Untargeted and Targeted Fingerprinting of Extra Virgin Olive Oil Volatiles by Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry: Challenges in Long-Term Studies

Untargeted and Targeted (*UT*) Fingerprinting of Extra Virgin Olive Oil Volatiles by Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry: Challenges in Long-Term Studies

Federico Stilo[1], Erica Liberto[1], Stephen E. Reichenbach[2,3], Qingping Tao[3], Carlo Bicchi[1] and Chiara Cordero[1*]

[1]Dipartimento di Scienza e Tecnologia del Farmaco, Università degli Studi di Torino, Turin, Italy

[2]Computer Science and Engineering Department, University of Nebraska, Lincoln, NE, USA

[3]GC Image LLC, Lincoln, NE, USA

*Corresponding author:

Dr. Chiara Cordero - Dipartimento di Scienza e Tecnologia del Farmaco, Università di Torino, Via Pietro Giuria 9, I-10125 Torino, Italy – e-mail: chiara.cordero@unito.it; phone: +39 011 6707172; fax: +39 011 2367662

Abstract

Comprehensive two-dimensional gas chromatography coupled with mass spectrometric detection (GC×GC-MS) offers an information-rich basis for effective chemical fingerprinting of food. However, GC×GC-MS yields 2D-peak patterns (i.e., sample 2D fingerprints) whose consistency may be affected by variables related to either the analytical platform or to the experimental parameters adopted for the analysis.

This study focuses on the complex volatile fraction of extra-virgin olive oil and addresses 2D-peak patterns variations, including MS signal fluctuations, as they may occur in long-term studies where pedo-climatic, harvest year or shelf-life changes are studied.

2D-pattern misalignments are forced by changing chromatographic settings and MS acquisition. All procedural steps, preceding pattern recognition by template matching, are analyzed and a rational work-flow defined to accurately re-align patterns and analytes metadata.

Signal-to-noise ratio (SNR) detection threshold, reference spectra and similarity match factor threshold are critical to avoid false-negative matches. Distance thresholds and polynomial transform are key parameters for effective template transform. In targeted analysis (supervised work-flow) with optimized parameters method accuracy achieves 92.5% (i.e., % of true-positive matches) while for combined untargeted and targeted (*UT)* fingerprinting (unsupervised work-flow), accuracy reaches 97.9 %. Response normalization also is examined, evidencing good performance of multiple internal standard normalization that effectively compensates for discriminations occurring during injection of highly volatile compounds. The resulting work-flow is simple, effective, and time efficient.

Key words

Comprehensive two-dimensional gas chromatography coupled to time-of-flight mass spectrometry; extra-virgin olive oil volatiles; template matching; combined untargeted and targeted (UT) fingerprinting; data alignment in long-term studies

Introduction

Comprehensive two-dimensional gas chromatography (GC×GC) is one of the most informative separation techniques for chemical characterization of complex fractions of volatiles from food [1–3]. It enables highly effective fingerprinting [4] and, when combined with mass spectrometric detection (GC×GC-MS), it has the intrinsic potential to provide a detailed profiling, giving access to higher level information encrypted in complex patterns of volatiles, for example: sample origin, technological signature, and aroma[2,5–8].

Each analytical run produces dense and multidimensional data, so that elaboration and interpretation of chemical information is challenging. Moreover, 2D-peak patterns representing the sample 2D fingerprint, are defined by a series of variables also related to the analytical platform and to the experimental parameters adopted for the analysis. The choice of flow modulation instead of thermal/cryogenic modulation, MS detection by fast scanning quadrupoles *vs.* time-of-flight MS, low-resolution MS *vs.* high-resolution MS as well as GC×GC stationary phase combination, columns lengths and diameters, carrier gas linear-velocities, modulation period (P_M) and oven temperature programming greatly impact on 2D-patterns signature and informational density.

Although most of these parameters, once fixed after method development and optimization, are kept constant (e.g., column set-up, carrier gas flows, and modulation parameters) or can be standardized as the MS tuning and optimization, some others represent a source of random variability that must be considered when fingerprinting and pattern recognition studies extend over time and/or across different platforms.

For mono-dimensional (1D) GC-MS applications, possible strategies for chromatographic alignment and data normalization are: (*a*) linear retention indexing (van Den Dool and Kratz or Kovats indices) or retention time locking methods based on pressure/flow adjustments (i.e., retention time locking [9–11]) to accurately locate target analytes along the analytical run; (*b*) chromatographic realignment [12–14]; (*c*) internal standardization for response normalization by single or multiple Internal Standards (IS) addition; and (*d*) external standard normalization by adopting single or multiple External Standards (ES). These strategies are effective and routinely adopted in peak-features based applications[4]. However, for GC×GC, these 1D-GC

strategies could be ineffective especially for retention inconsistencies that result from two, almost independent, separation steps. On the other hand, the peculiar nature of 2D-peak patterns offers the possibility of exploiting pattern recognition algorithms for prompt and effective fingerprinting. So, strategies for pattern alignment and normalization are needed.

Pattern recognition approaches based on peak-region features[4], implemented with the smart template concept in commercial software[15], use different transform functions capable of recognizing 2D peak patterns based on retention times coordinates, and establishing correspondences between 2D-peaks, or 2D-peak-regions, from a *reference* pattern to those in an *analyzed* pattern even in presence of retention times shifts[15–17] and/or when severe misalignments occur because of different modulation principles [18,19].

Pattern correspondences are at the basis of the re-alignment of untargeted/targeted 2D-peaks or 2D-peak-regions across a samples-set to enable fingerprinting investigations[2]. Furthermore, the specificity and reliability of pattern matching can be improved by including constraint functions, operating on the third dimension of the data, i.e., the MS signature. Typical functions are those that limit positive correspondences to 2D-peaks with spectral similarity match above a certain threshold or, more simply, for 2D-peaks that comply for specific m/z relative ratios between informative fragments of the spectrum.

In a scenario where food chemical fingerprinting has to be extended over long time-frames, as for example to cover different harvest years or shelf-life modifications, strategies and tools for data re-alignment and normalization are necessary[20,21] together with more rational strategies and intuitive operative protocols/work-flows to guide analysts over the actual limits posed by analytical data misalignment.

This study addresses 2D-peak patterns variations occurring in long-term studies that might affect the effectiveness of combined untargeted and targeted fingerprinting (*UT* fingerprinting). The procedural steps preceding template matching function are analyzed to define a rational work-flow enabling consistent pattern recognition. These steps, supervised by the analyst, aim at selecting the key-parameters to generate a targeted template for marker 2D-peaks or a *reliable* or *consensus* template covering the sample chemical dimensionality [5], in particular: (*a*) 2D-peaks response thresholds (Signal-to-Noise ratio - SNR and Volume-to-Noise ratio – VNR); (*b*) MS reference signature to be used for spectral constraints; (*c*) MS similarity thresholds;

and (*d*) template matching transforms are considered, with their settings varied to compensate for 2D-peak patterns variability so as to achieve effective and comprehensive chemical fingerprinting.

The complex 2D patterns of volatiles from Extra Virgin Olive oil of different quality are studied. Volatiles are sampled by headspace solid-phase microextraction (HS-SPME) and analyzed by GC×GC-MS in a platform equipped with a loop-type thermal modulator adopting L_2 cryogenics. The 2D-patterns are obtained in a one-year study during which misalignments and inconsistencies were introduced by varying column lengths and restrictions, modulation period (P_M), and operating with the time-of-flight (TOF) MS with different optimized parameters.

Materials and methods

Reference compounds and solvents

Pure reference standards of α- and β-thujone and methyl-2-octynoate used as Internal Standards (ISs), *n*-alkane standards (n-C7 to n-C25) used for linear retention index (I^T_S) calibration and pure reference compounds for targeted analytes identity confirmation were supplied by Merk (Sigma-Aldrich srl Italy, Milan, Italy). Cyclohexane (HPLC grade) for n-alkane standards and pure dibutyl phthalate for ISTDs working solutions were from Merk.

Olive Oil samples

Extra Virgin Olive oils (EVO oils), supplied by the University of Granada (Spain), Prof. Luis Cuadros-Rodríguez, were obtained from olives of the *Picual* cultivar, harvested in the regions of Granada *Altipiano* named *Baza* and *Benamaurel*, and grown under differing production and irrigation practices [7]. Samples were obtained by mixing olives from five different trees from the same plot in duplicate batches (A and B). Olives for oil production were collected at four different ripening stages: November 10-12, 2014; November 24-28, 2014; December 16-17, 2014; and January 12-15, 2015, and classified by oil quality (Extra Virgin - EVOO, Virgin-VOO or *Lampante- LOO*). Samples acronyms and characteristics are summarized in Table 1. Oil qualifications were by a certified laboratory (ISO 17025:2018) [7] and according to Commission Regulation

(EEC) No 2568/91 of 11 July 1991. Some quality indices are reported in Table 1, including the sensory panel test results.

Headspace solid-phase microextraction sampling devices and conditions

Volatiles were extracted from samples by headspace (HS) solid-phase microextraction (SPME). DVB/CAR/PDMS d_f 50/30 µm 1 cm length fiber (Supelco, Belle-fonte, PA, USA) was chosen based on previous studies [8] and conditioned before use as recommended by the manufacturer. The ISs were pre-loaded onto the SPME device [22,23], before sampling, in a 20 mL headspace vial a 5.0 µL of α/β-thujone and methyl-2-octynoate at 100 mg L^{-1} in dibutyl phtalate. ISs were equilibrated at 40°C and pre-loaded by exposing the SPME device for 5 min.

Sampling was carried out on 0.100 ± 0.005 g of oil samples precisely weighed in 20 mL headspace vials. The very low amount of sample was chosen to comply with HS linearity conditions for most of the key-analytes responsible of samples discrimination [7]. Sampling was at 40°C for 60 min. After extraction, the SPME device was introduced into the S/SL injection port of the GC×GC system kept at 260°C for 5 min. Each sample was analyzed in triplicate.

Instrument set-up and analysis conditions

GC×GC analyses were performed on an Agilent 7890 GC unit (Agilent Technologies, Wilmington DE, USA) coupled to a Markes BenchTOF-Select™ featuring Tandem ionization™ (Markes International, Llantrisant, UK). The GC transfer line was at 270°C. TOF MS tuning parameters were set for single ionization at 70 eV and for tandem ionization at 70 and 12 eV; the scan range was set at 35-350 m/z with a spectra acquisition frequency of 100 Hz for single eV and 50 Hz/channel for tandem ionization.

The system was equipped with a two-stage KT 2004 loop type thermal modulator (Zoex Corporation, Houston, TX) cooled with liquid nitrogen controlled by Optimode v2.0 (SRA Intruments, Cernusco sul Naviglio, Milan, Italy). Modulation periods (P_M) and hot jet pulse times are detailed in Table 2, along with other parameters. A Mass Flow Controller (MFC) reduced the cold-jet stream from 45% to 8% of the total flow with a linear function along the run duration. A fused silica capillary loop (1.0 m × 0.1 mm id) was used in the modulator slit.

The column set was configured as follows: ^{1}D SolGel-Wax column (100% polyethylene glycol; 30 m × 0.25 mm d_c × 0.25 µm d_f) from SGE Analytical Science (Ringwood, Australia) coupled with a ^{2}D OV1701 column (86% polydimethylsiloxane, 7% phenyl, 7% cyanopropyl; 1 m × 0.1 mm d_c × 0.10 µm d_f) from Mega (Legnano, Milan, Italy). A capillary restriction towards MS was used to generate a differential pressure-drop influencing actual carrier gas linear velocities along the column in *Set-up 2* (1 m × 0.1 mm d_c deactivated silica). The GC Split/Splitless (S/SL) injector port was set at 260°C and operated in split mode with a split ratio 1:20.

The carrier gas was helium at a constant nominal flow of 1.3 mL/min. The oven temperature programming was set as follows: 40°C (2 min) to 240°C (10 min) at 3.5°C/min. Carrier gas average linear velocities ($^{1}\bar{u}$ and $^{2}\bar{u}$), pressure settings, and hold-up times are reported in **Table 2** and were obtained by basic calculations with a reference temperature of 60°C.

For I_T^S determination, 1µL of the n-alkanes sample solution was injected with a split ratio 1:50.

Data were acquired by TOF-DS™ (Markes International, Llantrisant, UK) and processed by GC Image ver 2.8 (GC Image, LLC Lincoln, NE, USA).

UT fingerprinting work-flow

The distribution of detectable analytes over the 2D chromatographic space in a GC×GC separation is at the basis of pattern recognition based on the *smart template* concept [24]. The template is a pattern of 2D-peaks and/or graphic objects built over a *reference* image(s) (single or cumulative image [25]) and then used to recognize similar patterns of 2D-peaks in an *analyzed* image(s)[26]. Each template object (2D-peak and/or graphic) can carry various metadata including: compound chemical name, retention times, I^T_S, mass spectra, informative ions and their relative ratios, constraint functions to limit peak correspondences above certain thresholds, and qualifier functions.

In the presence of temporal inconsistencies and detector fluctuations *peak-region* features[27] are of great help. Peak-regions attempt to define one chromatographic region around each individual peak thereby achieving the one-feature-to-one-analyte selectivity but with greater robustness than can be achieved with single 2D-peak detection[28]. 2D-peaks and peak-regions are features adopted in the combined targeted and untargeted fingerprinting (*UT* fingerprinting) strategy[7,18,25,29].

UT fingerprinting establishes a group of *reliable* peaks, positively matched across *all* or *most-of* chromatograms in set[30], and then uses them to align chromatograms[16] for their combination into a single, composite chromatogram. From the composite chromatogram, 2D-peaks (i.e., the combination of the re-aligned responses in the 2D retention-times plane) are detected and their outlines are recorded to define peak-region objects. The set of reliable 2D-peaks and peak-regions objects are collected in the *feature* template, or *consensus* template*,* covering the whole sample-set variability and capable of cross-corresponding chemical feature patterns among samples.

To note, within all detected analytes, the sub-group of targeted compounds can be highlighted by completing their metadata fields (compound name, ion ratios, I^T_S) and computed together with untargeted features during the data processing.

A schematic of the UT fingerprinting process is illustrated in the Supplementary Material – Supplementary Figure 1 together with some details on targeted and untargeted 2D peaks and peak-regions.

Results and discussion

In spite of the great potential of GC×GC in exploiting the chemical dimensionality of olive oil volatile fraction, just a few studies are available in this field and none of them address challenges posed by long-time frame studies. Vaz Freire *et al.* [31] adopted an image-features approach[28] to investigate characteristic distributions of volatiles. An open-source image analysis software (Image J, National Institutes of Health) was used to extract detector response information from 2D regions over the separation space. By Principal Component Analysis (PCA) image-features with a high discrimination potential were selected and targeted profiling was then combined to locate known analytes within most informative 2D regions.

Studies aimed at defining geographical origin indicators or cultivar markers are those by Cajka et al.[32] who adopted GC×GC-TOF-MS to identify 44 compounds able to discriminate extra-virgin olive oils based on their different geographical origin and production year and by Lukić et al. [33] who applied a peak-features approach to reveal compositional differences between oils obtained by different olive cultivars and geographical areas. They considered, as potentially informative, both untargeted and targeted analytes as

they were extracted from the raw data-set on the basis of relative retention and spectral features. Magagna et al. [7] were the first who developed an integrated strategy for *UT* fingerprinting based on template matching, to define olive ripening indicators while Purcaro et al.[8] combined targeted and untargeted analysis to delineate chemical blueprints of olive oil aroma defects.

In this scenario, intuitive and easily applicable strategies to extend the breath of comprehensive profiling and fingerprinting of olive oil volatiles over wider time frames, e.g. over harvest years or across the shelf-life of a product, are of great interest especially in the perspective of collecting data for authenticity and typicity databases as in the case of the Italian *Violin* project[34] aimed at the valorisation of the Italian extra-virgin olive oil.

The next sections will illustrate the strategy adopted to generate misaligned patterns and the subsequent work-flow designed to re-align templates for targeted and untargeted peak-features covering the entire volatile fraction of extra-virgin olive oil. Results are critically discussed in term of accuracy (i.e., true positive matches) and data inter-batch transferability (i.e., response normalization).

Pattern misalignment challenges

Chromatographic pattern distortions and misalignments were induced, generating the worst-case scenario, by changing the following parameters: (*a*) columns were from different commercial batches; (*b*) a post-column restriction was added generating a pressure drop between column inlet and outlet, influencing carrier gas linear velocities in both analytical dimensions; (*c*) the P_M was set at 4 or 3.5 s generating an absolute ^{2}D retention misalignment. All the other parameters, carrier gas nominal flow, oven temperature programming and injection conditions were kept constant. Analytical conditions for the two resulting set-ups (i.e., *Set-up 1* and *Set-up 2*) are summarized in Table 2 while Figure 1 shows the colorized plot of an oil sample obtained from olives at early ripening stage (*Baza-1-A*) analyzed by the two set-ups (Figure 1A for *Set-up 1* and Figure 1B for *Set-up 2*). Pattern differences are visible and are related to the different chromatographic efficiencies (peak-width – ^1W and ^2W – Table 2 data) that impact resolution and to the absolute retentions that affect system orthogonality[35–37].

207 The global misalignment between peaks patterns is visualized in Figure 2 and evaluated by calculating
208 analytes relative retention in both separation dimensions against reference peaks[19]. *(Z)-3-hexen-1-ol* acetate,
209 which elutes in the middle area of the chromatographic plane, was arbitrarily chosen as refence/centroid,
210 while *phenol,* the last-eluting marker analyte, was used to normalize analytes relative position *i*.
211 The ^{1}D relative retention (^{1}D RR) is calculated by Equation 1:
212 $$^{1}\text{D RR} = (^{1}t_{Ri} - {}^{1}t_{R\text{(Z)-3-Hexen-1-ol acetate}}) / {}^{1}t_{R\text{phenol}} \quad \text{Eq. 1}$$
213 where $^{1}t_{Ri}$ corresponds to the first dimension retention time expressed in minutes for the targeted peak *i*, *(Z)-*
214 *3-Hexen-1-ol acetate* is the reference peak, and *phenol* is the last eluting peak.
215 The ^{2}D relative retention (^{2}D RR) is calculated through Equation 2:
216 $$^{2}\text{D RR} = (^{2}t_{Ri} - {}^{2}t_{R\text{(Z)-3-Hexen-1-ol acetate}}) / P_M \quad \text{Eq. 2}$$
217 where $^{2}t_{Ri}$ corresponds to the second dimension retention time expressed in minutes for the targeted peak *i*,
218 *(Z)-3-hexen-1-ol acetate* is the reference peak, and P_M the modulation time [19].
219 As visualized in Figure 2, there is a dramatic impact on the ^{2}D absolute and relative retention. This
220 effect is primarily due to the different P_M applied (4 vs. 3.5 seconds) and to the actual carrier gas linear
221 velocities ($^{1}\bar{u}$ and $^{2}\bar{u}$) and operative pressures (initial head-pressure and mid-point pressure). Analytes falling
222 in the third quadrant show an higher ^{2}D *k* in *Set-up 1* whereas this effect is minimized as a function of the
223 increasing ^{1}D *k* (retention).
224 Interestingly, the two patterns, although misaligned on the normalized retention times space, keep
225 coherent the group-type separation for homologous series. Normal alkanes (*n*-alkanes), shown with green
226 indicators, mostly in the second quadrant; linear aldehydes, shown with orange indicators, spanning mostly
227 across the first and third quadrants; and short chain fatty acids, shown with cyan indicators, appearing in the
228 forth quadrant, all are rationalized over the 2D space.
229 The next step of the study addresses detector response variations and examines threshold
230 parameters for 2D-peaks descriptors to adopt for consistent template matching.
231

Supervised work-flow for reliable targeted template construction

MS detector response fluctuations due to tuning, optimization, or by other factors directly impact on absolute response and background noise intensity. Such performance issues also may affect template matching effectiveness and analytes identity confirmation, as a consequence of the varying quality and reliability of 2D-peak spectra adopted as reference for matching. In this study, TOF MS was set differently: in *Set-up 1* it multiplexed between high and low ionization energies (Tandem Ionization ™ - 70 and 12 eV) at 50 Hz acquisition rate for each channel whereas in *Set-up 2* TOF MS operated in single electron energy acquisition mode (70 eV) at 100 Hz. Therefore, MS was tuned differently [38] and output signals exhibited different absolute response (total ion current) and background noise intensity.

The signals resulting from the same sample (*Baza*-1-A), whose patterns are illustrated in Figure 1A and 1B, have the following characteristics:

- background noise sampled in the middle of the chromatogram within a 50x50 acquisition-point window reported an average absolute intensity of 75,500 counts (RSD% ^{1}D 0.6 and ^{2}D 5.98) for *Set-up 1* and 167,000 counts (RSD% ^{1}D 0.5 and ^{2}D 1.57) for *Set-up 2*. Supplementary data visually illustrates performance evaluation operations (Supplementary material – Supplementary Figure 2 – SF2);
- after background subtraction[39], the average intensity was 38,000 counts (RSD% ^{1}D 1.19 and ^{2}D 8.9) for *Set-up 1* and 88,000 counts (RSD% ^{1}D 0.9 and ^{2}D 2.49) for *Set-up 2*;
- the number of detected 2D-peaks above a SNR value of 15 were 770 for *Set-up 1* and 500 for *Set-up 2*;
- within the detected 2D-peaks, SNR values ranged between 15-13,000 in *Set-up 1* and between 15-3,000 in *Set-up 2*;
- volume-to-noise ratio (VNR) values ranged between 100-14,100 in *Set-up 1* and 100-6,700 in *Set-up 2*.

Experimental results indicate that MS, operating with a single ionization energy at 70 eV, results in higher absolute and relative background noise (e.g., 1.9 times) compared to the tandem ionization settings. Interestingly, the noise fluctuations are greater along the ^{1}D (RSD% values) where column bleeding increases as a function of temperature programming. Background noise subtraction has almost the same effect, in terms of noise suppression, and in both cases, signal intensity is halved compared to the initial values.

With respect to peak detection, *Set-up 1* exhibited better chromatographic efficiency (Table 3 ${}^{1}W_{0.1}$ and ${}^{2}W_{0.1}$) and resulted in a higher number of detected peaks over SNR≥15, with a wide range of variation, i.e., 15-13,000, whereas in *Set-up 2* maximum SNR achieved only a value of 3,000. On the other hand, VNR, which corresponds to the ratio of analyte 2D-peak volume to the standard error – SE ($\sigma/\sqrt{n}$), is not so influenced by peak-width as SNR. The dynamic range of the MS response with *Set-up 1* is 10 times greater (up to 14,100 vs. up to 6,700). Although with *Set-up 1* the number of detected peaks over a SNR of 15 was higher, a greater volume standard deviation (σ) was computed.

Based on the differences observed in the absolute response, spectral quality fluctuations were expected especially for low-intensity or threshold peaks. Therefore, the next step was to define threshold parameters for template construction with the objective of achieving 100% true-positive matches (accuracy) in presence of random variability over the application of the method in a short term and with the same method set-up. Therefore, these tests were performed between analytical replicates of the same oil sample to define benchmark values and then validated over a new sample acquired in the same conditions. Tests were done on the three analytical replicates of *Baza-4-A* acquired by *Set-up 1*; validation was on the three replicates of *Bena-4-A* acquired by the same set-up.

Threshold values for candidate peaks populating a template were set for SNR and NIST Similarity match factor (direct match factor - DMF); as reference, spectral signatures were tested for the average spectrum (named *blob spectrum*) and the highest modulation spectrum (named *peak spectrum*). SNR values were varied step-wise in a range between 10 and 100, covering 2D peaks with a percent response between 0.01 and 0.04, while the DMF threshold was set at 800 or 700[40]. Templates were built with ten 2D-peaks with SNR within the fixed range and homogeneously distributed over the 2D space. Experimental results, expressed as % of true-positive matches, are reported in Table 4.

2D Peaks with SNR values below 50 are connoted by inconsistent MS spectral signatures resulting in false-negative matches even when DMF threshold is lowered from 800 to 700. For these peak, neither the *blob spectrum* nor the *peak spectrum* are sufficiently reliable to carry consistent information for identity confirmation. For 2D-peaks with a SNR within 50-100 in the reference chromatogram, the rate of positive

matches increases from 10% to 70% when MS constraints are kept at 800 DMF and *blob spectrum* considered. The rate of true-positive matches reaches 100% with the combination of lower DMF threshold at 700 and *peak spectrum* taken as reference MS. Note, in these cases no false positives were revealed, meaning that correspondences were just established between peaks generated by the same chemical entity.

Results suggest that a SNR cut-off should be defined, based on 2D data particulars, to limit inconsistencies at targeted identity confirmation level. The validation of these settings was by applying the templates resulting from the reference sample *Baza-4-A* to *Bena-4-A* replicates on the same column set-up. Results, reported in Table 4, confirmed the need of applying a SNR threshold of at least 50 with better performances at DMF threshold of 700 with a reference template spectrum collected from the highest modulation (*peak spectrum*.) In this case, some template peaks were unmatched (true-negative matches) because they were not detected in the analyzed sample (below method Limit of Detection LOD).

Rules for template peaks thresholds and reference spectra were then applied to build a reference *targeted template* of known analytes. This fully supervised approach aimed at characterizing the distribution of marker compounds known for their role in defining olive oil sensory quality or to refer about olives ripening status[7,8]. It was also the complement of the *UT* fingerprinting process and included targeted peaks listed in Table 3. Analytes identifications were by combining ^{1}D retention data (I^{T}_{S}) with MS fragmentation pattern similarity above 900 DMF adopting commercial [40,41] and in-house databases or, when possible, by authentic standards.

The template of 126 2D-peaks was built by inspecting samples patterns obtained with *Set-up 1*; reference peaks inclusion was limited to those analytes showing a SNR≥100, the reference spectra was from the highest modulation (e.g., *peak spectrum*), and the MS constraint was set at 700 DMF and 700 Reverse Match Factor (RMF)[40].

Results for targeted template matching are summarized in Table 5. The average matching within *Set-up 1* samples achieved 97% (122 over 126 peaks). Further comments will follow for template matching for *Set-up 2* patterns.

Template transformation

Once template construction was established with simple rules for confident identification and effective matching, the next step was the selection of matching algorithm (transform) and related parameters to effectively transform the template over the peak pattern showing severe misalignment. To approach this challenge, global polynomial transformation were tested, as it was successful in complex re-alignment problems such as those posed by method translation from thermal to differential flow modulated platforms [16,19]. Global, low-degree transformation functions (second-degree or third-degree polynomials) are successful when a sufficient number of alignment points, at least 10 for affine and 30 for second-degree polynomial, are available to guide the pattern re-alignment[16].

The strategy here applied included the re-alignment of the targeted template built over *Set-up 1* analyses on those from *Set-up 2*. The first step included the adjustment of the *distance threshold* parameters for the ^{1}D and ^{2}D, which correspond to the horizontal and vertical distance threshold that limits the after-transformation distance between template and candidate 2D-peaks. These distances are expressed in inter-sample distances (i.e., pixel dimensions). To compensate for the greater ^{2}D misalignment the 2D distance threshold was step-wise incremented from a minimum value of 5 up to 25. On the ^{1}D, where the misalignment was minimal, a threshold of 10 was sufficient to avoid false-negative matches.

At the same time, affine and polynomial second-degree matching transforms were tested for performance. The *Set-up 1* targeted template was applied to the first (arbitrarily chosen) sample pattern (i.e., *Baza-4-A*) obtained with *Set-up 2*. At first, the number of true-positive matches was higher for the second-degree transform, so the iterative process of *match-and-transform* was continued. Iterating the process of matching and template transform allows the template to be adapted to the actual pattern while increasing, step-by-step, the number of matched peaks up to the maximum number that corresponds to all targeted analytes actually detected/confirmed in the analyzed pattern. In practice, this operation increments the number of alignment points available step-by-step, thus improving the quality of the global template transformation at each step.

Experimental results for the application of the targeted template adapted to *Set-up 1* over the patterns of *Set-up 2* resulted in a 65 positive matches over 126 template peaks (51% - 5.4% RSD); after transformation by taking these 65 alignment points, the successive matching step achieved 95 positive matches with a 75% of matched peaks; then, the next step matched 110 peaks, 87%. The maximum number of matched peaks, shown in Table 5, was 121 (96%) and was achieved after one additional matching step. In practice, for a full and effective re-alignment of the targeted template, a variable number of iterations between 3 and 5 was applied.

All such results are listed in Table 5; benchmark values for maximum template matching performance are those corresponding to the application of the targeted template to *Set-up 1* samples (first column). On average matching performance was better for *Set-up 1* patterns (97% of true-positive matches), although the loss of accuracy on *Set-up 2* pattern was just of 5% (92.46% *vs.* 97.02%). These results present a solid foundation for the application of this experimented strategy to a fully-unsupervised approach as that for the reliable template construction.

The next section illustrates the process of *feature template* construction over samples patterns from *Set-up 1* and its successive alignment over *Set-up 2* patterns. Accuracy results are discussed as % of true-positive matches [42].

UT fingerprinting: feature template construction

The feature template was built over a subset of *Set-up 1* chromatograms with the first analytical replicate of all analyzed samples. The 2D chromatograms were pre-processed for rasterization, background subtraction, and 2D peaks detection above a SNR threshold of 10. Data processing was then conducted within a component of the GC Image software suite (Image Investigator™) using the previously validated settings:

- SNR≥100 as threshold value for template peaks;
- *peak spectrum* as MS reference to upload in the template;
- DMF similarity threshold at 700;
- matching transform parameters with ^{1}D distance threshold of 10 and ^{2}D of 25.

Additional settings, specific for this process, included an option for reliable peak inclusion that was set as *most relaxed*: with this setting the algorithm considers as *reliable peaks* all 2D-peaks that match across at least half of the chromatograms. Reliable peaks are fundamental as they are used as alignment points for the transform function when the *feature template* is used to cross-align a large set of chromatograms including those obtained with a different set-up.

The final *feature template* accounted for 257 reliable peaks and 1500 peak-regions. The reliable template built over *Set-up 1* chromatograms is shown in Supplementary material – Supplementary Figure 3 – SF3. Matching constraints for MS spectrum similarity were applied on 2D-peak features and results are summarized in Table 5. The average % of matching for *Set-up 1* chromatograms was 99.75 (±0.45 RSD%); when the feature template is transformed to match for *Set-up 2* patterns, the % of matching is slightly lower 97.95 (±0.99 RSD%) but evidences the high accuracy of the process.

Once re-aligned almost all chemical features detected in all samples patterns, proceeding in a sort of data fusion, the final step aimed at defining the best 2D-peak response descriptor for cross-sample analysis.

Response normalization and samples clustering

As evidenced by signal intensity evaluation and by 2D-peaks statistics (SNR and VNR distribution), pattern cross-alignment is not sufficient to compensate for random variations across measurement sessions and impacting detector response. Response normalization is mandatory to allow consistent cross-comparison of data set. The removal of unwanted intensity variation (i.e., normalization) is referred to as signal drift correction, removal of batch effect, scaling, and matrix effects removal, and can be approached differently as function of the study objectives[21]. Normalization is, in fact, a fundamental step because it may affect the outcomes of a study; the meaningfulness of differentially abundant analytes may vary depending on the normalization method[21,39].

In this study, we tested three simple approaches within those generally adopted in volatiles profiling studies[2,43]. The first included multiple IS normalization with α and β tujone and methyl octinoate that were pre-loaded into the SPME device before headspace sampling of olive oils[23]. As an alternative method, the

analyte % response (calculated on the 2D-Volume) was considered. It was obtained by: (a) normalizing analyte 2D-Volume over all detected 2D-peaks above fixed thresholds or (b) normalizing analyte 2D-Volume over all *UT* peaks included in the feature template. In this last approach 2D-peaks from column bleeding and from interferents were excluded.

Results are illustrated, for a selection of informative analytes covering different volatilities, polarities and amounts, in Figure 3. The bias is computed as Error %, according to Equation 3, and between 2D peaks response (Normalized 2D Volume over IS, % response over all 2D peaks detected or over UT peaks) taking *Set-up 2* as reference.

$$\text{Error \%} = ((Response_i\ Set\text{-}up\ 2 - Response_i\ Set\text{-}up\ 1)/\ Response_i\ Set\text{-}up\ 2)*100 \quad \text{Eq. 3}$$

Normalization by IS(s) performs, on average, more effectively than those on % response (Average Error % 11.2 vs. 16.2); it better compensates for response fluctuations derived by S/SL injection discriminations here impacting on highly volatiles (heptane and acetone) and due to the different operative head-pressures applied to the two set-up. For highly volatile analytes, the Error % computed for IS normalization drops below 16 while for % response it reaches 50-60%. On the other hand, % response indicators well compensate for detection variability on less abundant analytes as 1-octen-3 ol (0.01%), 2-butanone (0.01%), 3-penten-2-one (0.02%), 1-octanol (0.04%), and octanal (0.06%). Note, response variations also are influenced by chromatographic performance; 2D-peaks showing long tails (carbonyl derivatives or unsaturated alcohols) or distorted by overloading phenomena may be splitted in multiple 2D-peaks. In these cases, supervision is needed to merge all 2D-peaks belonging to the same analytes in a single one.

Although ISs normalization gave better results for analytes heavily discriminated, it requires a dedicated sample preparation with ISs pre-loading before sampling that may impact on the global analysis time. In addition, a careful selection of standards is necessary by focusing on compounds not present in the samples under study while covering the suitable range of volatility and polarity. In this scenario, % response normalization is attractive being simpler and less time consuming although it does not rule out the use of

quality control procedures as for example external standardization (ES) or multiple quality control samples analysis.

As a proof of concept, a PCA was run on normalized responses obtained from the two set-up. The data matrix consisted of 126 chemical variables corresponding to the targeted analytes listed in Table 2, and 16 samples corresponding to 2 plots × 4 ripening stages × 2 set-up. Technical replicates were averaged; therefore, the final data matrix was 126 × 16 dimensioned. Results are shown in Figure 4A; observations (samples) are rationally distributed over the Cartesian plane according to the ripening stage of olives (visible along the F1 from right to left) and in accordance to oil quality (i.e. extra-virgin, virgin or *lampante*). Results are in agreement with those from the original study by Magagna et al.[7] Measurements from the two set up are homogeneously distributed confirming the effectiveness of response normalization on the "batch effect". The latter is clearly visible in Figure 4B where the PCA is conducted on analytes 2D volumes without normalization. Here the two independent clusters belonging to set up 1 and set up 2 are clearly visible and well discriminated along the F1.

Acknowledgements

The research was supported by Progetto Ager – Fondazioni in rete per la ricerca agroalimentare. Project acronym *Violin* - Valorization of Italian olive products through innovative analytical tools.

Compliance with ethical standards Notes

Prof. Stephen E. Reichenbach and Dr. Qingping Tao have a financial interest in GC Image, LLC.

References

(1) Tranchida, P. Q.; Donato, P.; Cacciola, F.; Beccaria, M.; Dugo, P.; Mondello, L. Potential of Comprehensive Chromatography in Food Analysis. *TrAC - Trends Anal. Chem.* **2013**, *52*, 186–205.

(2) Cordero, C.; Kiefl, J.; Reichenbach;, S. E.; Bicchi, C. Characterization of Odorant Patterns by Comprehensive Two-Dimensional Gas Chromatography: A Challenge in Omic Studies. *Trends Anal. Chem.* **2018**.

(3) Cordero, C.; Kiefl, J.; Schieberle, P.; Reichenbach, S. E.; Bicchi, C. Comprehensive Two-Dimensional Gas Chromatography and Food Sensory Properties: Potential and Challenges. *Analytical and Bioanalytical Chemistry*. Springer Verlag 2015, pp 169–191.

(4) Reichenbach, S. E.; Tian, X.; Tao, Q.; Ledford, E. B.; Wu, Z.; Fiehn, O. Informatics for Cross-Sample Analysis with Comprehensive Two-Dimensional Gas Chromatography and High-Resolution Mass Spectrometry (GCxGC-HRMS). *Talanta* **2011**, *83* (4), 1279–1288.

(5) Cordero, C.; Liberto, E.; Bicchi, C.; Rubiolo, P.; Schieberle, P.; Reichenbach, S. E.; Tao, Q. Profiling Food Volatiles by Comprehensive Two-Dimensional Ga Schromatography Coupled with Mass Spectrometry: Advanced Fingerprinting Approaches for Comparative Analysis of the Volatile Fraction of Roasted Hazelnuts (Corylus Avellana L.) from Different Ori. *J. Chromatogr. A* **2010**, *1217* (37), 5848–5858.

(6) Kiefl, J.; Schieberle, P. Evaluation of Process Parameters Governing the Aroma Generation in Three Hazelnut Cultivars (Corylus Avellana L.) by Correlating Quantitative Key Odorant Profiling with Sensory Evaluation. *J. Agric. Food Chem.* **2013**, *61* (22), 5236–5244.

(7) Magagna, F.; Valverde-Som, L.; Ruíz-Samblás, C.; Cuadros-Rodríguez, L.; Reichenbach, S. E.; Bicchi, C.; Cordero, C. Combined Untargeted and Targeted Fingerprinting with Comprehensive Two-Dimensional Chromatography for Volatiles and Ripening Indicators in Olive Oil. *Anal. Chim. Acta* **2016**, *936*, 245–258.

(8) Purcaro, G.; Cordero, C.; Liberto, E.; Bicchi, C.; Conte, L. S. Toward a Definition of Blueprint of Virgin Olive Oil by Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2014**, *1334*, 101–111.

(9) Etxebarria, N.; Zuloaga, O.; Olivares, M.; Bartolomé, L. J.; Navarro, P. Retention-Time Locked Methods in Gas Chromatography. *J. Chromatogr. A* **2009**, *1216* (10), 1624–1629.

(10) Klee, M. S.; Quimby, B. D.; Blumberg, L. M. No Title. US Patent 5,987,959 A, 1999.

(11) Blumberg, L. M. No Title. US Patent 6,634,211 B1, 2003.

(12) Jiang, W.; Zhang, Z. M.; Yun, Y.; Zhan, D. J.; Zheng, Y. B.; Liang, Y. Z.; Yang, Z. Y.; Yu, L. Comparisons of Five Algorithms for Chromatogram Alignment. *Chromatographia* **2013**, *76* (17–18), 1067–1078.

(13) Niu, W.; Knight, E.; Xia, Q.; McGarvey, B. D. Comparative Evaluation of Eight Software Programs for Alignment of Gas Chromatography-Mass Spectrometry Chromatograms in Metabolomics Experiments. *J. Chromatogr. A* **2014**, *1374*, 199–206.

(14) Stanimirova, I.; Daszykowski, M. Exploratory Analysis of Metabolomic Data. *Compr. Anal. Chem.* **2018**, *82*, 227–264.

(15) Reichenbach, S. E.; Carr, P. W.; Stoll, D. R.; Tao, Q. Smart Templates for Peak Pattern Matching with Comprehensive Two-Dimensional Liquid Chromatography. *J. Chromatogr. A* **2009**, *1216* (16), 3458–3466.

(16) Rempe, D. W.; Reichenbach, S. E.; Tao, Q.; Cordero, C.; Rathbun, W. E.; Zini, C. A. Effectiveness of Global, Low-Degree Polynomial Transformations for GCxGC Data Alignment. *Anal. Chem.* **2016**, *88* (20), 10028–10035.

(17) Reichenbach, S. E.; Rempe, D. W.; Tao, Q.; Bressanello, D.; Liberto, E.; Bicchi, C.; Balducci, S.; Cordero, C. Alignment for Comprehensive Two-Dimensional Gas Chromatography with Dual Secondary Columns and Detectors. *Anal. Chem.* **2015**, *87* (19), 10056–10063.

(18) Magagna, F.; Liberto, E.; Reichenbach, S. E.; Tao, Q.; Carretta, A.; Cobelli, L.; Giardina, M.; Bicchi, C.; Cordero, C. Advanced Fingerprinting of High-Quality Cocoa: Challenges in Transferring Methods from Thermal to Differential-Flow Modulated Comprehensive Two Dimensional Gas Chromatography. *J. Chromatogr. A* **2018**, *1535*, 122–136.

(19) Cordero, C.; Rubiolo, P.; Reichenbach, S. E.; Carretta, A.; Cobelli, L.; Giardina, M.; Bicchi, C. Method Translation and Full Metadata Transfer from Thermal to Differential Flow Modulated Comprehensive Two Dimensional Gas Chromatography: Profiling of Suspected Fragrance Allergens. *J. Chromatogr. A* **2017**, *1480*, 70–82.

(20) Mommers, J.; Ritzen, E.; Dutriez, T.; van der Wal, S. A Procedure for Comprehensive Two-Dimensional Gas Chromatography Retention Time Locked Dual Detection. *J. Chromatogr. A* **2016**, *1461*, 153–160.

(21) De Livera, A. M.; Sysi-Aho, M.; Jacob, L.; Gagnon-Bartsch, J. A.; Castillo, S.; Simpson, J. A.; Speed, T. P. Statistical Methods for Handling Unwanted Variation in Metabolomics Data. *Anal. Chem.* **2015**, *87* (7), 3606–3615.

(22) Setkova, L.; Risticevic, S.; Linton, C. M.; Ouyang, G.; Bragg, L. M.; Pawliszyn, J. Solid-Phase Microextraction-Gas Chromatography-Time-of-Flight Mass Spectrometry Utilized for the Evaluation of the New-Generation Super Elastic Fiber Assemblies. *Anal. Chim. Acta* **2007**, *581* (2), 221–231.

(23) Wang, Y.; O'Reilly, J.; Chen, Y.; Pawliszyn, J. Equilibrium In-Fibre Standardisation Technique for Solid-Phase Microextraction. *J. Chromatogr. A* **2005**, *1072* (1), 13–17.

(24) Reichenbach, S. E.; Carr, P. W.; Stoll, D. R.; Tao, Q. Smart Templates for Peak Pattern Matching with Comprehensive Two-Dimensional Liquid Chromatography. *J. Chromatogr. A* **2009**, *1216* (16), 3458–3466.

(25) Bressanello, D.; Liberto, E.; Collino, M.; Chiazza, F.; Mastrocola, R.; Reichenbach, S. E.; Bicchi, C.; Cordero, C. Combined Untargeted and Targeted Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography: Revealing Fructose-Induced Changes in Mice Urinary Metabolic Signatures.

505 *Anal. Bioanal. Chem.* **2018**, *410* (11), 2723–2737.
506 (26) GC Image™. *GC Image GCxGC Edition Users' Guide*; 2017.
507 (27) Cordero, C.; Liberto, E.; Bicchi, C.; Rubiolo, P.; Reichenbach, S. E.; Tian, X.; Tao, Q. Targeted and Non-
508 Targeted Approaches for Complex Natural Sample Profiling by GCxGC-QMS. *J. Chromatogr. Sci.* **2010**,
509 *48* (4), 251–261.
510 (28) Reichenbach, S. E.; Tian, X.; Cordero, C.; Tao, Q. Features for Non-Targeted Cross-Sample Analysis with
511 Comprehensive Two-Dimensional Chromatography. *J. Chromatogr. A* **2012**, *1226*, 140–148.
512 (29) Magagna, F.; Guglielmetti, A.; Liberto, E.; Reichenbach, S. E.; Allegrucci, E.; Gobino, G.; Bicchi, C.;
513 Cordero, C. Comprehensive Chemical Fingerprinting of High-Quality Cocoa at Early Stages of
514 Processing: Effectiveness of Combined Untargeted and Targeted Approaches for Classification and
515 Discrimination. *J. Agric. Food Chem.* **2017**, *65* (30), 6329–6341.
516 (30) Reichenbach, S. E.; Tian, X.; Boateng, A. A.; Mullen, C. A.; Cordero, C.; Tao, Q. Reliable Peak Selection
517 for Multisample Analysis with Comprehensive Two-Dimensional Chromatography. *Anal. Chem.* **2013**,
518 *85* (10), 4974–4981.
519 (31) Vaz-Freire, L. T.; da Silva, M. D. R. G.; Freitas, A. M. C. Comprehensive Two-Dimensional Gas
520 Chromatography for Fingerprint Pattern Recognition in Olive Oils Produced by Two Different
521 Techniques in Portuguese Olive Varieties Galega Vulgar, Cobrançosa e Carrasquenha. *Anal. Chim. Acta*
522 **2009**, *633* (2), 263–270.
523 (32) Cajka, T.; Riddellova, K.; Klimankova, E.; Cerna, M.; Pudil, F.; Hajslova, J. Traceability of Olive Oil Based
524 on Volatiles Pattern and Multivariate Analysis. *Food Chem.* **2010**, *121* (1), 282–289.
525 (33) Lukić, I.; Carlin, S.; Horvat, I.; Vrhovsek, U. Combined Targeted and Untargeted Profiling of Volatile
526 Aroma Compounds with Comprehensive Two-Dimensional Gas Chromatography for Differentiation of
527 Virgin Olive Oils According to Variety and Geographical Origin. *Food Chem.* **2019**, *270*, 403–414.
528 (34) Progetto Ager. *Violin Project*; 2016.
529 (35) Cordero, C.; Rubiolo, P.; Cobelli, L.; Stani, G.; Miliazza, A.; Giardina, M.; Firor, R.; Bicchi, C. Potential of
530 the Reversed-Inject Differential Flow Modulator for Comprehensive Two-Dimensional Gas
531 Chromatography in the Quantitative Profiling and Fingerprinting of Essential Oils of Different
532 Complexity. *J. Chromatogr. A* **2015**, *1417*, 79–95.
533 (36) Cordero, C.; Rubiolo, P.; Sgorbini, B.; Galli, M.; Bicchi, C. Comprehensive Two-Dimensional Gas
534 Chromatography in the Analysis of Volatile Samples of Natural Origin: A Multidisciplinary Approach to
535 Evaluate the Influence of Second Dimension Column Coated with Mixed Stationary Phases on System
536 Orthogonality. *J. Chromatogr. A* **2006**, *1132* (1–2), 268–279.
537 (37) Venkatramani, C. J.; Xu, J.; Phillips, J. B. Separation Orthogonality in Temperature-Programmed
538 Comprehensive Two-Dimensional Gas Chromatography. *Anal. Chem.* **1996**, *68* (9), 1486–1492.
539 (38) Markes International. Select-EV: The next Generation of Ion Source Technology. *Technical Note*. 2016.

(39) Reichenbach, S. E.; Ni, M.; Zhang, D.; Ledford, E. B. Image Background Removal in Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2003**, *985* (1–2), 47–56.

(40) NIST/EPA/NIH Mass Spectral Library with Search Program Data Version: NIST V17.

(41) Adams, R. P. *Identification of Essential Oil Components by Gas Chromatography—Mass Spectroscopy*; Allured Publishing: New York, 1995.

(42) Kiefl, J.; Cordero, C.; Nicolotti, L.; Schieberle, P.; Reichenbach, S. E.; Bicchi, C. Performance Evaluation of Non-Targeted Peak-Based Cross-Sample Analysis for Comprehensive Two-Dimensional Gas Chromatography-Mass Spectrometry Data and Application to Processed Hazelnut Profiling. *J. Chromatogr. A* **2012**, *1243*, 81–90.

(43) Sgorbini, B.; Cagliero, C.; Liberto, E.; Rubiolo, P.; Bicchi, C.; Cordero, C. E. I. Strategies for Accurate Quantitation of Volatiles from Foods and Plant-Origin Materials: A Challenging Task. *J. Agric. Food Chem.* **2019**, No. 67, 1619–1630.

Caption to Figures

Figure 1: colorized plot of the *Baza-1-A* sample analyzed with *Set-up 1* (**1A**) and *Set-up 2* (**1B**). Red colored areas indicate the available separation space while green areas include all targeted peaks elution area. For details, see the text.

Figure 2: dispersion graph resulting from the relative position of targeted peak analytes from the two set-ups in the normalized retention times space: homologous series *n*-alkanes - green indicators, linear saturated aldehydes - orange indicators, an short chain fatty acids - cyan indicators.

Figure 4: Error %, according to Equation 3, calculated between 2D peaks response (Normalized 2D Volume over IS, % response over all 2D peaks detected or over UT peaks) taking *Set-up 2* as reference for a selection of informative analytes. Red dotted lines indicate boundaries for acceptance at ± 20%.

Caption to Tables

Table 1: List of samples together with acronym, harvest region, harvest stage, quality parameters according to COMMISSION REGULATION (EEC) No 2568/91 of 11 July 1991, sensory evaluation results (Md: median of defects – Mf: median of fruity notes) and commercial classification.

Table 2: *Set-up 1* and *2* columns characteristics, settings and operative pressures.

Table 3: List of all targeted analytes together with their elution order (#Rank) in the two set-ups, retention times (^1t_R and ^2t_R), relative standard deviation (RSD%) calculated over all analyzed samples, and peak-widths at 10% of peak height estimated on each dimension ($^1W_{0.1}$ (min) and $^2W_{0.1}$ (sec)).

Table 4: template matching results based on *Set-up 1* templates including 2D peaks with SNR ranging from 10 to 100. Similarity DMF threshold applied is 800 or 700 and reference spectrum is *blob* (average 2D-peak spectrum) or *peak* (highest modulation spectrum). The upper part of the table refers to benckmark values obtained by applying *Set-up 1* templates over replicated analyses of the same sample (i.e., *Baza-4-A*); the lower part of the table refers about results of *Set-up 1* templates over replicated analyses of another sample (i.e., *Bena-4-A*).

Table 5: template matching results for targeted template (supervised work-flow) from *Set-up 1* applied to *Set-up 2* samples and for feature template (unsupervised and automatic work-flow) built over *Set-up 1* samples and applied on *Set-up 2* samples.

588 **Table 1**: List of samples together with acronym, harvest region, harvest stage, quality parameters according
589 to COMMISSION REGULATION (EEC) No 2568/91 of 11 July 1991, sensory evaluation results (Md: median of
590 defects – Mf: median of fruity notes) and commercial classification.
591

Sample Acronym	Region	Harvest stage	Acidity (%)	Peroxide index (mEq O_2/kg)	K_{232}	K_{270}	ΔK	Md	Mf	Classification
Baza-1-A	Baza	November 10-12	0.2	5	1.84	0.2	0	0	5	EVOO
Baza-2-A	Baza	November 24-28	0.2	3	1.6	0.2	0	0	4.1	EVOO
Baza-3-A	Baza	December 16-17	0.2	5	1.17	0.2	0	> 0.00	1.3	VOO
Baza-4-A	Baza	January 12-15	0.4	11	1.11	0.1	0	> 0.00	0	LOO
Baza-1-B	Baza	November 10-12	0.2	4	1.92	0.2	0	0	5.2	EVOO
Baza-2-B	Baza	November 24-28	0.1	3	1.65	0.2	0	0	3.8	EVOO
Baza-3-B	Baza	December 16-17	0.2	6	1.28	0.1	0	> 0.00	1.7	VOO
Baza-4-B	Baza	January 12-15	0.4	13	1.12	0.1	0	> 0.00	0	LOO
Bena-1-A	Benamaurel	November 10-12	0.2	5	1.61	0.2	0	0	4.4	EVOO
Bena-2-A	Benamaurel	November 24-28	0.2	4	1.53	0.2	0	0	4.3	EVOO
Bena-3-A	Benamaurel	December 16-17	0.2	8	1.19	0.1	0	0	3.1	EVOO
Bena-4-A	Benamaurel	January 12-15	0.4	19	1.05	0.1	0	> 0.00	0	LOO
Bena-1-B	Benamaurel	November 10-12	0.1	4	1.64	0.4	0	0	4.2	EVOO
Bena-2-B	Benamaurel	November 24-28	0.2	3	1.48	0.2	0	0	4.3	EVOO
Bena-3-B	Benamaurel	December 16-17	0.2	6	1.51	0.1	0	0	2.9	EVOO
Bena-4-B	Benamaurel	January 12-15	0.2	14	1.05	0.1	0	> 0.00	0	LOO

592

593 **Table 2:** *Set-up 1* and *2* columns characteristics, settings and operative pressures.

594

	Set-up 1	***Set-up 2***
^{1}D Columns	^{1}D: SolGelWax™ (30 m, 0.25mm d_c, 0.25 µm d_f) Batch N° 1238274C06	^{1}D: SolGelWax™ (30 m, 0.25mm d_c, 0.25 µm d_f) Batch N° 1315621E03
^{1}D Carrier gas settings	He carrier @ 1.3 mL/min - constant flow conditions Average velocity ($^1\bar{u}$): 15.3 cm/s Initial head-pressure (relative) 234 kPa Outlet pressure (absolute) 285 kPa Hold-up 3.27 min	He carrier @ 1.3 mL/min - constant flow conditions Average velocity ($^1\bar{u}$): 12.8 cm/s Initial head-pressure (relative) 290 kPa Outlet pressure (absolute) 349 kPa Hold-up 3.89 min
^{2}D Columns	^{2}D: OV1701 Mega (1.0 m, 0.10 mm d_c, 0.10 µm d_f) Loop-capillary: deactivated fused silica (1.0 m, 0.10 mm d_c,) Restriction toward MS: none	^{2}D: OV1701 Mega (1.0 m, 0.10 mm d_c, 0.10 µm d_f) Loop-capillary: deactivated fused silica (1.0 m, 0.10 mm d_c,) Restriction toward MS: deactivated fused silica (1 m, 0.10 mm d_c,)
^{2}D Carrier gas settings	He carrier @ 1.3 mL/min - constant flow conditions Average velocity ($^2\bar{u}$): 157 cm/s Mid-point pressure (relative) 184 kPa Hold-up 1.28 s	He carrier @ 1.3 mL/min - constant flow conditions Average velocity ($^2\bar{u}$): 128 cm/s Mid-point pressure (relative) 248 kPa Hold-up 2.35 s
Modulation	P_M: 3.5 s - Hot-Jet pulse time: 250 ms	P_M: 4s - Hot-Jet pulse time: 250 ms

595

596

Table 3: List of all targeted analytes together with their experimental I^T_S, identification criterion (a) authentic reference compound or (b) I^T_S ±20 and spectral similarity direct match ≥ 850, elution order (#Rank) in the two Set-up; retention times (^1t_R and ^2t_R) and relative standard deviation (RSD%) calculated over all analyzed samples, peak-width at 10% of peak height estimated on each dimension ($^1W_{0.1}$ (min) and $^2W_{0.1}$ (sec)).

Compounds	Exp I^T_S	Identification	*Set-up 1*							*Set-up 2*						
			# Rank	^1t_R (min)	RSD%	^2t_R (sec)	RSD%	$^1W_{0.1}$ (min)	$^2W_{0.1}$ (sec)	*# Rank*	^1t_R (min)	RSD%	^2t_R (sec)	RSD%	$^1W_{0.1}$ (min)	$^2W_{0.1}$ (sec)
Hexane	778	a	1	4.36	0.77	0.33	3.53	0.12	0.05	*1*	4.80	8.45	1.06	5.80	0.40	0.15
2,2,4-Trimethylpentane	789	b	2	4.63	0.73	0.56	0.00	0.10	0.06	*2*	5.44	3.53	0.83	3.07	0.36	0.13
Heptane	795	a	3	4.78	0.00	0.58	0.00	0.10	0.05	*3*	5.53	3.61	0.60	5.38	0.38	0.16
Cyclohexane	802	b	4	4.96	0.00	0.49	2.34	0.08	0.04	6	*5.88*	*3.49*	0.96	*7.99*	*0.29*	*0.11*
1,3-Pentadiene	813	b	5	5.25	0.00	0.42	0.00	0.37	0.17	4	*5.60*	*0.00*	*0.54*	*4.26*	*0.53*	*0.17*
Propanal	819	a	6	5.44	0.62	0.28	0.00	0.25	0.12	5	*5.82*	*0.00*	0.52	*7.62*	*0.40*	*0.15*
Octane	830	a	7	5.70	0.59	1.06	0.00	0.10	0.09	*8*	6.28	1.18	1.68	1.04	0.33	0.15
Acetone	831	a	8	5.74	0.59	0.29	4.03	0.25	0.11	*10*	6.67	0.00	0.74	8.48	0.53	0.37
1-Octene	850	b	9	6.26	0.54	1.01	1.15	0.12	0.08	*7*	6.26	0.62	1.01	3.78	0.40	0.23
Tetrahydrofuran	858	b	10	6.46	0.52	0.48	0.00	0.23	0.16	*9*	6.33	1.82	0.48	13.50	0.36	0.18
Butanal	867	a	11	6.67	0.51	0.42	0.00	0.23	0.13	*11*	7.47	2.36	0.91	8.41	0.42	0.23
Ethyl acetate	873	a	12	6.86	0.49	0.43	2.71	0.18	0.05	*13*	7.74	0.01	1.08	0.05	0.33	0.16
2,3-Dimethylheptane	893	b	13	7.43	0.45	1.76	1.14	0.23	0.17	*12*	7.73	0.86	1.99	11.30	0.38	0.13
2-Methylbutanal	895	a	14	7.44	0.01	0.60	0.21	0.22	0.07	*15*	8.27	8.41	1.16	8.24	0.43	0.24
Ethanol	913	a	15	7.78	0.43	0.26	0.00	0.19	0.09	*14*	7.78	2.47	0.38	3.53	0.42	0.29
1-Methoxyhexane	932	b	16	8.42	0.40	1.15	1.01	0.14	0.09	*17*	9.09	0.43	2.21	2.66	0.29	0.31
2-Ethylfurane	932	b	17	8.42	0.40	0.54	0.00	0.29	0.07	*18*	9.11	0.74	1.10	3.64	0.38	0.17
2-Methylnonane	947	b	18	8.85	0.38	2.29	2.02	0.31	0.24	*16*	9.04	0.43	2.85	2.84	0.49	0.23
2,3-Butanedione	955	a	19	8.94	0.38	0.37	3.09	0.37	0.05	*19*	9.18	3.71	0.45	11.80	0.42	0.13
Pentanal	956	a	20	9.03	0.00	0.65	0.02	0.30	0.23	*21*	10.00	0.04	1.26	0.05	0.37	0.16
3-Methylnonane	960	b	21	9.20	0.37	2.33	0.50	0.31	0.27	*20*	9.65	0.69	2.48	12.00	0.51	0.19
Acetonitrile	980	b	22	9.69	0.00	0.27	0.04	0.15	0.10	*22*	10.27	0.03	0.86	0.29	0.30	0.32
(Z)-1-Methoxy-3-hexene	991	b	23	10.01	0.34	1.03	1.12	0.14	0.09	*24*	10.66	0.36	1.99	4.07	0.24	0.24
Decane	997	a	24	10.18	0.01	2.51	0.03	0.18	0.08	*23*	10.48	0.06	3.31	0.09	0.32	0.15
1-Penten-3-one	1000	a	25	10.25	0.33	0.56	0.00	0.21	0.06	*25*	10.96	0.70	1.11	3.74	0.42	0.23
Propan-1-ol	1011	a	26	10.65	0.00	0.29	0.04	0.19	0.08	*28*	11.43	0.02	0.65	0.11	0.34	0.18
α-Pinene	1011	a	27	10.66	0.00	1.62	0.01	0.20	0.06	*27*	11.18	0.01	2.77	0.17	0.28	0.16
(E)-2-Butenal	1017	a	28	10.95	1.11	0.50	4.00	0.21	0.15	*26*	11.05	1.51	0.59	6.55	0.40	0.15
Toluene	1020	a	29	10.99	0.31	0.68	2.94	0.18	0.07	*29*	11.61	0.67	1.37	4.68	0.33	0.17
1-Decene	1032	b	30	11.41	0.00	2.06	0.01	0.17	0.05	*30*	11.98	0.38	2.98	0.53	0.43	0.20
2,3-Pentanedione	1033	a	31	11.50	0.00	0.52	0.00	0.21	0.06	*32*	12.19	0.32	1.16	0.27	0.36	0.23
4-Methyldecane	1051	b	32	12.04	0.56	2.83	0.41	0.29	0.31	*31*	12.02	0.56	2.93	2.44	0.53	0.38

Hexanal	1059	a	33	12.39	0.27	0.86	0.00	0.21	0.07	*33*	13.09	1.28	1.72	6.23	0.42	0.19
Isobutyl alcohol	1061	a	34	12.46	0.27	0.34	0.00	0.16	0.04	*34*	13.10	0.88	0.78	6.51	0.29	0.17
2,4,6-Trimethyldecane	1090	b	35	13.26	0.25	3.11	0.37	0.29	0.33	*38*	13.99	1.93	3.26	8.52	0.44	0.35
β-Pinene	1093	a	36	13.59	0.00	1.58	0.00	0.17	0.05	*36*	13.77	0.01	1.70	0.22	0.25	0.12
Undecane	1097	a	37	13.75	0.00	2.85	0.01	0.19	0.09	*42*	14.30	0.01	3.61	0.09	0.32	0.16
3-Penten-2-one	1100	a	38	13.88	0.00	0.60	0.00	0.27	0.14	*35*	13.73	1.70	0.69	6.19	0.42	0.22
1-Methoxy-2-propanol	1101	a	39	13.90	0.01	0.39	0.03	0.38	0.09	*39*	14.10	0.02	0.48	0.74	0.41	0.22
3-Methylbutyl acetate	1101	a	40	13.90	0.24	1.02	0.00	0.14	0.10	*37*	13.95	3.01	1.34	9.21	0.69	0.18
1-Methoxy-1-propanol	1103	b	41	13.98	0.24	0.39	2.94	0.39	0.16	*40*	14.12	1.84	0.66	10.90	0.36	0.32
Ethylbenzene	1106	b	42	14.12	0.00	0.86	0.00	0.16	0.07	*41*	14.24	1.51	1.31	7.54	0.42	0.30
1,4-Dimethylbenzene	1115	b	43	14.41	0.00	0.87	1.32	0.16	0.17	*43*	14.59	0.53	0.95	10.60	0.47	0.25
1-Butanol	1115	a	44	14.41	0.00	0.36	0.00	0.21	0.06	*44*	14.98	1.81	1.22	8.66	0.36	0.23
1,3-Dimethylbenzene	1121	b	45	14.64	0.00	0.84	0.00	0.14	0.06	*45*	15.09	0.44	1.75	6.31	0.64	0.35
Butyl 2-methylpropanoate	1128	b	46	14.94	0.00	1.36	0.02	0.12	0.08	*47*	15.62	0.01	2.12	0.05	0.33	0.15
1-Penten-3-ol	1130	a	47	15.01	0.22	0.36	0.00	0.16	0.05	*46*	15.53	0.86	0.99	12.50	0.36	0.24
2-Methylpropyl butyrate	1141	a	48	15.40	0.00	1.33	0.87	0.31	0.21	*48*	15.74	1.30	1.55	8.44	0.38	0.15
ß-Myrcene	1147	a	49	15.65	0.22	1.33	0.87	0.14	0.09	*49*	15.74	1.07	1.55	6.74	0.34	0.32
Heptanal	1162	a	50	16.28	0.00	1.02	0.00	0.19	0.08	*52*	17.07	2.64	2.00	6.10	0.44	0.18
1,3-Xylene	1163	b	51	16.33	0.00	0.83	1.39	0.18	0.11	*50*	16.61	0.97	1.34	4.40	0.44	0.29
2-Ehylhexanal	1166	a	52	16.45	0.00	1.33	0.87	0.14	0.09	*53*	17.17	3.19	2.30	13.60	0.51	0.29
3-Methyl-2-butenal	1172	a	53	16.68	0.00	0.61	1.88	0.16	0.07	*51*	16.98	2.24	0.97	14.10	0.47	0.32
3-Methyl-1-butanol	1177	b	54	16.86	0.00	0.79	1.47	0.16	0.07	*54*	17.50	3.19	1.45	13.60	0.51	0.29
2-Methyl-1-butanol	1177	b	55	16.86	0.00	0.41	0.03	0.16	0.08	*56*	17.69	0.01	1.10	0.25	0.38	0.25
Limonene	1182	a	56	17.09	0.00	1.41	0.82	0.16	0.09	*55*	17.57	0.44	2.87	19.40	0.36	0.26
Eucalyptol	1190	a	57	17.37	0.00	1.54	0.01	0.32	0.15	*58*	17.94	0.01	2.43	0.12	0.37	0.24
(E)-2-Hexenal	1193	a	58	17.50	0.00	0.81	1.43	0.31	0.12	*57*	17.80	2.43	1.50	3.66	0.31	0.19
Dodecane	1197	a	59	17.68	0.00	3.02	0.01	0.19	0.09	*59*	18.18	0.01	3.68	0.08	0.33	0.16
Terpinene	1229	a	60	18.92	0.18	1.35	0.85	0.12	0.09	*60*	19.20	0.34	1.71	3.07	0.20	0.30
(E)-ß-Ocimene	1234	a	61	19.08	0.00	1.27	1.82	0.16	0.11	*62*	19.49	0.35	2.64	5.06	0.27	0.21
1-Pentanol	1236	a	62	19.15	0.18	0.11	6.74	0.33	0.17	*61*	19.33	2.60	0.92	7.57	0.69	0.31
Styrene	1244	b	63	19.57	0.01	0.62	0.01	0.31	0.22	*63*	19.98	0.01	1.39	0.10	0.58	0.43
1-Dodecene	1249	b	64	19.74	0.17	2.67	0.87	0.27	0.32	*64*	19.98	1.75	2.74	12.80	0.33	0.30
Hexyl acetate	1250	a	65	19.78	0.00	1.13	1.02	0.18	0.09	*67*	20.59	0.97	2.35	15.40	0.42	0.22
3-Hydroxy-2-butanone	1256	a	66	19.93	0.17	0.63	1.84	0.37	0.21	*66*	20.58	1.16	1.02	4.55	0.36	0.21
1,2,3-Trimethylbenzene	1261	b	67	20.18	0.00	0.96	0.00	0.27	0.09	*65*	20.35	2.01	1.06	3.73	0.38	0.27
2-Ethyl-2-hexenal	1278	a	68	20.77	0.00	1.14	0.02	0.17	0.09	*68*	21.34	0.02	2.13	0.08	0.48	0.32
(Z)-3-Hexenyl acetate	1300	a	69	21.43	0.16	0.94	0.00	0.14	0.09	*69*	21.95	0.47	1.98	1.01	0.27	0.33
N,N-Dimethylformamide	1302	b	70	21.53	0.00	0.50	0.00	0.21	0.15	*71*	22.28	1.95	1.47	13.30	0.36	0.37
(Z)-2-Heptenal	1306	a	71	21.66	0.16	0.90	0.00	0.25	0.08	*70*	22.24	1.55	1.38	5.59	0.29	0.29
6-Methyl-5-hepten-2-one	1319	a	72	22.17	0.00	0.91	1.26	0.18	0.08	*72*	22.86	0.45	1.87	1.22	0.20	0.29
(Z)-3-Hexen-1-ol	1346	a	73	23.74	0.00	0.42	0.00	0.21	0.11	*73*	23.73	0.85	1.03	6.28	0.22	0.24
(E,E)-2,4-Hexadienal	1369	a	74	24.38	0.00	0.58	0.00	0.18	0.05	*76*	25.39	0.26	1.22	2.84	0.33	0.49
Nonanal	1370	a	75	24.44	0.00	1.24	1.61	0.29	0.10	*75*	24.77	1.48	1.77	14.10	0.49	0.31

(E)-2-Hexen-1-ol	1373	a	76	24.56	0.00	0.40	0.00	0.21	0.09	*74*	24.71	2.20	0.91	3.37	0.31	0.31
(E)-3-Octen-2-one	1382	a	77	24.85	0.00	0.96	0.01	0.28	0.10	*77*	25.41	0.01	1.82	0.03	0.50	0.33
α-Thujone (ISTD)	1402	a	78	25.55	0.00	1.24	0.02	0.25	0.16	*79*	26.47	0.02	2.20	0.13	0.31	0.31
(E)-2-Octenal	1415	a	79	25.73	0.00	0.98	2.04	0.27	0.14	*78*	25.83	1.13	1.01	11.90	0.80	0.24
1-Octen-3-ol	1417	a	80	25.78	0.00	0.34	0.00	0.18	0.08	*81*	26.83	0.01	1.15	0.09	0.20	0.29
β-Thujone (ISTD)	1424	a	81	26.26	0.00	1.21	0.01	0.15	0.09	*80*	26.59	0.01	2.21	0.10	0.21	0.20
1-Heptanol	1429	a	82	26.43	0.00	0.53	2.19	0.18	0.05	*84*	27.19	0.28	1.17	2.65	0.27	0.30
1-Ethenyl-4-ethyl-benzene	1430	b	83	26.48	0.00	0.86	0.00	0.37	0.13	*82*	26.87	0.72	1.81	1.68	0.51	0.25
Furfural	1434	a	84	26.77	0.00	0.34	0.00	0.17	0.06	*86*	27.54	0.01	1.17	0.22	0.32	0.30
(E,E)-2,4-Heptadienal	1437	a	85	26.89	0.00	0.69	1.67	0.14	0.06	*87*	27.80	1.06	1.44	0.00	0.62	0.25
5-Methyl-2-(1-methylethyl)-cyclohexanone	1446	b	86	27.13	0.00	1.31	0.88	0.21	0.13	*83*	27.05	0.28	1.37	2.41	0.33	0.25
Butyl-2-ethylhexanoate	1453	a	87	27.42	0.00	1.89	1.22	0.39	0.29	*85*	27.33	0.88	1.97	7.99	0.40	0.19
2-Ethyl-1-hexanol	1460	a	88	27.72	0.00	0.60	0.00	0.21	0.11	*88*	28.06	0.02	1.46	0.23	0.34	0.20
(Z)-Hepten-4-ol	1469	a	89	28.18	0.00	0.46	0.02	0.19	0.08	*90*	28.93	0.01	1.17	0.23	0.39	0.15
Decanal	1474	a	90	28.35	0.00	1.30	1.54	0.25	0.09	*89*	28.60	1.50	1.89	14.30	0.33	0.25
(E)-2-Hepten-1-ol	1476	a	91	28.41	0.00	0.44	0.02	0.41	0.16	*91*	29.20	0.00	0.97	0.03	0.55	0.33
3,5-Octadien-2-one	1485	a	92	28.88	0.00	0.77	0.01	0.21	0.12	*92*	29.52	0.02	1.56	0.10	0.60	0.31
Benzaldehyde	1494	a	93	29.03	0.12	0.49	2.34	0.18	0.04	*95*	29.95	0.47	1.06	15.40	0.44	0.22
6-Undecanone	1505	a	94	29.38	0.00	1.53	0.01	0.38	0.14	*96*	29.96	0.01	2.30	0.05	0.44	0.25
Propanoic acid	1506	a	95	29.40	0.01	0.14	0.06	0.35	0.09	*97*	30.29	0.02	0.96	0.12	0.47	0.23
(E)-2-Nonenal	1509	a	96	29.58	0.00	1.05	1.10	0.41	0.17	*93*	29.66	1.42	1.16	16.50	0.36	0.15
Linaloool	1514	a	97	29.75	0.00	0.67	1.73	0.18	0.06	*94*	29.90	0.26	0.72	2.90	0.38	0.12
1-Octanol	1525	a	98	30.16	0.00	0.57	2.01	0.14	0.05	*98*	30.69	0.25	1.57	2.40	0.16	0.22
Nonyl acetate	1567	b	99	31.15	0.00	1.33	0.01	0.25	0.16	*99*	30.96	0.00	1.29	0.06	0.40	0.28
5-Methyl-2(5H)-furanone	1589	b	100	31.46	0.11	0.46	0.00	0.54	0.07	*100*	32.49	0.24	1.02	3.08	0.62	0.25
Undecanal	1598	a	101	32.08	0.00	1.37	1.69	0.53	0.28	*102*	32.76	0.12	2.72	0.71	0.78	0.50
Butanoic acid	1610	a	102	32.57	0.10	0.16	0.00	0.39	0.09	*101*	32.64	0.42	1.45	7.74	0.47	0.23
Butyrolactone	1611	a	103	32.61	0.00	0.44	0.00	0.35	0.09	*103*	32.78	0.12	0.94	3.34	0.24	0.27
1-Nonanol	1628	a	104	33.66	0.00	0.63	1.82	0.16	0.05	*104*	34.22	0.11	1.43	2.20	0.20	0.23
Ethyl benzoate	1637	a	105	34.03	0.10	0.67	1.71	0.12	0.06	*105*	34.53	0.84	1.35	10.90	0.40	0.28
5-Ethyl-2(5H)-furanone	1643	b	106	34.28	0.20	0.41	10.20	0.25	0.16	*106*	35.04	0.61	1.24	8.38	0.36	0.33
(Z)-3-Nonen-1-ol	1647	a	107	34.44	0.00	0.57	0.02	0.21	0.07	*107*	35.04	0.01	1.11	0.26	0.44	0.30
Dodecanal	1684	a	108	35.64	0.00	1.42	1.41	0.51	0.21	*108*	36.25	0.38	2.49	8.17	0.47	0.36
α-Muurolene	1701	a	109	35.99	0.00	1.61	1.89	0.16	0.13	*110*	36.86	0.28	2.52	12.80	0.40	0.30
3,4-Dimethyl-2,5-furandione	1701	b	110	35.99	0.00	0.54	0.02	0.25	0.11	*111*	36.96	0.01	1.41	0.12	0.55	0.28
Pentadecanoic acid	1710	a	111	36.16	0.00	0.17	0.21	0.32	0.20	*109*	36.80	0.01	0.94	0.14	0.60	0.34
1,4-Cyclohex-2-enedione	1710	b	112	36.16	0.00	0.43	0.03	0.33	0.19	*112*	37.13	0.00	1.28	0.21	0.49	0.22
α-Farnesene	1725	a	113	36.98	0.00	1.43	0.02	0.15	0.06	*113*	37.13	0.01	2.16	0.11	0.28	0.15
(E,E)-2,4-Decadienal	1732	a	114	37.20	0.09	0.91	2.55	0.37	0.20	*114*	37.64	0.10	0.95	3.20	0.64	0.35
Methyl salicylate	1743	b	115	37.57	0.00	0.57	2.04	0.25	0.11	*115*	38.11	1.57	1.16	9.73	0.56	0.41
2-(2-Butoxyethoxy)-ethanol	1754	a	116	37.92	0.00	0.52	0.00	0.25	0.16	*116*	38.39	0.70	1.31	5.33	0.31	0.35
δ-Pentalactone	1767	b	117	38.33	0.00	0.52	0.00	0.39	0.16	*117*	39.36	0.00	1.18	2.74	0.60	0.32

3-Phenyl-2-propenal	1794	b	118	39.20	0.00	0.52	0.00	0.21	0.07	*118*	40.26	0.10	1.14	3.51	0.40	0.29
Hexadecanoic acid	1805	a	119	39.51	0.09	0.20	0.00	0.31	0.22	*119*	40.47	0.10	0.83	7.78	0.58	0.43
(Z)-6,10-Dimethyl-5,9-undecadien-2-one	1821	b	120	40.02	0.00	1.13	2.05	0.35	0.20	*120*	40.70	0.09	2.12	0.99	0.31	0.29
Butyl benzoate	1830	b	121	40.31	0.00	0.82	0.02	0.41	0.27	*121*	41.27	0.00	1.69	0.14	0.59	0.30
Benzyl alcohol	1832	a	122	40.37	0.00	0.28	0.00	0.28	0.14	*122*	41.29	0.00	0.99	0.26	0.46	0.27
Phenylethyl alcohol	1869	a	123	41.44	0.16	0.34	0.00	0.29	0.11	*123*	41.42	0.16	0.83	12.00	0.56	0.34
4-Phenyl-3-buten-2-one	1914	b	124	42.82	0.00	0.56	0.00	0.40	0.14	*124*	43.77	0.09	1.12	1.61	0.64	0.25
1-Dodecanol	1928	a	125	43.23	0.00	0.81	1.42	0.23	0.13	*125*	43.89	0.32	1.68	8.43	0.38	0.31
Phenol	1956	a	126	44.04	0.00	1.27	3.97	0.43	0.34	*126*	44.84	1.51	1.66	18.30	0.53	0.41
Average			-	-	*0.12*	-	*0.86*	*0.24*	*0.12*	-	-	*0.90*	-	*5.03*	*0.42*	*0.25*
Min			-	-	*0.00*	-	*0.00*	*0.08*	*0.04*	-	-	*0.00*	-	*0.00*	*0.16*	*0.11*
Max			-	-	*1.11*	-	*10.20*	*0.54*	*0.34*	-	-	*8.45*	-	*19.40*	*1.04*	*0.50*

600

601

Table 4: template matching results based on *Set-up 1* templates including 2D peaks with SNR ranging from 10 to 100. Similarity DMF threshold applied are 800 or 700 while reference spectrum is *blob* (average 2D-peak spectrum) or *peak* (highest modulation spectrum). The upper part of the table refers to benckmark values obtained by applying *Set-up 1* templates over replicated analyses of the same sample (i.e., *Baza-4-A*); lower part of the table refers about results of *Set-up 1* templates over replicated analyses of another sample (i.e., *Bena-4-A*).

***Baza-4-A* Set-up 1 – three replicates**

			Similarity threshold 800 – *Blob* MS		**Similarity threshold 700 – *Peak* MS**	
SNR	**% Response**	**Peaks n°**	**% Matching**	**Matched peaks n°**	**% Matching**	**Matched peaks n°**
10 ± 2	0.01	Column bleeding or interferences				
30 ± 2	0.02	10	10.00 %	1	40.00 %	4
50 ± 2	0.02	10	10.00 %	1	90.00 %	9
70 ± 2	0.03	10	40.00 %	4	100.00 %	10
90 ± 2	0.04	10	40.00 %	4	100.00 %	10
100 ± 2	0.04	10	70.00 %	7	100.00 %	10

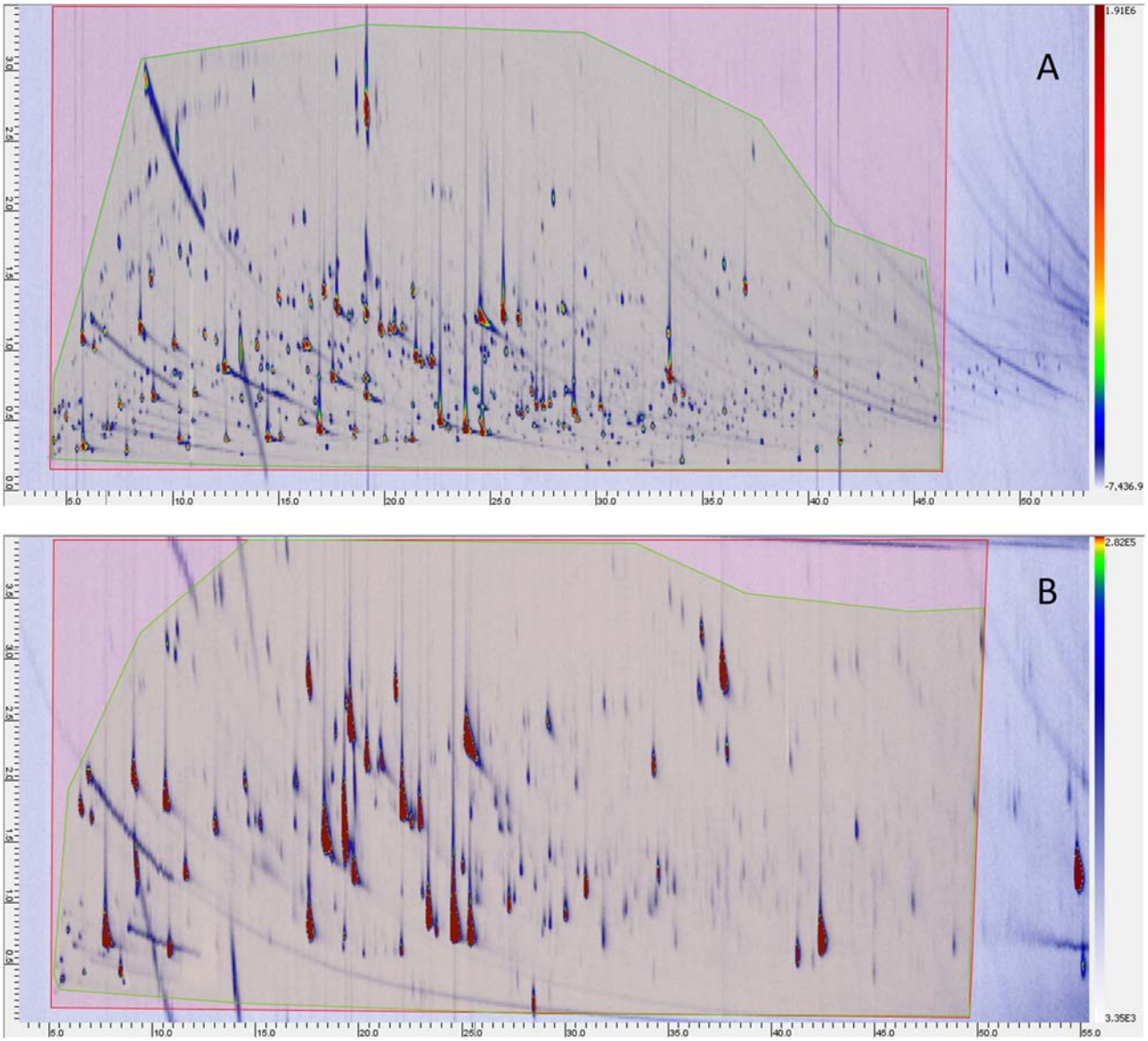
***Baza 4-A* Set-up 1 over *Bena-4-A* Set-up 1**

			Similarity threshold 800 – *Blob* MS		**Similarity threshold 700 – *Peak* MS**	
SNR	**% Response**	**Peaks n°**	**% Matching**	**Matched peaks n°**	**% Matching**	**Matched peaks n°**
10 ± 2	0.01	Column bleeding or interferences				
30 ± 2	0.02	10	10.00 %	1	10.00 %	1
50 ± 2	0.02	10	10.00 %	1	20.00 %	1
70 ± 2	0.03	8	25.00 %	2	87.50 %	7
90 ± 2	0.04	5	40.00 %	2	100.00 %	5
100 ± 2	0.04	7	42.86 %	3	100.00 %	7

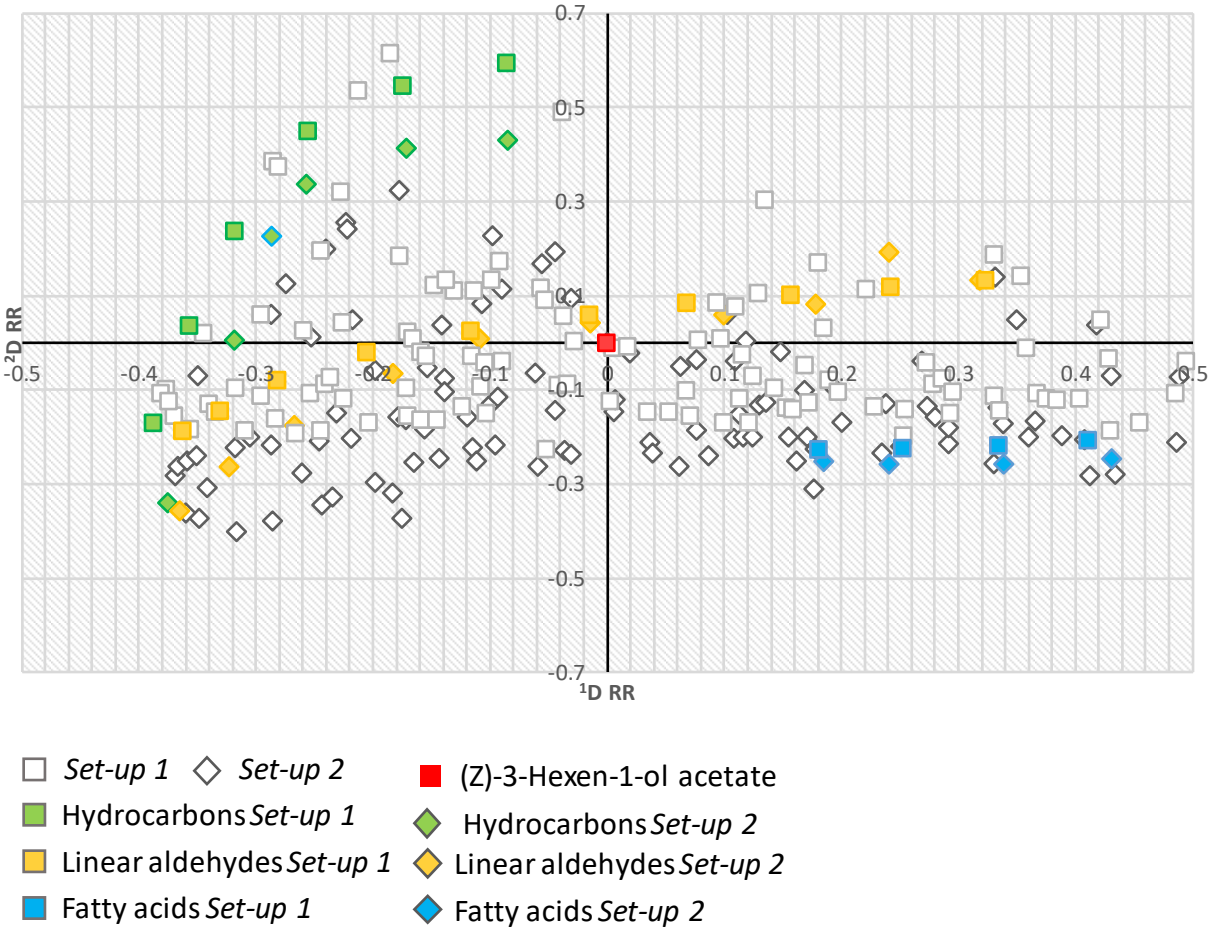
Table 5: template matching results for targeted template (supervised work-flow) from *Set-up 1* applied to *Set-up 2* samples and for feature template (unsupervised and automatic work-flow) built over *Set-up 1* samples and applied on *Set-up 2* samples.

	Targeted template (126 peaks)				***Feature* template (257 reliable peaks)**			
	Set-up 1		***Set-up 2***		***Set-up 1***		***Set-up 2***	
Samples	**%**	**Peaks n°**	**%**	**Peaks n°**	**%**	**Peaks n°**	**%**	**Peaks n°**
Baza-1	94.44	119	91.27	115	100.00	257	97.28	250
Baza-2	97.62	123	92.06	116	100.00	257	98.44	253
Baza-3	98.41	124	95.24	120	100.00	257	98.83	254
Baza-4	100.00	126	96.03	121	99.22	255	96.86	247
Bena-1	93.65	118	88.89	112	100.00	257	98.44	253
Bena-2	96.03	121	91.27	115	100.00	257	96.50	248
Bena-3	97.62	123	92.06	116	98.83	254	99.22	255
Bena-4	98.41	124	92.86	117	100.00	257	98.05	252
Average	*97.02*	*122*	*92.46*	*116*	*99.75*	*256*	*97.95*	*252*
RSD%	*2.22*	*2.22*	*2.47*	*2.47*	*0.45*	*0.46*	*0.99*	*1.22*

Figure 1



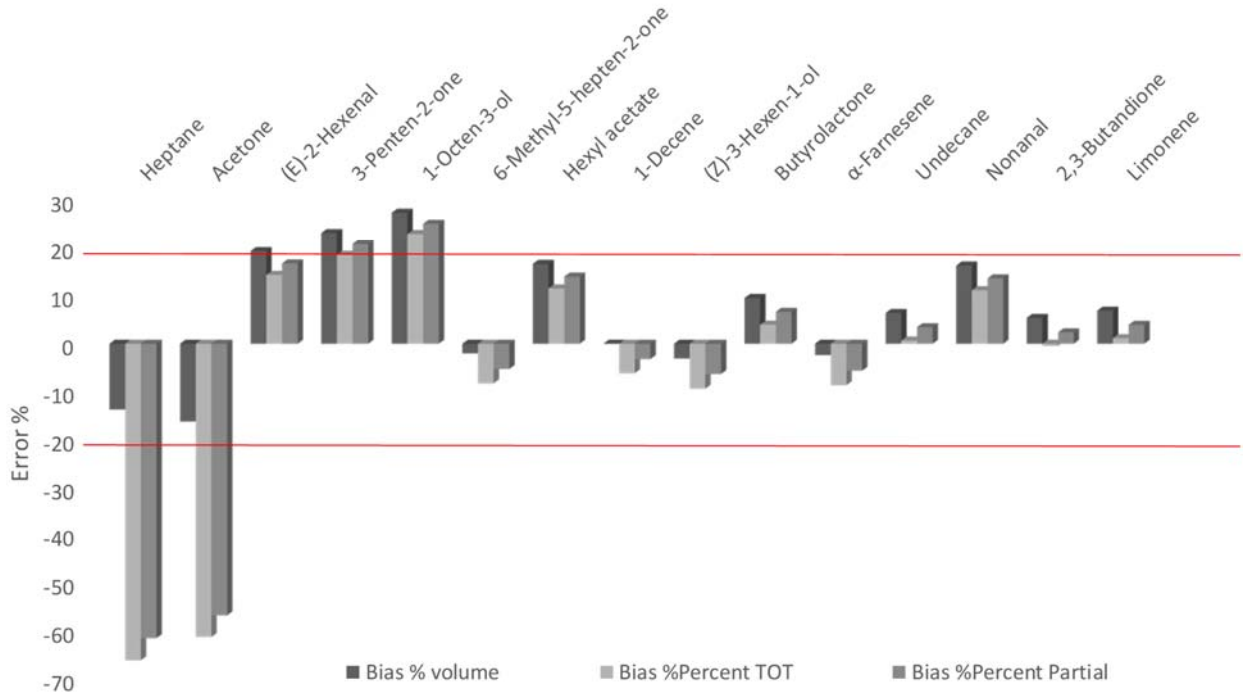
631 Figure 2



632

633

634 Figure 3

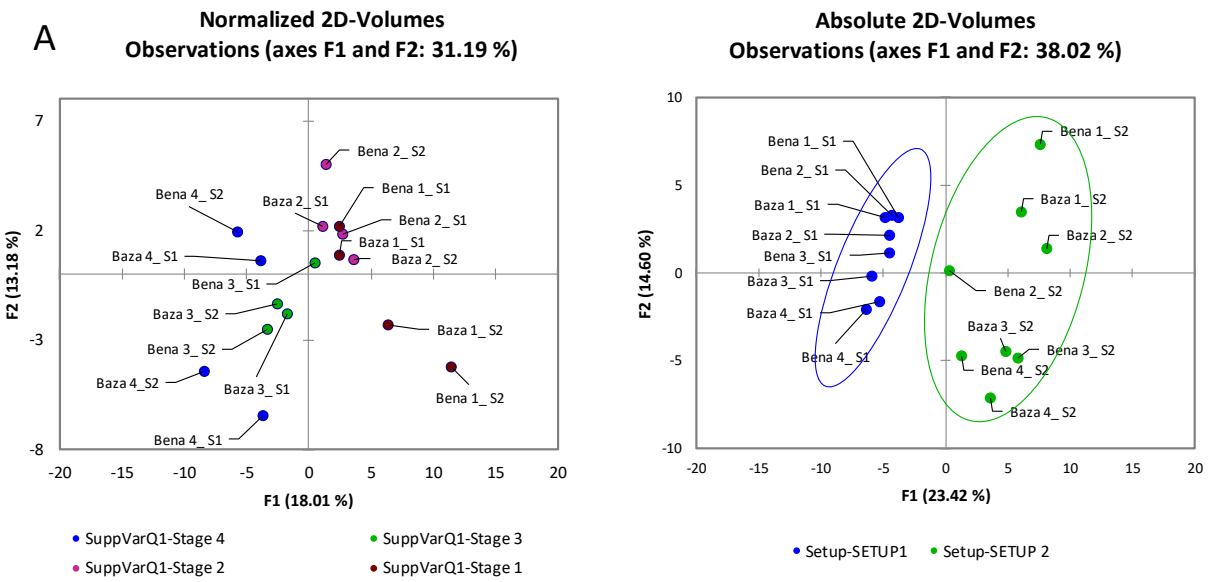


635

636

637

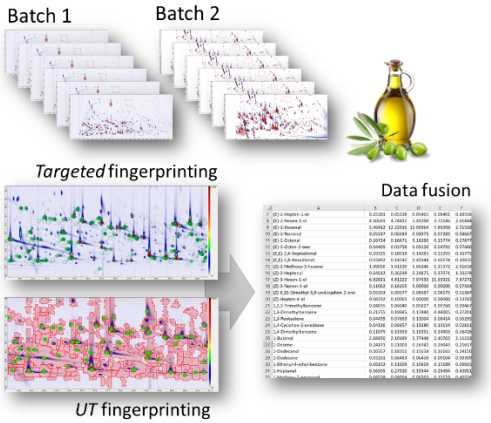
638 Figure 4



639

640

641 For Table of Contents Only



642
643
644