp.2), multiplicity-edited HSQC (hsqcedetgp), HMBC (hmbcgplpndqf) and HSQC-TOCSY (hsqcdietgpsi).

2.6. Chemical analysis by LC-MS to guide assignment of NMR signals

Polar extracts were generated as described above with the exception that non-deuterated solvents (LC-MS grade) were used. Extracts were analysed in positive ion mode on a Bruker Maxis II mass spectrometer (maXis, UHR-TOF, Bruker Daltonics, Bremen, Germany) equipped with an electrospray ion source. HPLC separation was achieved with an Agilent 1290 Infinity II LC system (Agilent Technologies, Santa Clara, CA USA) equipped with an Eclipse Plus-C_{18} column (2.1 × 50 mm, 1.8 μm, Agilent Technologies, Santa Clara, CA USA) in reversed-phase chromatography. 0.1% formic acid was added to the eluents (water/CH_3CN). The separation method was: 0–1 min, CH_3CN (1%), 1–12 min, CH_3CN (1–35%), 12–14 min, CH_3CN (35–99%), 16–18 min, CH_3CN (99%). The flow rate was 200 μL/min.

MS conditions were as follows: mass range 50–1000 *m/z* with rolling average of 2 and spectral rate of 2 Hz. Capillary voltage was 4.5 kV and end plate offset 500 V. The drying gas flow (N_2) was 8.0 L/min, the nebulizer pressure 2.0 Bar, and drying gas temperature 220 °C. The analysis was calibrated with sodium formate cluster ions, directly injected into the source at the start of the measurement run.

The open-source software MZmine (Pluskal et al., 2010) was used for processing of LC-MS data and identification of compounds. First step was mass detection using the centroid algorithm with a cut-off of 15.000 to remove noise. Chromatograms for each mass were created using the ADAP Chromatogram Builder module with an *m/z* tolerance of 0.001/*m/z*, a minimum group size of 5 scans and a group intensity threshold of 10000. Chromatograms were deconvoluted using a wavelets algorithm and deisotoped with an *m/z* tolerance of 0.001 and a retention time tolerance of 0.01 min. For identification of compounds a custom made database based on entries assigned to Origanum species in FooDB (FooDB, 2022) was used. For this, monoisotopic masses were extracted from FooDB and the following ion features generated using an in-house code: $[M+H]^+$, $[M+Na]^+$, $[M+K]^+$, $[M + CH_3CN + H]^+$, $[M + CH_3CN + Na]^+$, $[M+2Na–H]^+$. The identification *m/z* tolerance was 0.001.

2.7. Pre-analytical preparation of NMR data

NMR spectra were (pre-)processed in a batch-mode using NMRLab and MetaboLab software packages (Günther et al., 2000; Ludwig & Günther, 2011) within Matlab (version 9, The Mathworks). 2D *J*-RES spectra were processed with zero filling to 8 k real data points in the direct dimension and 128 points in the indirect dimension with application of a sine apodization function. The spectra were then tilted by 45° and symmetrized, and sum projections calculated along the indirect dimension. Each spectrum was calibrated by setting the TMSP-d_4 signal to 0 ppm.

In MetaboLab the 1D *J*-RES projections were baseline corrected using the same baseline points for all spectra after realigning all to the first spectrum. The spectral regions around the solvent signals at 3.32 ppm (MeOD) and 4.86 ppm (HDO) as well as the TMSP-d_4 signal, and its satellites, and spectral regions without signals (<0 ppm, >10 ppm) were excluded. No binning of the data was performed and ca. 4750 data points per spectrum were finally used for uni- and multivariate analysis. All multi-variate analyses were carried out using the *J*-RES and not the 1D-NOESY data because the former had a much better baseline, no need of phase correction and significantly reduced spectra overlap. The spectra underwent a probabilistic quotient normalization and auto- or Pareto scaling procedure.

2.8. Virtual and physical spiking

To estimate the amount of adulterant in the suspicious oregano samples, several physically spiked samples were prepared by direct weighing of the respective plant material for the extraction. Only non-suspicious oregano samples were randomly selected for spiking. The total amount of each spiked sample was 50 mg ± 1%. Physically spiked samples contain different ratios of oregano and foreign plant material in the range of 95/5 to 50/50 for two-component mixes, or 5/5/90 to 30/30/40 for three-component mixes. Numerically spiked samples were generated by first defining reference samples and then by adding processed spectra of references and adulterants.

PCA allowed straightforward discrimination of *O. vulgare* and *O. onites* samples. Reference spectra representative for each group were selected, and downstream models built, for each of these groups. Inclusion criteria were (i) non-suspicious NGS identification as pure *O. vulgare* or *O. onites/O.vulgare* without adulterants, (ii) small Hotelling's T^2 and Q residuals in a PCA model for each group. For numerical spiking of variable amounts of adulterant spectra into reference spectra scaling factors were applied for both keeping the sum always equal to one ($x_{adulterant} * spc_{adulterant} + x_{oregano} * spc_{oregano}$, where $x_{adulterant} + x_{oregano} = 1$). Three-component mixes were generated accordingly such that the sum of weighting factors x_i was always 1.

2.9. Multivariate NMR data analysis

All multivariate analyses were performed using PLS_Toolbox (Eigenvector Research, Inc., Manson, Washington) within Matlab. Models were assessed by venetian blinds cross-validation (CV) with ten random splits and on average 10% left-out data. Initial evaluation of ^{1}H NMR data was performed by PCA. Combined Hotelling's T^2 and Q residuals were used to identify outliers in the data set. These outliers, as well as samples identified as suspicious from genetic analysis, were excluded from further analysis.

Supervised models, partial least squares (PLS) regression and PLS discriminant analysis (PLS-DA) were constructed as orthogonalised analyses (OPLS/OPLS-DA: orthogonal option in PLS_Toolbox), which contain all the y-variance (indicators of categories related to observation groups) capturing direction in the first weight and loading. As indicators of the quality of discrimination models we report the sensitivity (true positive rate) and specificity (true negative rate) of the calibration (cal) and CV data sets following cross-validation. Further indicators for the

quality of supervised models are the root mean square error (RMSE) calculated as the distance between real and estimated y variable and the goodness of fit measure R-squared (R^2). These statistics were calculated for each class (DA models) or column in Y data (regression models) for the calibration (cal), CV and test data (pred). RMSEs close to 0 and R^2 close to 1 indicate excellent performance of models.

2.10. Univariate NMR data analysis and quantification of metabolites

To identify deviating positions in a spectrum of a suspicious sample, a one-sided two-sample *t*-test was performed against the set of authentic reference samples of the respective group the suspicious sample belongs to. Suspicious samples were identified by suspicious results from genetic analysis and/or suspicious combined Hotelling's T^2 and Q residuals of a PCA model. Significantly deviating positions in the spectrum were identified with a significance level of $\alpha = 0.01$, Bonferroni corrected with m = 4753 (equal to the number of data points).

To estimate concentrations of metabolites Chenomx NMR Suite 8.5, professional edition (Chenomx Inc., Edmonton, Canada) was used. A compound library was set up with *J*-RES spectra of commercially available metabolites and known concentrations. Individual compounds were simultaneously fitted to spectra of oregano extracts to estimate concentrations. Significant differences of four groups: (1) *O. onites,* Turkey (n = 11), 2) *O. vulgare,* Greece (n = 10), 3) *O. majorana*, Egypt (n = 10), and 4) *O. vulgare (*suspicious *O. majorana*), Chile/unknown (n = 5) were determined by one-way ANOVA.

3. Results

The selected analysis workflow is summarised in Fig. 1. All samples were first screened using NGS to authenticate the analytical material. In addition, the samples were measured with a 400 MHz NMR spectrometer. The results of these two orthogonal strategies were evaluated using univariate and multivariate calculations. In addition, mixtures were analysed to be able to estimate the ratios in which adulterations can be detected.

3.1. NMR method development

To establish a robust protocol for the analysis of oregano samples, we have optimised the extraction to maximise the number of compounds seen in the NMR spectra. Best results were obtained using a Precellys® homogenizer with a mixture of methanol and water (1:1 v/v) as extraction solvents. By employing deuterated solvents, we directly obtain NMR samples after a short filtration with minimal water suppression. Initially, 1D-NOESY spectra were acquired since this type of one-dimensional spectrum is commonly used for established food-screening protocols (Spraul et al., 2009). However, considering the amount of spectral crowding in these extracts at 400 MHz we chose one-dimensional projections of *J*-resolved spectra (*J*-RES) which considerably reduce overlap of signals, yield a much better baseline and omit the need for phase correction, which is important for subsequent automated statistical analyses. Using routines of the NMRLab software, *J*-RES spectra were processed and aligned for all subsequent analyses.

Fig. 2 shows a typical *J*-RES spectrum of an oregano and a marjoram sample with assignments for important metabolites. Initial assignments were guided by the FooDB database (FooDB, 2022), using a combination of LC-MS and NMR of a fractionated sample to obtain NMR assignments with the help of a range of NMR spectra, including $^{1}H–^{13}C$-HSQC, $^{1}H–^{13}C$-HMBC, $^{1}H–^{13}C$-HSQC-TOCSY. All metabolites shown in Fig. 2 were also confirmed by comparing with spectra of commercially available compounds. In the aliphatic region we detected sugars, amino acids, and terpenes. Additionally, a range of polyphenols, terpenes and flavonoide-glycols including rosmarinic acid, arbutin, salvianolic acid B, luteolin-7-glucoside, thymol and carvacrol have been identified in the aromatic region. Currently, we have 25–30 compounds with confirmed assignments and several additional candidate assignments without confirmation. Our analysis confirms previously published assignments of, for example, rosmarinic acid, salvianolic acid B, alanine, quinic acid, acetic acid, malic acid, aspartic acid and glucose (Mandrone et al., 2021). We believe that assignments published by Mandrone et al. for thymol and carvacrol have been interchanged (Fig. S1a). The same authors have proposed p-cymol and apigenin as markers for *O. onites*. Our analysis confirms the importance of these signals although the assignment to p-cymol and apigenin cannot be correct (Fig. S1b) (Mandrone

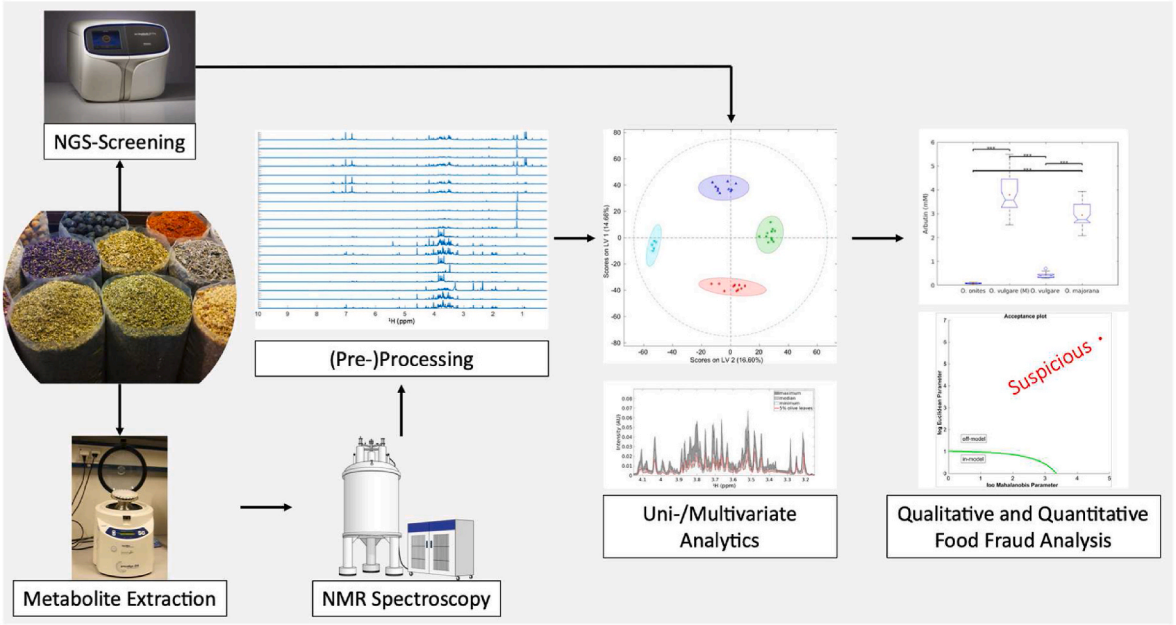

Fig. 1. Workflow describing the analytical procedure to determine and quantify food fraud in spices. Metabarcoding by NGS is used to screen and authenticate the analytical material. In parallel, metabolite extracts are generated and analysed on a 400 MHz NMR spectrometer. Applying uni- and multivariate analytics, our platform allows detection of suspicious samples and the identification and quantification of adulterants.

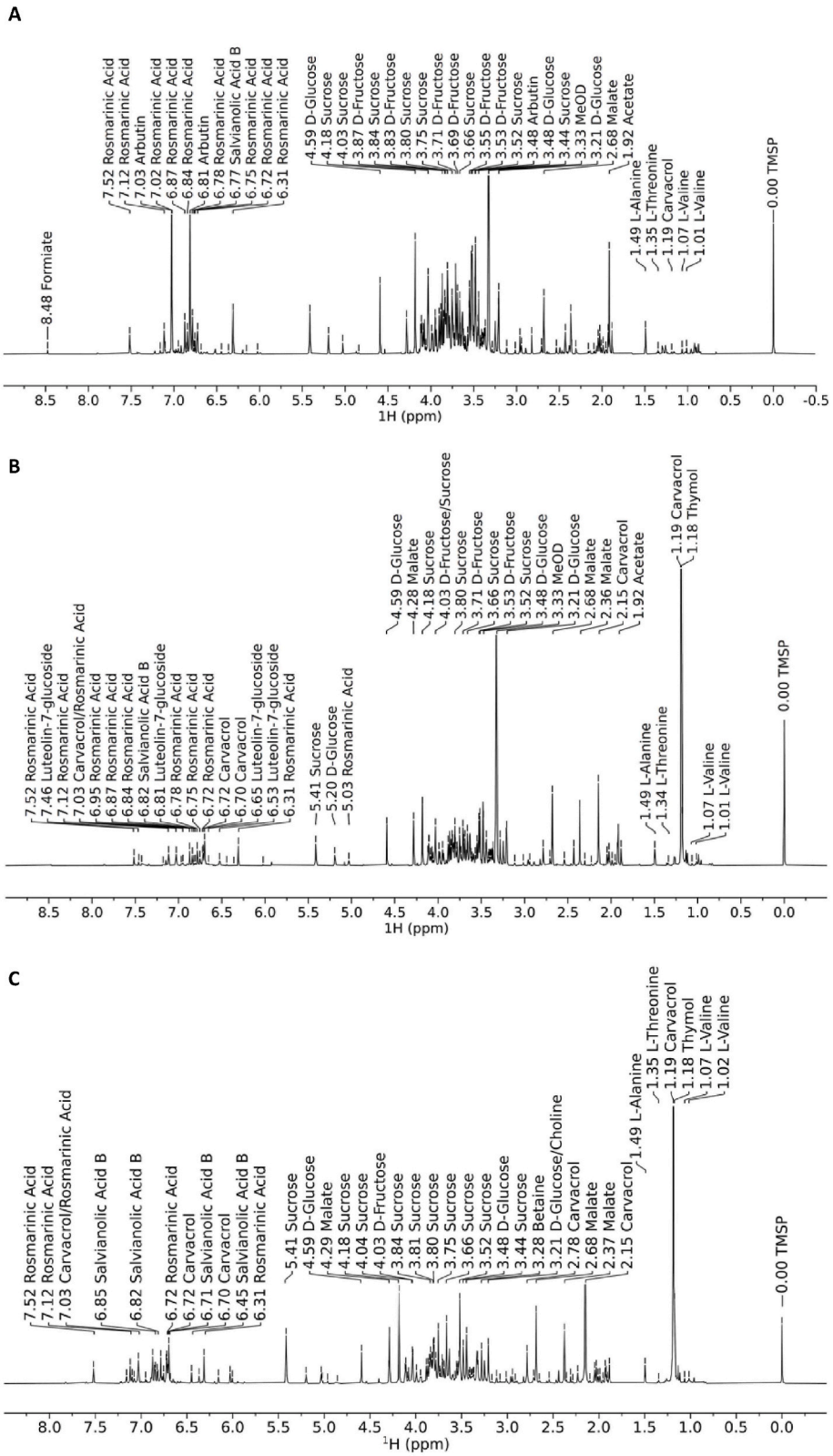

Fig. 2. 1D projection of a 2D ^{1}H *J*-RES spectrum of a representative *O. majorana,* EG **(A)**, *O. onites,* TR **(B)**, and *O. vulgare,* GR **(C)** extract.

et al., 2021). A spectral database was created for more than 90 oregano samples available on the German market.

3.2. Results of the NGS analyses

Using NGS (Thermo Fisher ion torrent technology) we have identified the presence of extraneous plant material. Most identification hits may just arise from non-intended weed contamination during farming,

for example bind weed is found in low amounts in most samples (Fig. S1). Plant species known to be used as fraudulent bulking material of oregano are also detected as shown in Fig. 3A. Thyme is found in most samples (87%), which may be due to an erroneous assignment by the NGS software. Apart of this, sweet marjoram is the most common addition found in oregano samples. We also found myrtle, olive, salvia, cistus and arbutus. We find that owing to their close phylogenic relationship, non-unique sequences in chloroplastic DNA are wrongly identified by the sequence alignment algorithm. It is also striking that only two samples were identified as pure *O. onites* (Table 1). Sequence homology between *O. onites* and *O. vulgare* is here also a likely reason.

This highlights the necessity of an orthogonal technique. Fig. 3B shows the discrimination obtained by unsupervised PCA analysis using raw NMR data showing six obviously separated groups, demonstrating the capacity of NMR to separate groups of different species. Colouring according to NGS and regional assignments demonstrates that the NMR PCA analysis correlates well with the NGS data. It is important to note that this classification is not only by species, but also by regional origin. The 6 classes are composed of 91 oregano and 38 marjoram samples. According to NGS German, Italian and Greek samples were pure *O. vulgare;* Turkish samples were *O. onites* and *O. vulgare*, whereby *O. onites* was present in larger amounts.; Chilean samples were composed of *O. vulgare* and *O. majorana* in different amounts. Finally, *O. majorana* samples originated largely from Egypt. Overall this shows that NMR can discriminate between those groups with a perfect separation in unsuperised PCA analysis. When using 3 principle components all samples that contain larger amounts of *O. majorana* (red and cyan) are clearly separated from the oregano samples along PC1. The Chilean samples (often labelled as oregano, but identified as at least partially marjoram) are further separated from pure marjoram along PC3. It is important to note that the two pure *O. onites* samples fall within the group of samples with identified *O. onites* and *O. vulgare* content. This suggests that within this group the *O. vulgare* percentage is rather small. Moreover, manual inspection of NGS data allow the hypothesis that the identification of *O. vulgare* is based on non-unique sequences, shared with *O. onites,* and therefore possibly false positive hits.

3.3. NMR analysis of oregano samples

NMR was used to search for specific biomarkers that can distinguish between oregano and adulterations. Although NMR assignments were aided by LC-MS, the focus here was the development of an NMR protocol to identify adulterations in oregano samples. As shown by Fig. 4 marjoram and oregano can be differentiated based on clearly identifiable markers, as seen in OPLS-DA loadings for the two groups (all species of oregano vs marjoram). Marjoram is characterised by high levels of arbutin which can cause liver damage as it is readily converted into hydroquinone glucuronide and hydroquinone sulfate in the gastrointestinal tract (Chakraborty et al., 1998; DrugBank, 2006), whereas oregano species show high levels of carvacrol known for its antifungal activity (Kordali et al., 2008; Kotan et al., 2014). These two substances alone are sufficient to distinguish between oregano and marjoram species. We also find that Italian oregano has larger concentrations of thymol than carvacrol (Fig. S2). Both are isomers of methyl-propane-phenol. Similarly, the different groups of oregano show different concentrations of characteristic metabolites as identified by ANOVA (Fig. 4 C-D and Fig. S2). Rosmarinic acid is lower in *O. onites* compared to all other groups, and salvianolic acid B is found at higher levels in *O. vulgare* from Greece.

3.4. Analysis of mixtures

Our further aim was to generate a platform to quantify amounts of botanical adulterants in oregano samples. For this, we have acquired NMR spectra of extracts of pure adulterants, including marjoram (*O. majorana*), olive leaves (*Olea europaea*), pink rock rose (*Cistus creticus*). For each of these we obtained spectra of several samples, in some cases with some variability between the impurity samples. For example, *Olea europaea* samples showed some variability in metabolomics spectra. We also looked at myrtle (*Myrtus communis*) and hazelnut (*Corylus avellana*) for which only one reference sample was available, respectively. Considering that genetic analysis identified a subset of samples which were sold as oregano to contain a considerable amount of *O. majorana* we first looked at combinations of *O. vulgare* and *O. majorana*. For this, we numerically summed up spectra of genetically pure *O. vulgare* with genetically pure *O. majorana*. In both groups, samples with large Q residuals or Hotelling's T^2 values in a preceding PCA analysis were excluded as outliers. We used OPLS regression to build models by correlating numerically spiked spectra with the percentage of oregano vs. marjoram. In Fig. 5A the ratio of marjoram: oregano increases from left to right on the Latent Variable 1 (LV1) axis of the scores plot. Grey dots indicate electronically spiked samples. The high variability along LV2 arising from the different oregano classes can be attributed to their regional origin, with pure Greek samples forming a separate group at low LV2 values (-20). As the content of *O. majorana* in mixed spectra increases the LV2 variation becomes smaller at higher values of LV1. Red dots represent physically prepared mixtures of oregano and marjoram samples, showing the same trend as numerical addition. Moreover, cyan dots represent suspicious samples of mostly Chilean origin, or others falling in the same group in Fig. 3. Using OPLS

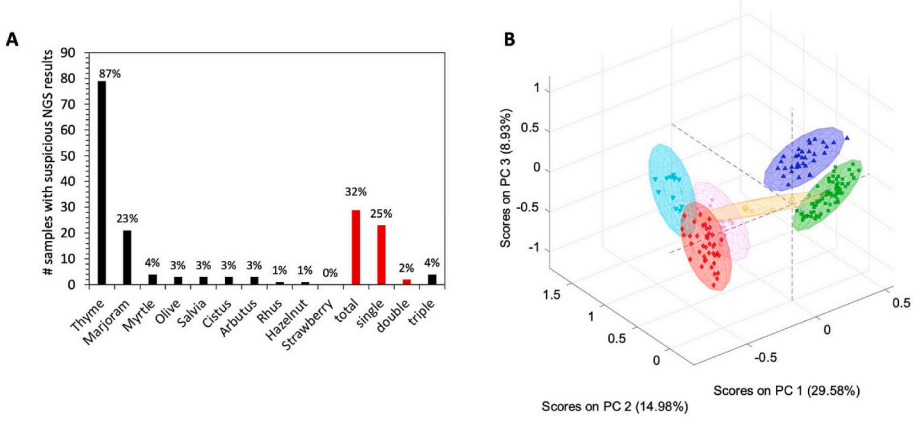

Fig. 3. A: Potential contaminants detected by NGS in oregano samples from different suppliers. The vertical axis displays the number of samples with positive NGS results for the respective species, above each bar percentage numbers indicate the fraction of oregano samples showing the respective species. Thyme is found in most samples (87%), which may be due to an erroneous assignment by the NGS software. If thyme is excluded, 29 samples (32%) were suspicious. In 25% one contaminant species was detected, in 2% two additional species, and in 4% three additional species. B: PCA score scatter plot based on ^{1}H *J*-RES NMR spectra of extracts of marjoram (*O. majorana*, ◆), mixes of *O. onites* and *O. vulgare* from Turkey (■), pure *O. vulgare* from Germany (●), Greece (▲), and Italy (★), and of *O. vulgare* samples with high ratios of *O majorana* primarily from Chile (▼). Oregano species and regional origin form clearly separated groups within the first 3 principal components. The three PCs account for 53.5% of the total variance of the data set ($Q^2 = 0.48$).

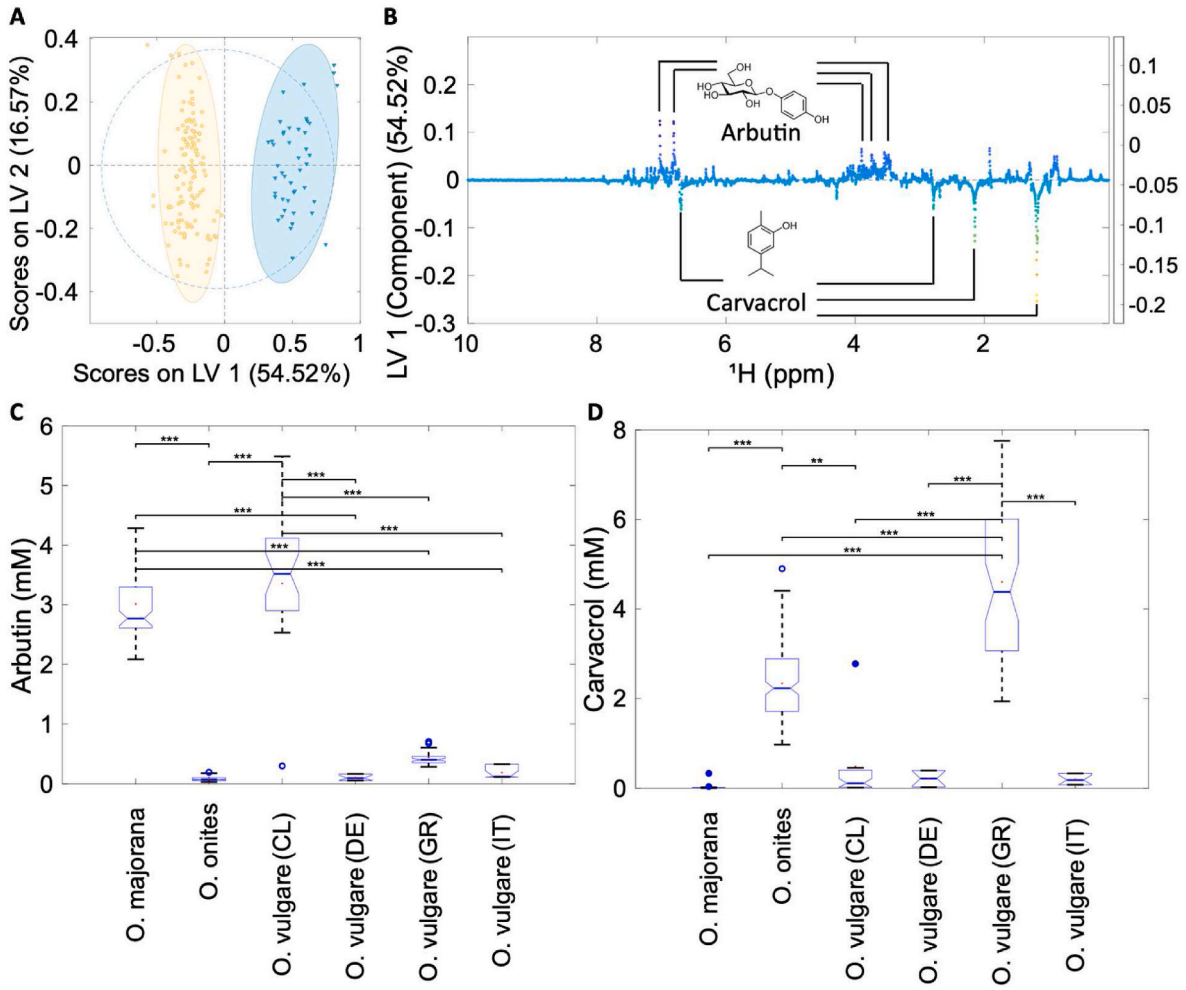

Fig. 4. A: Biomarker identification with an OPLS-DA model based on ^{1}H NMR spectra of marjoram and oregano extracts. Oregano samples in which *O. majorana* was detected by NGS were not included in model calibration. The score scatter plot shows perfect separation of oregano (●) and marjoram (▼) extracts along LV 1 with $R^2_{cal} = 0.971$, $R^2_{CV} = 0.965$, $Sens_{cal} = Sens_{CV} = 1$ and $Spec_{cal} = Spec_{CV} = 1$. B: The loadings of LV 1 allow identification of biomarkers that are responsible for discrimination between the species. Loadings are coloured according to weights on LV 1. Carvacrol is a biomarker for oregano. Arbutin is a marjoram biomarker. C–D: NMR-derived concentration differences of two biomarkers in six different groups of samples are represented as boxplots. Significant concentration differences in one-way ANOVA are marked with asterix (*α = 0.05, **α = 0.01, ***α = 0.001, Bonferroni corrected with m = 24, which equals the total number of quantified metabolites).

regression to correlate percentages of spiked spectra yields an excellent correlation (R^2 for physically spiked mixtures of 0.942). Physically mixed samples (red) are very well predicted by this model with a RMSE of 3.6%. According to this model, the unknown samples (panel C) contain up to 56% of *O. majorana*. The error in this analysis (standard deviation of biological replicates) is below 3% for more than 20% of adulteration, but rises to a maximum of 7% for smaller amounts of adulteration.

Similarly, we looked at combinations of *O. onites* with *Olea europaea*. This choice was triggered by an initial observation by NGS suggesting olive leave addition in three *O. onites* samples. One difficulty in this analysis was the spread observed in olive leaf samples causing a large spread in LV2 for higher LV1 values arising from different types of olive leaves which were not further classified. Nevertheless, we get a R^2 for the predicted values of 0.87, and a prediction error of only 4.9% for physically spiked samples. Interestingly, one of our off-the-shelf samples contained an amount of 81% of olives leaves, another 16%, both well within the detection limit of our method.

We also investigated whether NGS read percentages can be used as prediction for the actual weight percentage of a botanical contaminant in a sample. For this, olive leaves were added at defined weight percentages to seven randomly chosen oregano samples. The samples were analysed in parallel by NMR and by NGS. Although we find an increase in olive read percentages with increased weight percentage, the prediction of a contaminant based on read percentages lacks accuracy, especially compared to NMR (Fig. S3 shows OPLS regression models for NGS and for NMR).

As analytical methods improve, fraudulent sellers are likely to refer to more complex mixtures, using more than one adulterant, and we also find by NGS that such samples exist on the market. We therefore investigated three-fold mixtures of *O. onites* with *O. majorana* and *Olea europaea* (Fig. 6). Three-fold numerical mixtures were again used to build OPLS regression models (using 5 LVs), which predict additions of each species with R^2 values between 0.83 and 0.89 in physically spiked samples (two- and three-fold mixtures), which translates into errors of 4.5–6.6%. As anticipated, the more complex model based on three-fold numerical mixtures yields substantially better results when three different components are present (Fig. S4). This again establishes NMR metabolomics as highly suitable for purity analysis of oregano samples. These models can be further improved by increasing the spectral database for adulterant samples. Further extension of OPLS regression models up to 6-fold mixtures (using 5 LVs) shows clear feasibility, although at a somewhat lower accuracy (Fig. S5). For some cases such as cistus, our analysis suggests that the used cistus samples fall themselves

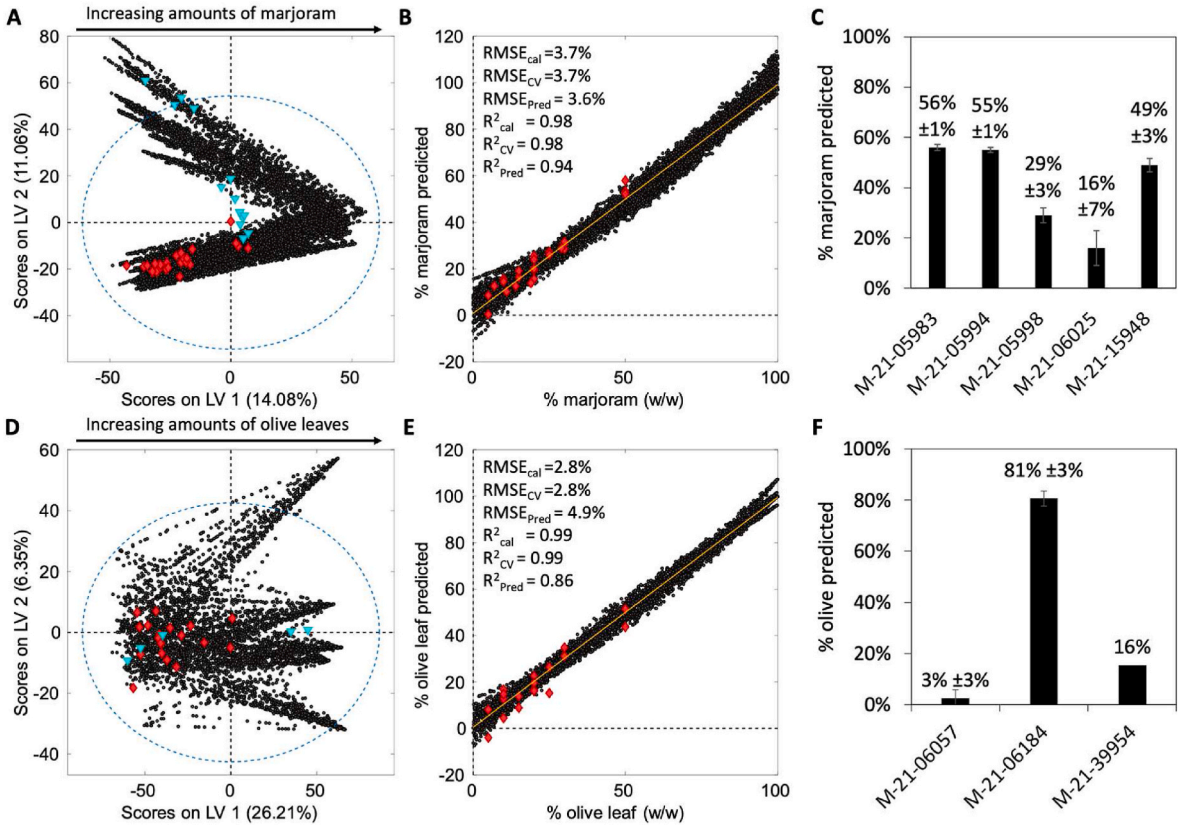

Fig. 5. A-C:Estimation of marjoram weight percentage with an OPLS regression model based on ^{1}H *J*-RES spectra of marjoram and oregano extracts (two-fold mixtures). A: Score scatter plot displaying the scores of the calibration (●), validation (◆), and test data set (▾) on LV 1 vs. LV 2 of a 5 LV model. The calibration data set contains virtually spiked spectra of each authentic *O. vulgare* sample with a single random authentic marjoram sample in the range 0–100% marjoram. The test data set comprises *O. vulgare* samples in which *O. majorana* addition is detected by NGS. B: Model-predicted marjoram weight percentage vs. the respective measured weight percentage for the calibration (●) and validation (◆) data sets. The prediction of the calibration data is determined using cross-validation with ten random splits and 10% left-out samples. The model statistics are displayed. C: Marjoram weight percentage in suspicious samples predicted by the model. Error bars represent the standard deviation of the prediction for two or three biological replicates. D–F: Estimation of olive leaf weight percentage with an OPLS regression model based on ^{1}H NMR spectra of olive leaf and oregano extracts (two-fold mixtures). The panels are analogous to the model described in A-C. The calibration data set contains virtually spiked spectra of each authentic *O. onites* sample with each olive leaf sample in a range of 0–100% of olive. The test data set comprises test samples labelled as oregano in which *Olea europaea* addition was detected by NGS.

into different classes with varying small-molecule composition. To improve statistical significance such models require larger NMR data bases for adulterants than currently available.

Commercial food quality analyses are usually based on univariate tests, using one or several substances for which a range is used for an individual species. For spices this is not easily achieved considering substantial signal overlap in NMR spectra. Nevertheless, we have a few compounds that are specific for oregano vs marjoram, or *O. onites* vs *O. vulgare*, or oregano species vs olive. To get an estimate of what can be achieved using individual discriminatory molecules, we have looked at the variation within raw NMR spectra of pure oregano species vs *O. majorana*. To identify whether a univariate test is possible, the adulterant needs to yield signals that can be detected as a significant positive or negative deviation from the reference data set. For this we have plotted the full range of signals in our data base (Fig. S6, grey colors for minimum, median and maximum) along with a *O. onites* sample mixed with *O. majorana*. In this plot, we can identify signals arising from arbutin even when only 5% of *O. majorana* is present. Similarly, we find such signals for olive leaves which allow for a univariate discrimination from oregano samples.

4. Discussion

Food fraud is a substantial problem across the food industry and occurrences of food fraud become more prevalent. One incident can cause losses between 2 and 15% of annual revenue (Galvin-King et al., 2018). Spices are particularly vulnerable. Consumers expect that the food products are correctly labelled, and this may also have health implications, for example to avoid allergic reactions against specific food products. Modern food industries endeavour to provide consumers with well-characterised products, which requires suitable analytical procedures to identify adulterants. Considering the market size with almost 500.000 tonnes of spices annually consumed alone in Europe, budgetary constraints are unlikely to prevent against the use of sophisticated analytical procedures. For NGS, NMR and LC-MS it will usually require highly skilled and certified expert laboratories to provide such tests. With the advance of NGS databases for good products, suppliers have started to rely on this technology to identify adulterants. This has led to false expectations and confusion as read numbers are often translated into concentrations. The results presented here clearly show that this is not justified. Based on our results, we can confirm the statements of the ESA in a recent white paper "NGS should not be used as a direct tool to authenticate herbs and spices nor to detect or quantify them as ingredients in mixtures." (European Spice Association (ESA), 2021).

Following this advice, NGS should be combined with orthogonal, non-targeted analytical methods, for spice authentication and particularly for adulterant quantification (European Spice Association (ESA), 2021). In principle, this can be achieved with near-infrared (NIR) spectroscopy, MS or NMR (Black et al., 2016; Ivanovic et al., 2021; Mandrone et al., 2021; McGrath et al., 2021; Rodionova & Pomerantsev,

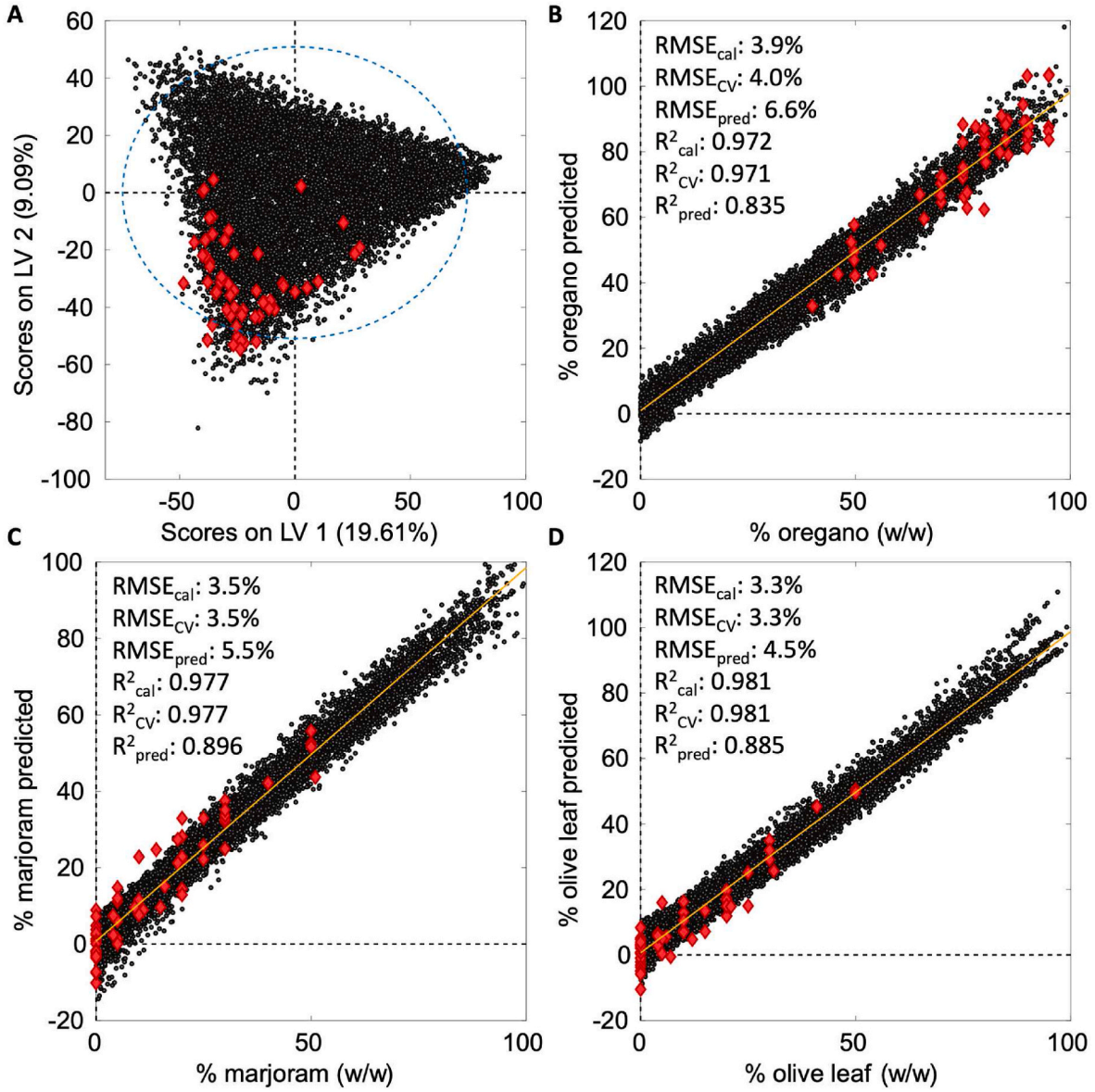

Fig. 6. Estimation of oregano, marjoram and olive leaf weight percentage with an OPLS regression model based on ^{1}H NMR spectra of marjoram, olive leaf and oregano extracts (three-fold mixtures). A: Score scatter plot displaying the scores of the calibration (●) and validation (◆) data set on LV 1 vs. LV 2 of a 5-LV model. The calibration data set contains virtually spiked spectra of each authentic *O. onites* sample with a combination of a marjoram and olive leaf sample in the range 0–100% marjoram or olive leaves with steps of 5%. B: Model predicted oregano weight percentage vs. measured oregano weight percentage for the calibration (●) and validation (◆) data sets. The prediction of the calibration data was determined by CV using ten random splits and 10% left-out samples. The model statistics are displayed. C: Model predicted marjoram weight percentage vs. measured marjoram weight percentage. D: Model predicted olive leaf weight percentage vs. measured olive leaf weight percentage.

2020; Wielogorska et al., 2018). Within this set NIR is arguably the cheapest technology, although with the limitation that only fingerprinting is feasible. While this may be sufficient for mixtures of two different components, it is desirable to gain further quantitative information. This leaves a choice of MS vs NMR. While LC-MS covers a substantially larger number of compounds, it is also subject to larger noise, especially for measurements that are spread over longer time periods. In contrast, NMR yields reproducible results well within 1% of error margin with minimal quality control. With the availability of modern food-screeners that require minimal maintenance at costs similar to mass spectrometers, NMR has become a competitive alternative.

In this study, we have shown that a simple workflow for NMR analysis can be used as a powerful and fast tool to identify and quantify adulterants in oregano samples above a limit of approximately 5%. Moreover, we can classify oregano according to regional origin and species. The data shown demonstrates that authentication is possible using a 400 MHz NMR instrument. Initially, NGS is required to screen and authenticate the analytical material and to calibrate the NMR database. With increasing sample numbers, the analytical platform improves, and sample screening only based on NMR will be possible. DNA sequencing of every sample will not be necessary but can still provide important additional information for suspicious samples, identified by NMR.

Applying NMR, we can quantify the percentage of an adulterant with error rates of approximately 3–7%. This can potentially be improved with a larger compound database or by combining multivariate and univariate approaches, i.e., by also using specific biomarkers for the discrimination. Although multivariate analysis is generally more powerful, it may be desirable to combine those with univariate tests for

specific markers. As analytical methods improve, fraudulent sellers are likely to refer to more complex mixtures, using more than one adulterant, and we also find by NGS that such samples exist on the market. Our analytical platform is also able to quantify percentages of more than one adulterant with reasonable accuracy.

In the case of mixtures including *O. majorana* but labelled as oregano, it may be questionable whether these products were accidently mislabelled or intentionally mixed, for example to reduce costs. Considering the similarity of the plants, the severity of "fraud" is of a lesser significance than the addition of olive leaves which by themselves were not produced for consumption. The finding that some oreganos had 80% of olive leaves is indeed striking and demonstrates the importance of such analytical tools to be available for the food industry.

The simplicity of the overall NMR workflow should encourage application for other food products. While it is already well established for juices, honey and wine, little is available for other products. The main effort is in building a suitable reference database for different botanical species or derived products. This will be difficult for plant products with a substantial range of varieties such as teas, but should be well possible for spices, coffees, and other plant products that are sold as pure commodities. The cost of building such databases should be well justified by the consumer benefit in a world of increased food consciousness.

Disclosures

LADR GmbH MVZ Dr. Kramer & Kollegen and the University of Lübeck have jointly developed this method and an NMR spectral library for oregano and common adulterants with the goal to derive a commercial analytical product. The spectral library will not be disclosed and is the property of LADR GmbH MVZ Dr. Kramer & Kollegen.

CRediT authorship contribution statement

Friedemann Flügge: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Tim Kerkow:** Methodology, Validation, Investigation, Resources, Data curation, Writing – review & editing. **Paulina Kowalski:** Methodology, Investigation. **Josephine Bornhöft:** Methodology, Investigation. **Eva Seemann:** Methodology, Investigation. **Marina Creydt:** Writing – review & editing. **Burkhard Schütze:** Conceptualization, Methodology, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Ulrich L. Günther:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

LADR GmbH MVZ Dr. Kramer & Kollegen and the University of Lübeck have jointly developed this method and an NMR spectral library for oregano and common adulterants with the goal to derive a commercial analytical product. The spectral library will not be disclosed and is the property of LADR GmbH MVZ Dr. Kramer & Kollegen.

Friedemann Flügge reports a relationship with LADR GmbH MVZ Dr. Kramer & Kollegen that includes: employment. Tim Kerkow reports a relationship with LADR GmbH MVZ Dr. Kramer & Kollegen that includes: employment. Paulina Kowalski reports a relationship with LADR GmbH MVZ Dr. Kramer & Kollegen that includes: employment. Josephine Bornhöft reports a relationship with LADR GmbH MVZ Dr. Kramer & Kollegen that includes: employment. Eva Seemann reports a relationship with LADR GmbH MVZ Dr. Kramer & Kollegen that includes: employment. Burkhard Schütze reports a relationship with LADR GmbH MVZ Dr. Kramer & Kollegen that includes: employment.

Data availability

The data that has been used is confidential.

Acknowledgments

This work has been funded by the European Union from the European Regional Development Fund (EFRE) and the federal government and state of Schleswig-Holstein, Germany. Project number: LPW-E/1.2.3/1708; Application number: 095 20 011; Project: "LADR-NLA".

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodcont.2022.109497.

References

Black, C., Haughey, S. A., Chevallier, O. P., Galvin-King, P., & Elliott, C. T. (2016). A comprehensive strategy to detect the fraudulent adulteration of herbs: The oregano approach. *Food Chemistry, 210*, 551–557. https://doi.org/10.1016/j.foodchem.2016.05.004

Chakraborty, A. K., Funasaka, Y., Komoto, M., & Ichihashi, M. (1998). Effect of arbutin on melanogenic proteins in human melanocytes. *Pigment Cell Research, 11*(4), 206–212. https://doi.org/10.1111/j.1600-0749.1998.tb00731.x

DrugBank. (2006). *DrugBank. DrugBank: A comprehensive resource for in silico drug discovery and exploration*. https://go.drugbank.com/drugs/DB11217. (Accessed 26 April 2022).

European Commission. Joint Research Centre. (2021). *Results of an EU wide coordinated control plan to establish the prevalence of fraudulent practices in the marketing of herbs and spices*. Publications Office. https://data.europa.eu/doi/10.2760/309557.

European Spice Association (ESA). (2018). *Quality minima document* (Vol. 5). Rev. https://www.esa-spices.org/index-esa.html/publications-esa

European Spice Association (ESA). (2021). *ESA white paper on plant metabarcoding NGS analysis applied to culinary herbs and spices*.

FooDB. (2022). *FooDB*. https://foodb.ca/. (Accessed 5 May 2022).

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

Godelmann, R., Fang, F., Humpfer, E., Schütz, B., Bansbach, M., Schäfer, H., & Spraul, M. (2013). Targeted and nontargeted wine analysis by ^{1}H NMR spectroscopy combined with multivariate statistical analysis. Differentiation of important parameters: Grape variety, geographical origin, year of vintage. *Journal of Agricultural and Food Chemistry, 61*(23), 5610–5619. https://doi.org/10.1021/jf400800d

Günther, U. L., Ludwig, C., & Rüterjans, H. (2000). NMRLAB—advanced NMR data processing in Matlab. *Journal of Magnetic Resonance, 145*(2), 201–208. https://doi.org/10.1006/jmre.2000.2071

Guyader, S., Thomas, F., Portaluri, V., Jamin, E., Akoka, S., Silvestre, V., & Remaud, G. (2018). Authentication of edible fats and oils by non-targeted 13 C INEPT NMR spectroscopy. *Food Control, 91*, 216–224. https://doi.org/10.1016/j.foodcont.2018.03.046

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

ISO 7925:1999. (1999). Dried oregano (Origanum vulgare L.)—whole or ground leaves—Specification. , October 1 https://www.iso.org/obp/ui/#iso:std:iso:7925:ed-2:v1:en.

Ivanovic, S., Mandrone, M., Simic, K., Ristic, M., Todosijevic, M., Mandic, B., & Godjevac, D. (2021). GC-MS-based metabolomics for the detection of adulteration in oregano samples. *Journal of the Serbian Chemical Society, 86*(12), 1195–1203. https://doi.org/10.2298/JSC210809089I

Kordali, S., Cakir, A., Ozer, H., Cakmakci, R., Kesdek, M., & Mete, E. (2008). Antifungal, phytotoxic and insecticidal properties of essential oil isolated from Turkish Origanum acutidens and its three components, carvacrol, thymol and p-cymene. *Bioresource Technology, 99*(18), 8788–8795. https://doi.org/10.1016/j.biortech.2008.04.048

Kotan, R., Cakir, A., Ozer, H., Kordali, S., Cakmakci, R., Dadasoglu, F., Dikbas, N., Aydin, T., & Kazaz, C. (2014). Antibacterial effects of Origanum onites against phytopathogenic bacteria: Possible use of the extracts from protection of disease caused by some phytopathogenic bacteria. *Scientia Horticulturae, 172*, 210–220. https://doi.org/10.1016/j.scienta.2014.03.016

Ludwig, C., & Günther, U. L. (2011). MetaboLab-advanced NMR data processing and analysis for metabolomics. *BMC Bioinformatics, 12*(1), 366.

Mandrone, M., Marincich, L., Chiocchio, I., Petroli, A., Gođevac, D., Maresca, I., & Poli, F. (2021). NMR-based metabolomics for frauds detection and quality control of oregano samples. *Food Control, 127*, Article 108141. https://doi.org/10.1016/j.foodcont.2021.108141

Martin, G. J., Martin, M. L., Mabon, F., & Michon, M. J. (1982). Identification of the origin of natural alcohols by natural abundance hydrogen-2 nuclear magnetic resonance. *Analytical Chemistry, 54*, 2380–2382.

McGrath, T. F., Haughey, S. A., Islam, M., Elliott, C. T., Kelly, S. D., Suman, M., Rindy, T., Taous, F., García-González, D. L., Singh, D., Nan Wai, S., Zitek, A., Jandric, Z., Almirall, J., Acosta, A., Thao, C., Islam, M., Hoo Fung, L., Kim, J.-K., … Handagiripathira, H. M. N. L. (2021). The potential of handheld near infrared spectroscopy to detect food adulteration: Results of a global, multi-instrument inter-laboratory study. *Food Chemistry, 353*, Article 128718. https://doi.org/10.1016/j.foodchem.2020.128718

Pluskal, T., Castillo, S., Villar-Briones, A., & Orešič, M. (2010). MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics, 11*, 395. https://doi.org/10.1186/1471-2105-11-395

Prosser, S. W. J., & Hebert, P. D. N. (2017). Rapid identification of the botanical and entomological sources of honey using DNA metabarcoding. *Food Chemistry, 214*, 183–191. https://doi.org/10.1016/j.foodchem.2016.07.077

Regulation (EU). (2017). 2017/625 of the European parliament and of the Council, 095 OJ L. http://data.europa.eu/eli/reg/2017/625/oj/eng.

Rodionova, O. Y.e., & Pomerantsev, A. L. (2020). Chemometric tools for food fraud detection: The role of target class in non-targeted analysis. *Food Chemistry, 317*, Article 126448. https://doi.org/10.1016/j.foodchem.2020.126448

Spiteri, M., Jamin, E., Thomas, F., Rebours, A., Lees, M., Rogers, K. M., & Rutledge, D. N. (2015). Fast and global authenticity screening of honey using 1H-NMR profiling. *Food Chemistry, 189*, 60–66. https://doi.org/10.1016/j.foodchem.2014.11.099

Spraul, M., Schütz, B., Rinke, P., Koswig, S., Humpfer, E., Schäfer, H., Mörtter, M., Fang, F., Marx, U., & Minoja, A. (2009). NMR-based multi parametric quality control of fruit juices: SGF profiling. *Nutrients, 1*(2), 148–155. https://doi.org/10.3390/nu1020148

Wielogorska, E., Chevallier, O., Black, C., Galvin-King, P., Delêtre, M., Kelleher, C. T., Haughey, S. A., & Elliott, C. T. (2018). Development of a comprehensive analytical platform for the detection and quantitation of food fraud using a biomarker approach. The oregano adulteration case study. *Food Chemistry, 239*, 32–39. https://doi.org/10.1016/j.foodchem.2017.06.083