NMR-based metabolomics for frauds detection and quality control of oregano samples

Manuela Mandrone, Lorenzo Marincich, Ilaria Chiocchio, Alessandra Petroli, Dejan Gođevac, Immacolata Maresca, Ferruccio Poli

PII: S0956-7135(21)00279-6

DOI: https://doi.org/10.1016/j.foodcont.2021.108141

Reference: JFCO 108141

To appear in: *Food Control*

Received Date: 15 January 2021

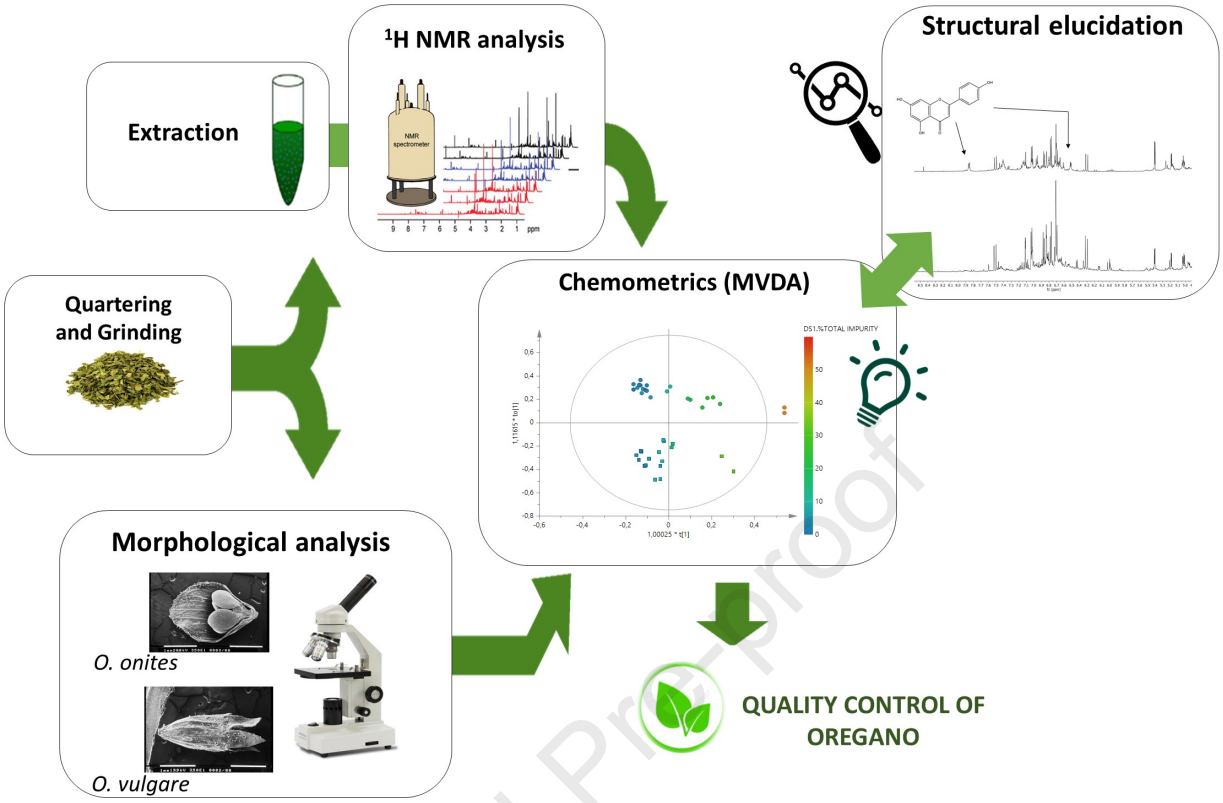
Revised Date: 10 March 2021

Accepted Date: 3 April 2021

Please cite this article as: Mandrone M., Marincich L., Chiocchio I., Petroli A., Gođevac D., Maresca I. & Poli F., NMR-based metabolomics for frauds detection and quality control of oregano samples, *Food Control*, https://doi.org/10.1016/j.foodcont.2021.108141.

Author Contributions

Conceptualization: Manuela Mandrone (M.M.), Ferruccio Poli (F.P.); **Data curation:** M.M., Lorenzo Marincich (L.M.), Ilaria Chiocchio (I.C.); **Plant material authentication:** F.P.; Immacolata Maresca (I.M.); **Morphological Analysis and Microscopy:** I.M. and L.M.; **Metabolomics and Structure Elucidation:** M.M, I.C., L.M., Dejan Gođevac (D.G.); **Methodology:** M.M., L.M., Alessandra Petroli (A.P.); **NMR Experimental design:** A.P., M.M.; **Investigation:** M.M, I.C., L.M., F.P., D.G.; **Software:** M.M., A.P., L.M.; **Funding acquisition:** F.P.; **Project administration:** M.M. and F.P; **Supervision:** M.M. and F.P.; **Writing – original draft:** M.M., L.M., I.C.; **Writing – review & editing:** M.M, F.P, D.G. All authors have read and agreed to the published version of the manuscript.



NMR-based metabolomics for frauds detection and quality control of oregano samples

Manuela Mandrone[a*], Lorenzo Marincich[a], Ilaria Chiocchio,[a] Alessandra Petroli[b], Dejan Gođevac[c], Immacolata Maresca[d], Ferruccio Poli[a]

[a]Department of Pharmacy and Biotechnology, University of Bologna, Via Irnerio, 42, 40126 Bologna, Italy

[b]Department of Industrial Chemistry, University of Bologna, Viale del Risorgimento, 4, 40136 Bologna, Italy

[c]Department of Chemistry, Institute of Chemistry, Technology, and Metallurgy, University of Belgrade, Njegoševa 12, 11000 Belgrade, Republic of Serbia

[d]Department of Life Sciences and Biotechnology, University of Ferrara, via Borsari 46, 44100 Ferrara, Italy

***Correspondence**

Dr. Manuela Mandrone, University of Bologna, Department of Pharmacy and Biotechnology, Via Irnerio 42, 40126 Bologna, Italy

E-mail: manuela.mandrone2@unibo.it Phone: +390512091294; Fax +39051242576

Abstract

In this work ^{1}H NMR metabolomics has been employed for quality control of oregano samples. NMR data and morphological analysis (MA) were combined by PCA, obtaining a model able to individuate non-marketable samples, and to distinguish between the two marketable oregano species (*Origanum vulgare* and *O. onites*) on the basis of their metabolomic profile. Through this approach distinctive biomarkers of the two species were found, namely apigenin and p-cymene for *O. onites*, and salvianolic acid B for *O. vulgare*. Furthermore, the percentage of the samples' impurity (evaluated by MA) and the metabolomic profiles were correlated by OPLS models, which showed that, in addition to the species-specific biomarkers, thymol and rosmarinic acid (common to both marketable species) strongly correlate to oregano degree of purity. Cistus was one of the most frequent contaminants, thus, a further OPLS model, able to detect the degree of cistus contamination in oregano samples, was also built.

Keywords

NMR-based metabolomics; quality control; biomarkers; apigenin; p-cymene; salvianolic acid B; oregano; cistus

Chemical compounds studied in this article:

Apigenin (PubChem CID: 5280443); Carvacrol (PubChem CID: 10364); p-Cymene (PubChem CID: 7463); Rosmarinic acid (PubChem CID: 5315615); Salvianolic acid B (PubChem CID: 11629084); Thymol (PubChem CID: 6989)

List of abbreviations

MA: morphological analysis

MVDA: multivariate data analysis

NMR: nuclear magnetic resonance

OPLS: orthogonal partial least squares

OPLS-DA: orthogonal partial least squares discriminant analysis

PCA: principal components analysis

VIP: variable influence on projection

1. Introduction

The global herbs and spices market is increasingly threatened by accidental or intentional adulteration, also facilitated by the complexity of the supply chains (Galvin-King et al., 2018). Herbs are generally sold chopped or powdered, making it relatively easy to counterfeit by means of adding cheaper bulking agents (i.e. chemicals, extraneous material from other plants or foreign species) (Van Ruth et al., 2018). In this scenario, the development of rapid, robust, cost-effective and environmentally friendly methods for authentication and quality control of spices and herbs takes on great importance (Wadood et al. 2020).

Oregano is one of the most adulterated herbs (Black *et* al., 2016; L. Drabova et al., 2019). Despite the high heterogeneity of *Origanum* genus, only a few species are accepted on the market. According to ISO 7925:1999, all *Origanum* species and sub-species, except *Origanum majorana* L. are considered marketable, while according to the European Spice Association (ESA) only *Origanum vulgare* L. and

Origanum onites L. (or products made of a mixture of the two) are considered 'true' oregano (List of culinary herbs and spices, ESA, 2018), with the limit of permitted impurities set at 2% (w/w) (Quality minima document, ESA, 2015). With respect to oregano authentication, the European Pharmacopoeia makes a further restriction: in addition to *O. onites*, it only considers a specific subspecies (subsp.) of *O. vulgare* as 'true' oregano, namely *O. vulgare* subsp. *hirtum* (Link) Ietsw (9^{th} European Pharmacopoeia, 2017).

Besides the frauds ascribable to the sale of non-marketable species, oregano is often contaminated with a significant amount of other plants such as olive tree leaves, myrtle, sumac, cistus, savory and others (Black et al. 2016, Marieschi et al., 2009; Marieschi et al., 2010; Marieschi et al., 2011a; Marieschi et al., 2011b).

Generally, companies dealing with large-scale sales buy oregano from third parties. The contractor sends a representative sample and the company analyzes it in order to establish, on the basis of the sample quality, whether to buy it or not. In this stage, in order to determine the authenticity and purity of samples, companies generally rely on morphological analysis performed by microscopy (MA). Nevertheless, this technique is time-consuming, and thus not suitable for quickly analyzing a high number of samples; furthermore, its precision strongly depends on the operator performing the analysis. This makes it of great importance to develop new efficient, quick and cost-affordable methods for oregano quality control, and to contribute to building databases, apt to facilitate fraud detection worldwide.

Marieschi et al. (2009, 2010, 2011a, 2011b) published several works focused on the assessment of oregano authenticity by sequence-characterized amplified region makers (SCARs), which also allowed them to recognize a number of specific adulterants. The technique proposed is extremely precise and accurate, however, it does not provide information about the phytochemical features of the samples,

which is a further element of interest to establish the quality of an herbal product. In fact, the presence of specific metabolites confers the characteristic taste and flavor to the herb. Moreover, several plant secondary metabolites are organ-specific, this implies, taking for instance the case of oregano, that a high-quality sample should be made of leaves and bracts predominantly, with the lowest possible amount of stem. Hence, the analysis of the phytochemical profile is an important additional step for the quality assessment of herbs, also considering that the metabolites production in plants is significantly affected by a number of factors, including seasonal variation, altitude, biotic and abiotic factors and so on (Mandrone et al 2021, Anđelković et al., 2017; Scognamiglio et al., 2015; Salomè et al., 2020).

In the field of phytochemistry applied to food or herb quality control, in addition to targeted analysis, new untargeted strategies, relying on an inductive approach, have recently been emerging. In this case, the analytical protocols used are set to detect the widest possible range of metabolites (metabolome) (Riedl et al., 2015). According to this approach, the insights are gained from comparison, on a pattern level, of a high number of samples by multivariate data analysis (MVDA). This workflow was successfully applied by Black et al. (2016) to determine oregano quality by both FTIR and LC-HRMS untargeted analysis. In this context, a further analytical option is offered by NMR-based metabolomics, increasingly employed for untargeted food and herbs quality control (Sobolev et al., 2019). NMR is non-destructive, quickly performed and environmental-friendly, and in addiction it provides numerous detailed chemical information in a single spectrum (Pontes et al., 2017).

This study explores the potentiality of ^{1}H NMR metabolomics to be employed for oregano quality control. In particular, spectral data and results obtained by MA (performed by stereomicroscope and optical microscope) were combined by MVDA. Predictive multivariate data models, able to quickly provide a wide spectrum of information on oregano samples, were obtained, and oregano biomarkers of quality easily detectable by ^{1}H NMR profiling were identified.

2. Methods and material

2.1 Chemicals

Deuterium oxide (D_2O, 99.90% D), CD_3OD (99.80% D) were purchased from Eurisotop (Cambridge Isotope Laboratories, Inc, France). Standard 3-(trimethylsilyl)-propionic-2,2,3,3-d_4 acid sodium salt (TMSP), sodium phosphate dibasic anhydrous and sodium phosphate monobasic anhydrous and all the other chemicals and solvents were purchased from Sigma-Aldrich Co. (St. Louis, MO, USA).

2.2 Oregano samples and *Cistus creticus*

Different Italian companies provided a total of twenty-seven samples of (supposed) oregano to be analyzed in order to assess purity and authenticity. According to what declared by the companies: 14 samples were cultivated in Turkey, 3 in Sicily, 3 in Albania, 2 in Peru, while for 5 samples this information was not available. Vouchers of crude drugs were deposited in Department of Pharmacy and Biotechnology, University of Bologna (via Irnerio 42, Bologna, Italy) and reported in Table 1, together with the results obtained by the MA performed in this work.

Cistus creticus (ex. *incanus*) L. was harvested on April 2017 in Ostuni (Brindisi, Italy; 40°43'31.153" N 17°35'27.844" E), the sample was identified by F.P., and voucher specimen (BOLO0602029) was retained at the Herbarium of Alma Mater Studiorum University of Bologna (SMA) (Via Irnerio 42, 40126, Bologna, Italy).

2.3 Morphological analysis

The identification of oregano species was performed according to the oregano taxonomic key presents in Ietswaart (1980). Morphological analysis (MA) was performed following the method reported by Marieschi et al. 2009. In particular, twenty-five milligrams of each dried sample were analyzed by a

stereoscopic microscope (Nikon SMZ-1000) and a microscope (Nikon Eclipse E600). The pictures were obtained by Scanning Electron Microscopy. Specifically, dried calices of *Origanum onites* and *Origanum vulgare* were mounted on aluminum stubs with double-stick carbon tape, sputter coated with gold nanoparticles (5 nm) using a Emitech K500 coater and observed with a Philips SEM 515 at 20.0 and 19.9 kV respectively.

2.4 Sample preparation for metabolomics

Samples were constituted by dried and shredded plant material. They were quartered and powdered using an electrical grinder (IKA, A11 basic, Merck, Italy) Then 50 mg were extracted with 1 mL of mixture (1:1) of phosphate buffer (90 mM; pH 6.0) in H_2O-d_2 (containing 0.1% TMSP) and MeOH-d_4 by ultrasonication (TransSonic TP 690, Elma, Germany) for 20 minutes at 45°C. After this procedure, samples were centrifuged for 10 min (17000 x *g*), then 700 µL of supernatant were transferred into NMR tubes. For each sample, two different extracts were prepared to test reproducibility.

2.5 NMR and ESI-MS measurement

^{1}H NMR spectra, *J*-resolved (J-res), ^{1}H-^{1}H homonuclear (COSY) and inverse detected ^{1}H-^{13}C correlation experiments (HMBC, HSQC) were recorded at 25°C on a Varian Inova instrument (equipped with a reverse triple resonance probe). For ^{1}H NMR profiling the instrument was operating at ^{1}H NMR frequency of 600.13 MHz, and H_2O-d_2 was used as internal lock. Each ^{1}H-NMR spectrum consisted of 256 scans (corresponding to 16 min) with the relaxation delay (RD) of 2 s, acquisition time 0.707 s, and spectral width of 9595.8 Hz (corresponding to δ 16.0). A presaturation sequence (PRESAT) was used to suppress the residual water signal at δ 4.83 (power = -6dB, presaturation delay 2 s). For 1D and 2D NMR analysis of pure compounds MeOH-d_4 was used as lock. *J*-res spectra has been measured in order to clarify or confirm splitting patterns and coupling constants values. This

153 approach is especially useful in metabolomics where a mixture of compounds is visible in a single ^{1}H
154 NMR spectrum making more difficult to establish these parameters (Emwas et al, 2019).

155 For ESI-MS analyses, dried pure compounds were dissolved in MeOH, and analyzed by WATERS ZQ
156 4000 (Milford, MA USA) mass spectrometer. in negative and/or positive ion modes according to the
157 more ionizable chemical groups of samples. A direct infusion of 20 μL/min, source temperature of 80
158 °C and desolvation (nitrogen) gas (flow rate of 200 L/h) were common parameters used in both
159 positive and negative ion modes. Capillary potential and source cone were 3.54 Kv and 20 V in
160 positive ion mode, and 2.53 Kv and 30 V in negative ion mode. The mass range was from 0 to 1000
161 *m*/*z*.

162 **2.6 Multivariate data analysis**

163 The spectra were manually phased and baseline corrected, and calibrated to the internal standard
164 trimethyl silyl propionic acid sodium salt (TMSP) at δ 0.0 using Mestrenova software (Mestrelab
165 Research, Spain). The regions of δ 5–4.5 and δ 3.34–3.30 were excluded from the analysis because of
166 the residual solvent signals. Then spectral intensities were reduced to integrated regions of equal width
167 (δ 0.04) corresponding to the region from δ 0.0 to 10.0, and normalized by total area.

168 The analysis of ^{1}H NMR profiles of extracts was performed based on an in-house library and
169 comparison with literature (Mandrone et al., 2017; Mandrone et al., 2019b).

170 SIMCA P+ software (v. 15.0, Umetrics, Sweden) was used in order to perform PCA, OPLS, OPLS-
171 DA, PLS-DA models. For multivariate analysis, data were subjected to Pareto scaling. Supervised
172 models were evaluated by the goodness of fit (R^2x (cum) and R^2y(cum)) and goodness of prediction
173 (Q^2(cum)), together with the parameters given by cross validation tests: permutation test (performed
174 using 200 permutations) and CV-ANOVA (Malzert-Fréon et al., 2010).

2.7 Purification and structure elucidation of biomarkers

In order to elucidate the structure of the metabolites of interest, different NMR-guided isolation procedures ware carried out for *O. onites*, *O. vulgare* and *Cistus creticus*. The NMR spectra of the isolated compounds and/or fractions are given in supporting material (Fig. S1 to S8).

2.6.1 Isolation of biomarkers from O. onites and O. vulgare

Forty g of *O. onites* (degree of purity of 99.88%) or *O. vulgare* (degree of purity of 98.74%) were extracted using 300 mL of CH_3OH:H_2O (1:1), sonicating for 30 min and centrifuging for 10 min (2469 x *g*). The obtained supernatant was dried in rotary evaporator at 40°C yielding 18% w/w for *O. onites* and 15% for *O. vulgare*. In both cases, the dry extract was suspended into 300 mL of water to undergo liquid-liquid partition with $CHCl_3$, EtOAc and *n*-BuOH used in sequence, repeating the procedure twice for each solvent. Fractions were dried in rotary evaporator, and for *O. onites* 5 g were obtained from the water fraction, 1.5 g from the *n*-BuOH fraction, 0.36 g from the EtOAc fraction, and 0.2 g from the $CHCl_3$ fraction. While for *O. vulgare* 4.50 g were obtained from the water fraction, 0.96 g from the *n*-BuOH fraction, 0.27 g from the EtOAc fraction and 0.22 from the $CHCl_3$ fraction.

The biomarkers of *O. onites* were found (by ^{1}H-NMR analysis) in the $CHCl_3$ and the *n*-BuOH fractions, while the water fraction contained the biomarker of *O. vulgare*. Fractions in EtOAc obtained from both species contained rosmarinic acid, whose structure was elucidated by NMR and ESI-MS experiments (Exarchou, et al. 2003). The $CHCl_3$ fraction obtained from *O. onites* was analyzed by ^{1}H NMR and ESI-MS (by direct infusion) allowing to confirm the structures of thymol and p-cymene (Exarchou, et al. 2003). For purification and structure elucidation, 200 mg of *n*-BuOH and water fractions from *O. onites* and *O. vulgare*, respectively, were further fractionated by column chromatography, using in both cases a chromatography column (1800 mm x 25 mm) filled with 220

198 g of Sephadex (LH-20) and, as eluent, methanol. The fractions were suspended in the minimum
199 amount of CH_3OH:H_2O (1:1) to be chromatographed. The flow rate was of 0.4 mL/min. Each
200 fraction was concentrated and analyzed by ^{1}H-NMR. Fraction 114 contained the biomarker of
201 interest of *O. onites*, which has been identified as apigenin (0.1 mg) (Exarchou, et al. 2003). In case
202 of *O. vulgare* fractions from 25 to 27 contained the biomarker of interest identified as salvianolic
203 acid B (10.9 mg) (Sun et al., 2009). Table S1 summarizes the results of the NMR experiments
204 performed in order to elucidate the structure of salvianolic acid B (^{1}H NMR, HMBC, HSQC,
205 COSY).

206 *2.7.2 Isolation of pinitol as biomarker of C. creticus*

207 In order to characterize the biomarker of *Cistus creticus* L., 9.8 g of dried and grounded plant
208 material were extracted using 300 mL of CH_3OH:H_2O (1:1); sonicated for 40 min; centrifuged for
209 10 min (2469 x *g*). This extraction procedure was repeated twice on the same plant material. The
210 extract was dried in rotary evaporator (yield = 30.6% w/w) and suspended in 300 mL of water to
211 undergo liquid-liquid partition with $CHCl_3$ and EtOAc (for two times each), obtaining 2.9 g of
212 material from the water fraction and 0.1 g from the EtOAc fraction. According to ^{1}H-NMR analysis,
213 water fraction contained the metabolite of interest, thus 300 mg of this fraction were suspended in
214 500 µL of water and injected in MPLC instrument (Reveleris®, Büchi, Switzerland) connected to
215 C_{18} column (4 g). A gradient of water (solvent A) and methanol (solvent B) was used as eluent. The
216 steps gradient was composed by an isocratic phase of 10 min (95% A and 5% B), a gradient from
217 95% A to 90% A in 10 min, an isocratic phase of 10 min (90% A and 10% B), a gradient from 90%
218 A to 80% A in 10 min, an isocratic phase of 10 min (80% A and 20% B), a gradient from 80% A to
219 0% A in 20 min. The flow rate was 3 mL/min, and the run length was 70 minutes. A total of 43

220 fractions were collected, and the metabolite of interest was found in was in fraction 3 and fraction 4.
221 NMR and MS analysis suggested that this compound might be pinitol.

222 Apigenin ^{1}H NMR (D_2O, 600 MHz): δ 7.80 (d, 2, J = 9.06 Hz, H-2', H-6'), 6.88 (d, 2, J = 9.06 Hz, H-
223 3', H-5'), 6.47 (s, 1, H-3), 6.27 (d, 1, J = 2.25 Hz, H-8), 6.07 (d, 1, J = 2.25 Hz, H-6). Negative ESI-
224 MS m/z: 269 $[M - H]^-$ calculated as 270.24 for $C_{15}H_{10}O_5$.

225 p-Cymene ^{1}H NMR (CD_3OD, 600 MHz): δ 7.34 (d, 2, J = 8.95 Hz), 7.12 (d, 2, J = 8.95 Hz), 2.79 (m,
226 1), 1.98 (s, 3), 0.99 (d, 6, J = 6.79 Hz). Negative ESI-MS m/z: 133 $[M - H]^-$ calculated as 134.21 for
227 $C_{36}H_{30}O_{16}$.

228 Rosmarinic acid ^{1}H NMR (CD_3OD, 600 MHz): δ 7.52 (d, 1, J = 15.90 Hz), 7.01 (d, 1, J = 1.91 Hz),
229 6.93 (dd, 1, J = 1.91, 8.21 Hz), 6.75 (d, 1, J = 8.21 Hz), 6.72 (d, 1, J = 1.91 Hz), 6.67 (d, 1, J = 8.21
230 Hz), 6.59 (dd, 1, J = 1.91, 8.21 Hz), 6.24 (d, 1, J = 15.90 Hz), 5.16 (dd, 1, J = 4.28, 8.51 Hz), 3.07 (dd,
231 1, J = 4.28, 14.39 Hz), 2.98 (dd, 1, J = 8.51, 14.39 Hz). Negative ESI-MS was performed on EtOAc
232 faction of *O. vulgare*, yielding m/z: 359 $[M - H]^-$ calculated as 360.31 for $C_{18}H_{16}O_8$.

233 Salvianolic acid B ^{1}H NMR (CD_3OD, 600 MHz): δ 6.94 (d, 1, J = 8.54 Hz), 6.91 (d, 1, J = 16.27 Hz),
234 6.86 (d, 1, J = 8.54 Hz), 6.85 (d, 1, J = 2.33 Hz), 6.84 (d, 1, J = 2.33 Hz), 6.82 (d, 1, J = 8.54 Hz), 6.81
235 (d, 1, J = 8.54 Hz), 6.74 (dd, 1, J = 8.54, 2.33 Hz), 6.70 (dd, 1, J = 8.54, 2.33 Hz), 6.35 (d, 1, J = 8.54
236 Hz), 6.17 (d, 1, J = 2.33 Hz), 6.00 (d, 1, J = 8.54, 2.33 Hz), 5.90 (d, 1, J = 5.82 Hz), 5.79 (d, 1, J =
237 16.27 Hz), 4.93 (dd, 1, J = 10.10, 3.54 Hz), 4.83 (dd, 1, J = 12.11, 3.94 Hz), 4.33 (d, 1, J = 5.82 Hz),
238 3.04 (dd, 2, J = 14.44, 3.54 Hz), 2.92 (dd, 2, J = 14.70, 12.11 Hz), 2.85 (dd, 2, J = 14.44, 10.10 Hz),
239 2.49 (dd, 2, J = 14.70, 3.94 Hz). Negative ESI-MS m/z: 717 $[M - H]^-$ calculated as 718.6 for
240 $C_{36}H_{30}O_{16}$.

241 Thymol ^{1}H NMR (CD_3OD, 600 MHz): δ 6.9 (d, 1, J = 7.55 Hz), 6.59 (d, 1, J = 1.23 Hz), 6.56 (dd, 1, J
242 = 1.23, 7.55 Hz), 2.73 (m, 1), 2.10 (s, 3), 1.16 (d, 6, J = 6.95 Hz). Negative ESI-MS was performed on
243 $CHCl_3$ faction of *O. onites*, yielding *m*/*z*: 149 $[M - H]^-$ calculated as 150.22 for $C_{10}H_{14}O$.

244 **3. Results and Discussion**

245 **3.1 Discrimination between *Origanum* species**

246 The samples analyzed in this work were provided by different companies, who, in turn, obtained them
247 from third parties. The samples were under evaluation by the companies in order to decide whether to
248 buy and put them on the market or not. These samples were subjected both to MA and ^{1}H NMR
249 profiling. Species, percentage of stem, degree of total impurity, and percentage of specific
250 contaminants were firstly assessed by MA, as usually done by the companies in order to verify the
251 quality and the authenticity of the samples (Table 1).

252 The most evident morphological difference between *O. vulgare* and *O. onites* is the shape of their
253 calices, *O. vulgare* is characterized by actinomorphic calix (radial symmetry), with five equal lobes
254 (teeth), while *O. onites* presents a cone shaped zygomorphic calyx (with only one plane of symmetry,
255 bilaterally symmetrical) open on one side (Fig. 1A) (Ietswaart, 1980). According to the MA analysis,
256 out of twenty-seven samples, twelve were *O. vulgare*, eleven *O. onites*, two samples resulted composed
257 by a mixture of the two species, and two were recognized as non-marketable species.

258 After recording the ^{1}H NMR spectra, a first overview of the results obtained was acquired by
259 performing Principal Components Analysis (PCA), using bucketed ^{1}H NMR spectra as *x* variables. The
260 results obtained by PCA analysis agreed with what observed by MA, except for the samples coming
261 from Sicily, thus they were excluded from the first PCA model developed (this result will be discussed
262 later in the text). Excluding all unknown species, sixteen components (PCs) maximized the explained

98.6% of the variance in the data set (given by R^2x(cum)), while the obtained Q^2(cum) was 89.6%, indicating very good predictability (Q^2 must be equal or higher than 50%). This PCA model facilitated the detection of similarities/differences among the metabolomic profiles of the samples. As highlighted by the score scatter plot (Fig. 1A), the model was able to make a distinction between the two marketable oregano species along the component t[1], and, consistently, the samples made up of the two species blended were placed in the middle of the plot (violet dots). On the other hand, the shift along the component t[2] was due to sample degree of purity, which decreases along positive t[2].

In order to facilitate the identification of the main biomarkers responsible for the differentiation between the two marketable species, an OPLS-DA (Orthogonal Partial Least Squares-Discriminant Analysis) model was build (Fig. 1B), using as discriminant classes the two species of oregano (*O. onites* and *O. vulgaris*) previously identified by MA. Considering the objective of this analysis, the model was performed excluding from the data set the samples constituted by the mix of the two species, and the samples heavily adulterated with other herbs (percentage of contamination higher than 15%).

The OPLS-DA analysis find a perfect fit to the response using two components, with goodness of fit (R^2y(cum)) of 100% and goodness of prediction (Q^2(cum)) of 98.5%. Considering that these two parameters have to be as close as possible to 100% (Malzert-Fréon et al., 2010), the obtained model is interpretable and strongly predictable. This was further confirmed by permutation test (200 permutations) (Fig. S9), giving Q^2(cum) of 99% and intercept on *y* axis of *Q*-line of -0.55; while R^2(cum) was 99% and intercept on *y* axis of *R*-line was 0.19, and significance testing of the model based on ANOVA of the cross-validated residuals (CV-ANOVA) giving $p = 4.64 \times 10^{-17}$ and $F = 313$.

Loading plot and *S*-plot of OPLS-DA model show the relationships between *x* variables and sample classes, in this case, *x* were ^{1}H NMR spectral signals and classes were oregano species (Fig. S9). In

order to elucidate the molecular structure of the metabolites responsible for these ^{1}H NMR signals, they were isolated by NMR-guided fractionation and subjected to 1D, 2D-NMR and ESI-MS analysis. After this procedure, salvianolic acid B was found to be a biomarker of O. *vulgare,* while the flavonoid apigenin and the terpene p-cymene resulted the biomarkers of *O. onites* (Fig. 2 and Fig. S10). To the best of our knowledge, this is the first detailed report of the main biomarkers useful to distinguish between these two oregano species. Although also Blake et al. 2016 developed a method to detect adulteration in oregano samples through both FTIR and LC-HRMS untargeted analysis, their results are only at a pattern level, while no biomarkers identification was reported in their work.

The significance of the identified biomarkers was confirmed by their variable influence on projection (VIP) scores: p-cymene (VIP=1.60; calculated for signal at δ 0.98); apigenin (VIP=1.58; δ 6.51); salvianolic acid B (VIP=1.67; δ 6.35).

Other aromatic signals (number 15 in Fig. 2) emerged from the OPLS-DA as characteristic of *O. onites*, however, the metabolite(s) responsible for these signals was not fully characterized in this work.

Besides the possibility to discriminate between *O. vulgare* form *O. onites*, metabolomics also individuated non-marketable samples. This was highlighted when the two samples, classified by MA as unknown plant material (thus non-marketable), were added to the data matrix and processed by PCA. These samples resulted, in fact, outliers (Fig. 1C), since their phytochemical profile was completely different from the marketable oregano species. In this specific case, the non-marketable samples were lacking the main components of the essential oil (among which thymol), and they showed a different aromatic pattern, with high content of amino acid tyrosine, which was not detected in marketable oregano species (Fig. S11). Conversely, rosmarinic acid was detected both in 'true oregano' and in these non-marketable samples.

As above described, the results obtained by PCA and successively by OPLS-DA were consistent with those obtained by MA, making reliable the discrimination of oregano species based on their ^{1}H NMR profiles. However, in this study, the results of MA and NMR-metabolomics disagreed on three samples (all three coming from Sicily), which were classified by MA as *O. vulgare* (based on morphological traits), these samples resulted extremely different from both *O. vulgare* and *O. onites* when analyzed by NMR-based PCA (Fig. 3A). Thus, according to metabolomics approach, these samples should be considered non-marketable. In particular, instead of thymol (which was not detected) they were enriched in carvacrol, which was not found in the NMR profiles of marketable oregano species (Fig. 3B-C). These two positional isomers were easily distinguished by ^{1}H NMR spectroscopy, in fact, the signal of the methyl group (in *para* position to the isopropyl group on the aromatic ring) of carvacrol is more downfield shifted (δ 2.22) compared to the spectrum of its positional isomer thymol (δ 2.14) (Fig. 3C). This spectral feature is due to the *β*-effect of the hydroxyl group. Conversely, the spectral signal of the same substituent in thymol resulted de-shielded, since the hydroxyl group decreases electron density due to smaller inductive and resonance effect.

These samples are likely a Sicilian subspecies of *O. vulgare* (or a chemotype), hardly recognizable by dichotomous keys. This result highlights the potentialities of NMR-based metabolomics approach for oregano quality control.

3.2 Assessment of samples degree of purity

Two independent OPLS models were built (one for each species of marketable oregano), using as *y* variable the percentage of total impurity found in the samples by MA (Fig. 4A-B).

For *O. vulgare*, the model was fitted by six components (R^2x(cum) = 90.4%, R^2y(cum) = 96.9%, and Q^2(cum) = 91.1%). Its predictability was confirmed by R^2(cum) and Q^2(cum) obtained by permutation test (200 permutations), which were 96.9% and 91.1%, respectively, while intercept on *y* axis of *Q*-line

was -1.38, of *R*-line intercept was 0.696. CV-ANOVA *F* and *p* were 9.11 and 0.002, respectively. The value of R^2 obtained in the observed *vs* predicted plot was 0.9695, indicating a consistency between prediction based on NMR profiling and observations made by MA (Fig. 4A). However, one sample (BO19UOR in Table 1) was considered less pure than what was established by MA, this result might be explained by the content of stem (found in this sample), which likely lower the amount of the biomarkers, leading the model to classify it as less pure.

For *O. onites*, the model was fitted by four components (R^2x(cum) = 79.2%, R^2y(cum) = 98.8%, and Q^2(cum) = 96.1%). Its predictability was confirmed by R^2(cum) and Q^2(cum) obtained by permutation test (200 permutations), which were 98.8% and 96%, respectively, while intercept on *y* axis of *Q*-line was -1.07, of *R*-line intercept was 0.478, and $F = 39.85$ and $p = 6.5 \times 10^{-8}$ by CV-ANOVA. The value of R^2 obtained for observed *vs* predicted plot was 0.9883, indicating, also in this case, a consistency between NMR profiling and MA results, with some discrepancies especially at very low level of impurity (Fig. 4B, S12).

The OPLS models, built for the two *Origanum* species, individuated thymol and rosmarinic acid as important biomarkers of purity. Thymol had a VIP coefficient (calculated for the signal at δ 2.14) of 2.19 for *O. onites* and 2.41 for *O. vulgare*. Rosmarinic acid VIP coefficient (δ 7.11) was 1.69 and 1.60 for *O. onites* and *O. vulgare*, respectively. These two compounds were easily recognizable by ^{1}H NMR profiling, however their molecular structures were confirmed by further NMR and ESI-MS analysis performed on pre-purified fractions obtained by liquid-liquid partition.

Thymol is one of the main components of oregano essential oil (Khan et al., 2019), and rosmarinic acid is reported to be the main phenolic acid found in oregano extracts (Ozkan et al., 2010). In addition to be important for oregano flavor, thymol is also relevant for its numerous biological activities (Dheer et al.,

353 2019). Rosmarinic acid is another important bioactive metabolite of oregano, especially for its
354 antioxidant potential (Guitard et al., 2016; Villalva et al., 2018).

355 Besides thymol and rosmarinic acid, p-cymene (VIP 1.02; δ 0.98) and apigenin (VIP 0.85; δ 6.51) were
356 also relevant to predict *O. onites* degree of purity. Conversely, salvianolic acid B (VIP 0.5; δ 6.35) was
357 not strongly correlated to *O. vulgare* degree of purity (Fig. S10 reports the molecular structures of all
358 the compounds mentioned).

359

360 **3.3 Detection of *Cistus creticus* impurity**

361 Cistus was the most frequent adulterant identified by MA from its characteristic trichomes (Fig. 5A).
362 The adulteration of oregano with cistus, specifically *Cistus creticus* (ex. *incanus*) L., has been also
363 reported by Marieschi et al., 2010. In order to get the adequate model for the prediction of cistus
364 contamination, samples with specific percentage of cistus were prepared in laboratory, spiking *O.*
365 *onites* and *O. vulgare* samples with diverse cistus percentage (from 2 to 60 %) (Fig. 5A). The NMR-
366 metabolomic profile of these samples was added to the dataset, previously constituted only by the
367 samples provided by the companies. An OPLS model was built using as *y* variable the percentage of
368 cistus (found in the samples by MA, and known in the samples prepared in laboratory). The model was
369 fitted by four components, with R^2x(cum) = 76.9 %, R^2y(cum) = 93.2%, Q^2(cum) = 88.9%, and it was
370 further validated by R^2(cum) and Q^2(cum) obtained by permutation test (200 permutations) (Fig. S13),
371 which were 93.2% and 88.9%, respectively, *Q*-line intercept on *y* axis was -0.518 and *R*-line intercept
372 was 0.198, while F = 47.2 and p = 6.77 x 10^{-20} given by CV-ANOVA. This OPLS model showed a
373 correlation between samples metabolomic profiles and the used *y* variable. The analysis of the *S*-plot
374 (Fig. 5B) highlighted a correlation between percentage of cistus contamination and the intensity of a
375 specific ^{1}H NMR bucket (from δ 3.45 to 3.49; VIP = 5.16). In fact, the spectral signal resonating at δ

3.47 was present only in ^{1}H NMR spectra of contaminated samples. *J-res* experiment confirmed that this signal was a singlet (Fig. S14). Hence, a sample of *Cistus creticus* L. was harvested and analyzed by ^{1}H NMR profiling, exhibiting a prominent signal at δ 3.47, which confirmed the result given by the OPLS model (Fig. 5C). Through a pre-purification procedure, performed on cistus extract, it was obtained a fraction containing a compound, whose positive ESI-MS yield *m*/*z*: 195 $[M + H]^{+}$, calculated as 194.18 for $C_7H_{14}O_6$ and negative ESI-MS was *m*/*z*: 193 $[M – H]^{-}$, corresponding to pinitol molecular weight. This cyclic alcohol, in fact, presents a singlet in NMR spectrum at δ 3.47, due to the methoxy group in position 3 of the ring.

4. Conclusions

Combining results obtained from morphological analysis and ^{1}H NMR metabolomics, it was possible to rapidly identify non-marketable species of oregano, as well as to discriminate between the two marketable ones (*Origanum vulgare* and *O. onites*). Apigenin and p-cymene were the main biomarkers of *O. onites*, while salvianolic acid B of *O. vulgare*. The developed PCA model was also able to individuate samples made up of a blend of these two species.

Moreover, sample degree of purity could be predicted by OPLS. Thymol and rosmarinic acid were very important indicators to predict general oregano purity, together with the specie-specific apigenin, p-cymene. A further model was built to detect the degree of cistus contamination.

Metabolomics proved to be a valuable approach for oregano quality control, giving information about oregano species and phytochemical composition of samples, which is an important data, also correlated to the organoleptic properties. The obtained results encourage the use of metabolomics to make predictions regarding oregano quality on the basis of samples NMR profile, supporting also the possibility of building NMR motabolomics-based databases, to be used worldwide for oregano quality control, giving a significant number of information in a considerably short period of time.

Acknowledgment

Thanks are due to Dr. David Baldo from Electron Microscopy Laboratory (DISTAL - Plant Pathology, University of Bologna, V.le Fanin 40, 40127, Bologna, Italy) for the picture with Scanning Electron Microscopy.

Thanks are due to Dr. Luca Zuppiroli from Department of Industrial Chemistry **(**University of Bologna, Viale del Risorgimento, 4, 40136 Bologna, Italy) for technical assistance with ESI-MS experiments.

Conflict of interest

None

References

American spice trade. Association, ASTA cleanliness specifications for spices, seeds and herbs, http://www.ina.or.id/knoma-hpsp/herb/HPSP-12-Ollop (Accessed September 9, 2020)

Anđelković, B., Vujisić, L., Vučković, I., Tešević, V., Vajs, V., & Gođevac, D. (2017). Metabolomics study of Populus type propolis. *Journal of Pharmaceutical and Biomedical Analysis*, *135*, 217–226. https://doi.org/10.1016/j.jpba.2016.12.003

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

Drabova, L., Alvarez-Rivera, G., Suchanova, M., Schusterova, D., Pulkrabova, J., Tomaniova, M., Kocourek, V., Chevallier, O., Elliot, C., & Hajslova, J. (2019). Food fraud in oregano: Pesticide

residues as adulteration markers. *Food Chemistry*, *276*, 726–734. https://doi.org/10.1016/j.foodchem.2018.09.143

Dheer, D., Singh, D., Kumar, G., Karnatak, M., Chandra, S., Prakash Verma, V., & Shankar, R. (2019). Thymol Chemistry: A Medicinal Toolbox. *Current Bioactive Compounds*, *15*(5), 454–474. https://doi.org/10.2174/1573407214666180503120222

Emwas, A.H., Roy, R., McKay, R.T., Tenori, L., Saccenti, E., Gowda, G.A., Raftery, D., Alahmari, F., Jaremko, L., Jaremko, M. and Wishart, D.S. (2019). NMR spectroscopy for metabolomics research. *Metabolites*, *9*(7), 123. https://doi.org/10.3390/metabo9070123

European Pharmacopoeia (2017). (9th edition) pp.1464–1466

European Spice Association, Quality Minima Document (2015). http://www.esa-spices.org/index-esa.html/publications-esa (Accessed September 9, 2020)

European Spice Association, List of culinary herbs and spices (2018). https://www.esa-spices.org/index-esa.html/publications-esa (Accessed September 9, 2020)

Exarchou, V., Godejohann, M., van Beek, T. A., Gerothanassis, I. P., & Vervoort, J. (2003). LC-UV-solid-phase extraction-NMR-MS combined with a cryogenic flow probe and its application to the identification of compounds present in Greek oregano. *Analytical Chemistry*, *75*(22), 6288–6294. https://doi.org/10.1021/ac0347819

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

ISO Online browsing platform. https://www.iso.org/obp/ui/#iso:std:29207:en (Accessed September 9, 2020)

Guitard, R., Paul, J. F., Nardello-Rataj, V., & Aubry, J. M. (2016). Myricetin, rosmarinic and carnosic acids as superior natural antioxidant alternatives to α-tocopherol for the preservation of omega-3 oils. *Food Chemistry*, *213*, 284–295. https://doi.org/10.1016/j.foodchem.2016.06.038

Ietswaart, J. H. (1980). *A taxonomic revision of the genus Origanum (Labiatae)* (Vol. 4). The Hague: Leiden University Press.

Khan, M., Khan, S. T., Khan, M., Mousa, A. A., Mahmood, A., & Alkhathlan, H. Z. (2019). Chemical diversity in leaf and stem essential oils of *Origanum vulgare* L. and their effects on microbicidal activities. *AMB Express*, *9*(1), 176. https://doi.org/10.1186/s13568-019-0893-3

Malzert-Fréon, A., Hennequin, D., & Rault, S. (2010). Partial least squares analysis and mixture design for the study of the influence of composition variables on lipidic nanoparticle characteristics. *Journal of Pharmaceutical Sciences*, *99*(11), 4603–4615. https://doi.org/10.1002/jps.22177

Mandrone, M., Scognamiglio, M., Fiorentino, A., Sanna, C., Cornioli, L., Antognoni, F., Bonvicini, F. & Poli, F. (2017). Phytochemical profile and α-glucosidase inhibitory activity of Sardinian *Hypericum scruglii* and *Hypericum hircinum*. *Fitoterapia*, *120*, 184–193. https://doi.org/10.1016/j.fitote.2017.06.020

Mandrone, M., Antognoni, F., Aloisi, I., Potente, G., Poli, F., Cai, G., Faleri, C., Parrotta, L., & Del Duca, S. (2019). Compatible and incompatible pollen-styles interaction in *Pyrus communis* L. show different transglutaminase features, polyamine pattern and metabolomics profiles. *Frontiers in Plant Science*, *10*, 741. https://doi.org/10.3389/fpls.2019.00741

Mandrone, M., Chiocchio, I., Barbanti, L., Tomasi, P., Tacchini, M. and Poli, F., 2021. Metabolomic study of sorghum (*Sorghum bicolor*) to interpret plant behavior under variable field conditions in view

461 of smart agriculture applications. *Journal of Agricultural and Food Chemistry*, *69*(3), 1132-1145.
462 https://doi.org/10.1021/acs.jafc.0c06533

463 Marieschi, M., Torelli, A., Poli, F., Sacchetti, G., & Bruni, R. (2009). RAPD-based method for the
464 quality control of Mediterranean oregano and its contribution to pharmacognostic techniques. *Journal*
465 *of Agricultural and Food Chemistry*, *57*(5), 1835–1840. https://doi.org/10.1021/jf8032649

466 Marieschi, M., Torelli, A., Poli, F., Bianchi, A., & Bruni, R. (2010). Quality control of commercial
467 Mediterranean oregano: Development of SCAR markers for the detection of the adulterants *Cistus*
468 *incanus* L., *Rubus caesius* L. and *Rhus coriaria* L. *Food Control*, *21*(7), 998–1003.
469 https://doi.org/10.1016/j.foodcont.2009.12.018

470 Marieschi, M., Torelli, A., Bianchi, A., & Bruni, R. (2011). Development of a SCAR marker for the
471 identification of Olea europaea L.: A newly detected adulterant in commercial Mediterranean oregano.
472 *Food Chemistry*, *126*(2), 705–709. https://doi.org/10.1016/j.foodchem.2010.11.030

473 Marieschi, M., Torelli, A., Bianchi, A., & Bruni, R. (2011). Detecting Satureja montana L. and
474 Origanum majorana L. by means of SCAR–PCR in commercial samples of Mediterranean oregano.
475 *Food Control*, *22*(3-4), 542–548. https://doi.org/10.1016/j.foodcont.2010.10.001

476 Ozkan, G., Baydar, H. and Erbas, S. (2010). The influence of harvest time on essential oil composition,
477 phenolic constituents and antioxidant properties of Turkish oregano (*Origanum onites* L.). *Journal of*
478 *the Science of Food and Agriculture*, *90*(2), 205–209. https://doi.org/10.1002/jsfa.3788

479 Pontes, J. G. M., Brasil, A. J. M., Cruz, G. C., de Souza, R. N., & Tasic, L. (2017). NMR-based
480 metabolomics strategies: plants, animals and humans. *Analytical Methods*, *9*(7), 1078–1096.
481 https://doi.org/10.1039/C6AY03102A

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

Sun, Y., Zhu, H., Wang, J., Liu, Z. and Bi, J., 2009. Isolation and purification of salvianolic acid A and salvianolic acid B from *Salvia miltiorrhiza* by high-speed counter-current chromatography and comparison of their antioxidant activity. *Journal of Chromatography B*, *877*(8-9), pp.733–737. https://doi.org/10.1016/j.jchromb.2009.02.013

Salomé-Abarca, L. F., Mandrone, M., Sanna, C., Poli, F., van der Hondel, C. A., Klinkhamer, P. G., & Choi, Y. H. (2020). Metabolic variation in *Cistus monspeliensis* L. ecotypes correlated to their plant-fungal interactions. *Phytochemistry*, *176*, 112402. https://doi.org/10.1016/j.phytochem.2020.112402

Scognamiglio, M., D'Abrosca, B., Esposito, A., & Fiorentino, A. (2015). Chemical composition and seasonality of aromatic mediterranean plant species by NMR-based metabolomics. *Journal of Analytical Methods in Chemistry*, *2015*. https://doi.org/10.1155/2015/258570

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

Van Ruth, S. M., Luning, P. A., Silvis, I. C., Yang, Y., & Huisman, W. (2018). Differences in fraud vulnerability in various food supply chains and their tiers. *Food Control*, *84*, 375–381. https://doi.org/10.1016/j.foodcont.2017.08.020

Villalva, M., Jaime, L., Aguado, E., Nieto, J. A., Reglero, G., & Santoyo, S. (2018). Anti-inflammatory and antioxidant activities from the basolateral fraction of Caco-2 cells exposed to a rosmarinic acid

503 enriched extract. *Journal of Agricultural and Food Chemistry*, *66*(5), 1167–1174.
504 https://doi.org/10.1021/acs.jafc.7b06008

505 Wadood, S. A., Boli, G., Xiaowen, Z., Hussain, I., & Yimin, W. (2020). Recent development in the
506 application of analytical techniques for the traceability and authenticity of food of plant origin.
507 *Microchemical Journal*, *152*, 104295. https://doi.org/10.1016/j.microc.2019.104295

508

Fig. 1 A) Morphological differences between *O. vulgare* e *O. onites* calix, and NMR-based PCA score scatter plot, *O. onites* (blue dots) is distinguished by *O. vulgare* (red dots) along the component t[1], samples composed by a mixture of both species (violet dots) are placed in the middle. **B) OPLS-DA score scatter plot using *O. onites* and *O. vulgare* as model classes** (blended samples and samples with percentage of contamination higher than 15% were excluded from this model). **C) PCA score scatter plot including two samples (green dots) resulted non-marketable by MA,** which are outliers. Each sample was analyzed in duplicate (blended samples were excluded from the dataset).

Fig. 2 ^{1}H NMR spectra of *O. vulgare* (on top) and *O. onites* (bottom). Extended spectral regions are shown on top of the figure. Numbers indicate the diagnostic signals of the main metabolites: 1 = p-cymene, 2 = thymol, 3 = alanine, 4 = quinic acid, 5 = acetic acid, 6 = malic acid, 7 = aspartic acid, 8 = sucrose, 9 = β-glucose, 10 = rosmarinic acid, 11 = α-glucose, 12 = salvianolic acid B, 13 = apigenin, 14 = formic acid, 15 = unknown compound.

Fig. 3 A) PCA score scatter plot where unknown samples (yellow dots) are discriminated from the two marketable oregano species. **B) Structures of thymol and carvacrol**, resulting the main compounds determining the discrimination of these samples **C) Aliphatic region of ^{1}H NMR spectra of marketable oregano**, thymol characteristic spectral signals are present only in the marketable species, while the unknown species presents carvacrol instead of thymol.

Fig. 4 OPLS score scatter plot (on top) and observed *vs* predicted plot (bottom). Percentage of total impurity of **A)** *O. vulgare* and **B)** *O. onites* was used as *y* variable. Samples were analyzed in duplicate. Samples constituted by unknown species or a mixture of the two marketable species were excluded from this analysis.

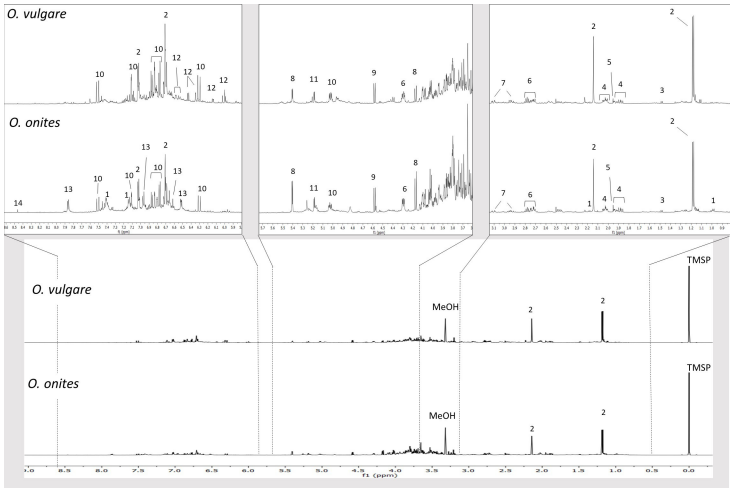
Fig. 5 A) Different trichomes of cistus (on the right) and oregano (on the left) and OPLS score scatter plot (*y* variable was percentage of cistus). The shift along component t[1] is due to oregano species: *O. onites* samples (dots) are placed on positive t[1], and *O. vulgare* (squares) are placed on negative t[1] of the model. Percentage of cistus contamination is given along the component t[0]. **B) OPLS *S*-Plot,** which indicates the ^{1}H NMR bin at δ 3.45-3.49 as strongly correlated to cistus contamination. **C) Extended ^{1}H**

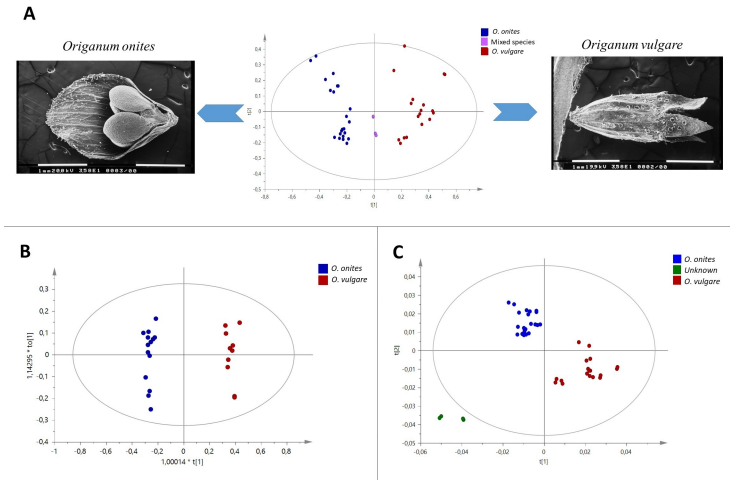
NMR spectral region of *Cistus creticus* showing an evident signal at δ 3.47, due to the methoxy group of pinitol.

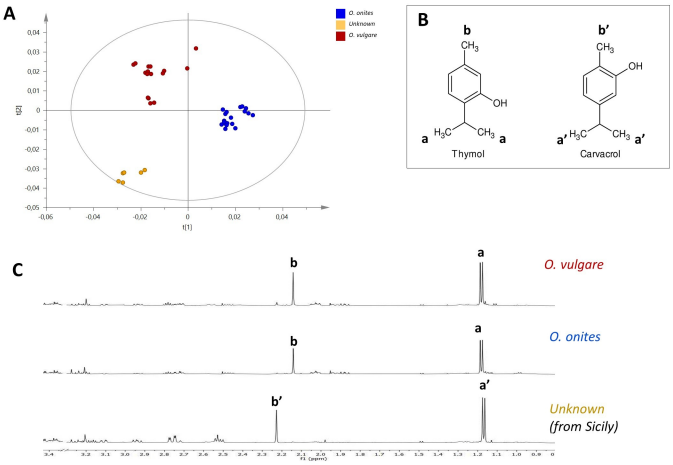
Table 1. Vouchers, origin, and results of the morphological analysis (MA) performed on the samples analyzed in this work. Samples were provided by different companies, who received, in turn, from third parties, who give the information related to sample origin. Samples defined as 'unknown' species have to be considered non-marketable as oregano. Samples coming from Sicily were identified as *O. vulgare* by MA while as 'unknown' species by NMR-metabolomics (this information has been reported in the Table).

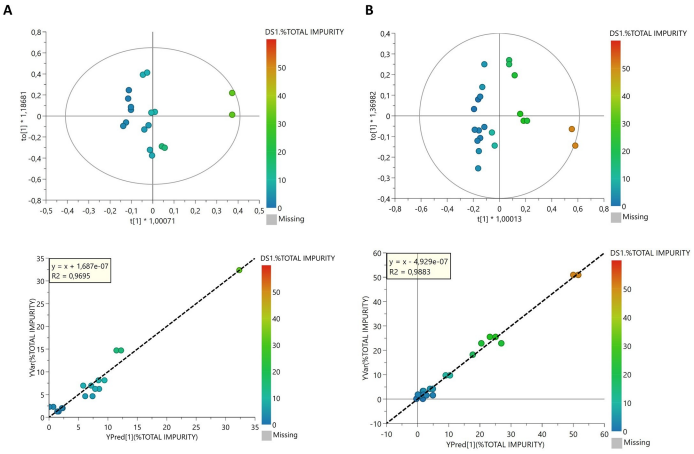
Vouchers	Origin	total oregano (%)	leaf and flower (%)	stem (%)	total impurity (%)	other Labiatae (%)	other plants (%)	cistus (%)	non-plant material (%)	olive leaves (%)	Species
BO 1A001	Turkey	85.28	81.06	4.22	14.72	13.62	1.1	0	0	0	*O. onites* + *O.vulgare*
BO 2BOR	Sicily	97.68	94.40	3.28	2.32	0	2.32	0	0	0	*O. vulgare* (by MA); unknown (by metabolomics)
BO 3COR	Turkey	96.6	87.56	9.04	3.4	2.6	0.8	0	0	0	*O. onites*
BO 4DOR	Turkey	90.16	83.20	6.96	9.84	9.1	0.44	0	0.3	0	*O. onites*
BO 5EOR	Peru	99.72	97.40	2.32	0.28	-	0	0	0.28	0	unknown
BO 6FOR	Turkey	99.88	96.60	3.28	0.12	0	0	0	0.12	0	*O. onites*
BO 7GOR	Turkey	98.4	94.12	4.28	1.6	0	1.6	0	0	0	*O. onites*
BO 8HOR	Turkey	98.26	91.58	6.68	1.74	1.7	0.04	0	0	0	*O. onites*
BO 9IOR	Not declared	49.14	43.44	5.70	50.86	0.52	0	50.34	0	0	*O. onites*
BO 10LOR	Turkey	95.82	91.82	4.00	4.18	4.02	0	0.16	0	0	*O. onites*
BO 11MOR	Turkey	98.58	90.40	8.18	1.42	0.7	0.2	0.52	0	0	*O. onites*
BO 12NOR	Turkey	81.7	75.14	6.56	18.3	5.9	10.54	1.86	0	10.54	*O. onites*
BO 13OOR	Turkey	74.46	69.88	4.58	25.54	7.6	7.42	10.52	0	7.42	*O. onites*
BO 14POR	Turkey	77.18	74.14	3.04	22.82	3.98	7.6	11.24	0	7.6	*O. onites*
BO 15QOR	Turkey	93.82	88.12	5.70	6.18	4.7	1.48	0	0	0	*O. vulgare*
BO 16ROR	Turkey	67.64	62.94	4.70	32.36	0	1.4	30.96	0	0	*O. vulgare*

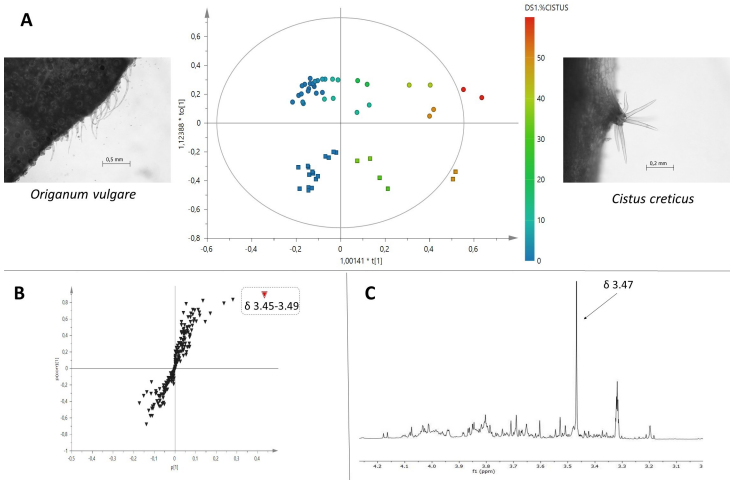
BO 17SOR	Albania	85.26	75.76	8.83	14.74	6.04	8.53	0.16	0	0	*O. vulgare*
BO 18TOR	Not declared	95.36	86.52	8.60	4.64	2.4	2.1	0	0	0	*O. vulgare*
BO 19UOR	Albania	98	86.70	11.30	2	1	1	0	0	0	*O. vulgare*
BO 20VOR	Peru	99.94	97.24	2.70	0.06	-	0	0	0.06	0	unknown
BO 21ZOR	Albania	98.74	90.12	8.62	1.26	0.74	0.52	0	0	0	*O. vulgare*
BO 22AAOR	Turkey	98.96	88.96	10.00	1.04	0.18	0.86	0	0	0	*O. onites* + *O.vulgare*
BO 23ABOR	Not declared	96.66	89.86	6.80	2.34	0	2.34	0	0	0	*O. vulgare*
BO 24ACOR	Sicily	100	97.42	2.58	0	0	0	0	0	0	*O. vulgare* (by MA); unknown (by metabolomics)
BO 25ADOR	Not declared	91.8	86.08	5.72	8.2	7.12	0	0	0	0	*O. vulgare*
BO 26AEOR	Not declared	93.02	80.88	12.14	6.98	6.18	0	0	0	0	*O. vulgare*
BO 27AFOR	Sicily	97.48	95.10	2.38	2.52	0	2.2	0	0.3	0	*O. vulgare* (by MA); unknown (by metabolomics)











Highlights

- Oregano is one of the most counterfeited spices
- Nuclear Magnetic Resonance-based metabolomics detect frauds in oregano samples
- Metabolomics provided species-specific biomarkers of *Origanum onites* and *O. vulgare'*
- Metabolomics allowed to predict oregano degree of purity

Declaration of interests

☒ 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.

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