Adding extra-dimensions to hazelnuts primary metabolome fingerprinting by comprehensive two-dimensional gas chromatography combined with time-of-flight mass spectrometry featuring tandem ionization: Insights on the aroma potential

Adding extra-dimensions to hazelnuts primary metabolome fingerprinting by comprehensive two-dimensional gas chromatography combined with time-of-flight mass spectrometry featuring tandem ionization: insights on the aroma potential

Marta Cialiè Rosso[1], Maria Mazzucotelli[1], Carlo Bicchi[1], Melanie Charron[2], Federica Manini[2], Roberto Menta[2], Mauro Fontana[2], Stephen E. Reichenbach[3,4] and Chiara Cordero[1*]

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

[2]*Soremartec Italia Srl,* Ferrero Group, Alba (CN), Italy

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

[4]GC Image LCC, 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

Abstract

The information potential of comprehensive two-dimensional gas chromatography combined with time of flight mass spectrometry (GC×GC-TOFMS) featuring tandem hard (70 eV) and soft (12 eV) electron ionization is here applied to accurately delineate high-quality hazelnuts (*Corylus avellana* L.) primary metabolome fingerprints. The information provided by tandem signals for untargeted and targeted 2D-peaks is examined and exploited with pattern recognition based on template matching algorithms. EI-MS fragmentation pattern similarity, base-peak *m/z* values at the two examined energies (i.e., 12 and 70 eV) and response relative sensitivity are adopted to evaluate the complementary nature of signals.

As challenging bench test, the hazelnut primary metabolome has a large chemical dimensionality that includes various chemical classes such as mono- and disaccharides, amino acids, low-molecular weight acids, and amines, further complicated by oximation/silylation to obtain volatile derivatives.

Tandem ionization provides notable benefits including larger relative ratio of structural informing ions due to limited fragmentation at low energies (12 eV), meaningful spectral dissimilarity between 12 and 70 eV (direct match factor values range 222-783) and, for several analytes, enhanced relative sensitivity at lower energies. The complementary information provided by tandem ionization is exploited by untargeted/targeted (*UT*) fingerprinting on samples from different cultivars and geographical origins. The responses of 138 *UT*-peak-regions are explored to delineate informative patterns by univariate and multivariate statistics, providing insights on correlations between known precursors and (key)-aroma compounds and potent odorants. Strong positive correlations between non-volatile precursors and odorants are highlighted with some interesting linear trends for: 3-methylbutanal with isoleucine (R^2 0.9284); 2,3-butanedione/2,3-pentanedione with monosaccharides (fructose/glucose derivatives) (R^2 0.8543 and 0.8860); 2,5-dimethylpyrazine with alanine (R^2 0.8822); and pyrroles (1H-pyrrole, 3-methyl-1H-pyrrole, and 1H-pyrrole-2-carboxaldehyde) with ornithine and alanine derivatives (R^2 0.8604). The analytical work-flow provides a solid foundation for a new strategy for hazelnuts quality assessment because aroma potential could be derived from precursors' chemical fingerprints.

Key-words

comprehensive two-dimensional gas chromatography, *UT* fingerprinting, template matching, time-of-flight mass spectrometry featuring tandem ionization, high quality hazelnuts, aroma potential

1. Introduction

Since its introduction, comprehensive two-dimensional gas chromatography (GC×GC) coupled with mass spectrometry (MS) has shown great potential in profiling and fingerprinting investigations [1] in food [2–6]. 2D patterns of separated compounds well describe sample chemical dimensionality [7].

Although GC×GC is the technique of election to investigate food volatiles, it also enables effective studies on semi-volatiles and on non-volatiles suitably, after derivatization to improve thermal stability and volatility. Examples of research in this direction are the characterization of lipid fractions by GC×GC-MS/FID [8,9], fatty acids methyl esters (FAMEs) profiling [10–12], and primary metabolites in function of harvesting and storage practices [13,14].

The current study aims at developing a methodology based on GC×GC-TOF MS that combines profiling and fingerprinting strategies [1] based on 2D-data patterns, to comprehensively map hazelnuts key-aroma compounds precursors and relevant primary metabolites (amino acids, reducing sugars and polyalcohols, organic acids etc..) from raw fruits. To deal with the great complexity of primary metabolome, variable electron ionization (EI) MS is examined. The patented technology, termed tandem ionization™ [Select eV™ - US patent number 9,786,480], operates with variable-energy EI across single analytical runs. The acquisition is therefore done by time-switching between two ionization energies so that tandem data streams are generated and acquired iteratively. In the ion source, a high potential difference accelerates electrons away from the filament while reducing their energy before their entrance in the ion chamber [15]. This results in a more efficient ionization and a reduced loss of sensitivity at low-energy EI with enhanced intensity for structure-diagnostic ions [16].

Recent studies, have demonstrated that tandem ionization combining 14 and 70 eV provided successful discrimination of isomeric species in complex mixtures of motor oil samples [17] and in blood volatiles [18]. Additional benefits are derived from dedicated tandem data processing procedures. Freye *et al*. [19] applied tile-based Fisher ratio analysis on fused tandem signals to detect analytes spiked at 50 mg/kg level in diesel samples. Cordero *et al.* [16] extended the combined untargeted/targeted fingerprinting (*UT* fingerprinting) approach [20,21] to cocoa volatilome analysis. They confirmed that tandem signals, when summed together, gave better performances, in terms of samples discrimination and classification. In addition, analytes with a reduced fragmentation at 12 and 14 eV gave higher signal-to-noise ratio at lower ionization voltages also extending the dynamic range of measurements.

Challenges posed by hazelnuts primary metabolome fingerprinting are related to the large chemical dimensionality [22], that includes different chemical classes as mono- and disaccharides, amino

acids, low-molecular weight acids, and amines, and are furtherly complicated by oximation/silylation to obtain volatile derivatives. However, the information embedded in primary metabolites chemical signatures could be of high interest if correlated with hazelnuts hedonic qualities. Several aroma active compounds [23–26] have known mechanisms of formation from non-volatile precursors [25–27].

Intriguing empirical evidence on the instability, and/or evolution, of the volatile metabolome during shelf-life and storage of hazelnuts was discussed by Cialiè Rosso *et al.* [28]. In that study, GC×GC-MS, combined with advanced fingerprinting based on template matching, highlighted informative patterns of odorants strongly correlated, in raw hazelnuts, to post-harvest drying practices and storage conditions within a 12-month shelf-life [28]. At the same time, the aroma potential of nuts was studied by mapping volatiles fingerprints after lab-scale roasting across the shelf-life (e.g., 12 months after harvest). Results obtained with two origins (Turkish *Ordu* and Italian Nocciola Romana or Tonda Gentile Romana PDO - IT/PDO/0005/0573), showed that volatile patterns were greatly impacted by storage conditions. Samples stored under modified atmosphere (99% N_2 – 1% O_2) at 5°C were characterized by higher abundance of alkyl pyrazines, Streker aldehydes, and ketones compared to samples kept in a storage atmosphere richer in oxygen (78% N_2 – 21% O_2) and/or at ambient temperature (18°C). Most components are Maillard reaction products and/or formed by thermal degradation of reducing sugars, so it was hypothesized that storage would have impacted the distribution/availability of some non-volatile precursors in the fruits.

In this study, for the first time, GC×GC-TOF MS featuring tandem ionization is tested for its effectiveness on hazelnuts primary metabolome fingerprinting. In particular, spectral similarity/dissimilarity between high (70 eV) and low (12 eV) ionization energy is examined, including complementary information potential, sensitivity, and dynamic range of response. The analytical strategy is then validated for its informative potential and reliability in providing robust data for correlation studies with potent odorants distribution.

2. Experimental

2.1 Chemicals and reference solutions

Pure standards of *n*-alkanes (from n-C9 to n-C25) for system evaluation and linear retention indexes (I^T) determination and α-tujone for volatiles internal standardization (IS) were from Merck (Milan, Italy).

The mixture of *n*-alkanes for the I^T solution was prepared in cyclohexane at a concentration of 100 mg/L; internal standard (IS) α-thujone solution was prepared in diethyl phthalate (Sigma Aldrich 99% of purity) at a concentration of 100 mg/L.

Pure standards for identity confirmation of pyruvic acid, lactic acid, malonic acid, acetoacetic acid, phosphoric acid, succinic acid, glyceric acid, fumaric acid, malic acid, citric acid, Alanine-Ala, Asparagine-Asn, Aspartic acid-Asp, Cysteine-Cys, Glutamic acid-Glu, Glycine-Gln, Isoleucine-Ile, Leucine-Leu, Lysine-Lys, Methionine-Met, Ornithine-Orn, Phenylalanine-Phe, Proline-Pro, Serine-Ser, Threonine-Thr, Tryptophan-Trp, Tyrosine-Tyr, Valine-Val, glycerol, xylitol, mannitol, myo-inositol, fructose, glucose, saccharose, and the internal standards (IS), 4- chlorophenylalanine (quality control - QC for derivatization) and 1,4-dibromobenzene (QC for GC normalization), were from Merck (Milan, Italy).

Derivatization reagents and HPLC grade solvents: O-methylhydroxylamine hydro- chloride (MOX), (N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA) methanol, pyridine, *n*-hexane, di- chloromethane, and toluene were from Merck (Milan, Italy).

2.2 Hazelnut samples

Commercial grade samples of raw hazelnuts (*Corylus avellana* L.) with uniform caliber of 13 mm and harvested in 2016 were supplied by Soremartec Italia Srl (Alba-CN, Italy). They were from different geographical areas: (*a*) mono-cultivar Nocciola Romana (TR), also known as Tonda Gentile Romana, a Protected Denomination Origin (PDO) product (EU Quality registration code IT/PDO/0005/0573) here defined as *Roman*; (*b*) mono-cultivar Tonda Gentile delle Langhe (TGL) or Nocciola del Piemonte, a Protected Geographical Indication (IGP) product here defined as *Piedmont* and, (*c*) Turkish blend harvested in the Ordu (OR) region with *Tombul*, *Palaz* and *Çakildak* cultivars here defined as *Ordu*.

Raw hazelnuts were grinded after freezing with liquid nitrogen and stored at -18 °C before analysis. Lab-scale roasting was by a ventilated oven for 15 minutes at 160 °C on 100 g aliquots; these conditions were set in a previous study and optimized to develop major key-odorants [26] and roasting marker compounds [29].

2.3 Primary metabolites extraction and derivatization

2.3.1 Defatting

Micro-scale defatting was conducted in sealed glass vials (20 mL volume) on 1.00 g of ground hazelnuts aliquots with 5.0 mL of *n*-hexane as extraction solvent at ambient temperature and with the aid of ultrasound (US - 40 KHz ± 5%). Steps of 15' were set and exhaustiveness was verified after 6-7 successive extractions with fresh *n*-hexane aliquots. To avoid any contamination of the hydrophilic phase and to ensure efficient derivatization in the next sample preparation step, 10 extractions were set in the final protocol for samples defatting. A schematic diagram of the extraction/derivatization procedure is provided as **Supplementary Figure 1 – SF1** together with experimental results on fat fraction extraction yields on a representative sample (**Supplementary Figure 2 – SF2**). The defatted hazelnut powder was then kept under a gentle stream of nitrogen to remove residual solvent and stored at -18°C before extraction and derivatization.

2.3.2 Extraction of primary metabolites

An aliquot of 0.100 g of defatted hazelnuts was placed in a centrifuge glass tube with 5.0 mL of H_2O/CH_3OH (98:2 v/v) mixture. Extraction was conducted at ambient temperature and with the aid of ultrasound (US - 40 KHz ± 5%) for 15'. Exhaustiveness was verified after 3-10 successive extractions, depending on the targeted metabolite, with fresh solvent aliquots. A schematic diagram of the extraction/derivatization procedure is provided as **Supplementary Figure 1 – SF1** together with experimental results on targeted primary metabolites extraction yields **Supplementary Figure 2- SF2**. After extraction, centrifugation was carried out at 5,500 rpm for 10 min, the supernatant was then carefully collected and filtered with Nylon HPLC filters with 20 µm pores.

2.3.3 Derivatization

1 mL of the water-methanol extract from the first four extractions were collected together, spiked with 20 µL of 4-chlorophenylalanine solution (4 mg/mL in CH_3OH) and dried under a gentle stream of nitrogen into 1.5 mL glass vials. 45 µL of MOX solution (20 mg/mL in pyridine) were added and metoximation reaction was carried out at 60 °C for 2 h. Lastly, 60 µL of (N,O-bis(trimethylsilyl)trifluoroacetamide - BSTFA were added to the reaction mixture. The sylilation reaction was carried out at 60 °C for 1 h. At the end of the derivatization step, the reaction mixture was spiked with 20 µL of 1,4-dibromobenzene (IS 1 g/L in CH_2Cl_2); an additional 75 µL of CH_2Cl_2 were added up to a final volume of 200 µL.

For primary metabolites identity confirmation, 1.00 mL of primary metabolites standards mixture (listed in *section 2.1*) was submitted to the derivatization procedure and analyzed under conditions described in *section 2.5*.

2.4 Headspace Solid Phase Microextraction devices and sampling conditions

Automated HS-SPME sampling was performed using a MPS-2 multipurpose sampler (Gerstel, Mülheim a/d Ruhr, Germany) installed on the GC×GC–TOF-MS system. SPME fibers, divinylbenzene/carboxen/polydimethyl siloxane (DVB/CAR/PDMS) df50/30m – 2 cm, were from Supelco (Bellefonte, PA, USA). Fibers were conditioned before their use as recommended by the manufacturer. The IS (α-thujone) used for peak response normalization was pre-loaded into SPME fiber before sampling by exposing the extraction device (e.g., the SPME fiber) to 5 µL of ISTD standard stock solution for 20 min at 50°CC. Hazelnut samples were frozen before milling, using liquid nitrogen, to ensure homogeneous particle size distribution thus stored at −80°C until analyzed. Samples were exactly weighed (1.50 g) in headspace glass vials (20 mL) and submitted to headspace sampling for 40 min at 50°C.

2.5 GC×GC-TOF MS featuring Tandem Ionization: instrument set-up and experimental conditions

GC×GC analyses were performed on an Agilent 7890B GC unit coupled with a Bench TOF-Select™ system (Markes International, Llantrisant, UK) featuring Tandem EI. Hard ionization at 70 eV was set for identity confirmation while 12 eV was applied to explore spectral complementarity. The ion source and transfer line were set at 290°C. The MS optimization option was set to operate in Tandem Ionization with a mass range between 35 and 550 *m/z*; data acquisition frequency was 50 Hz per channel; filament voltage was set at 1.60 V.

The system was equipped with a two-stage KT 2004 loop thermal modulator (Zoex Corporation, Houston, TX) cooled with liquid nitrogen controlled by Optimode™ V.2 (SRA Instruments, Cernusco sul Naviglio, MI, Italy). The hot jet pulse time was set at 280 ms, modulation period was 5 s, and cold-jet total flow was progressively reduced with a linear function from 30% of Mass Flow Controller (MFC) at initial conditions to 5% at the end of the run.

2.6 GC×GC columns and settings for primary metabolites profiling

The column set was configured as follows: ^{1}D DB-5 column (95% polydimethylsiloxane, 5% phenyl; 30 m × 0.25 mm d_c, 0.25 µm d_f) coupled with a ^{2}D OV1701 column (86% polydimethylsiloxane, 7% phenyl, 7% cyanopropyl; 2 m × 0.1 mm d_c, 0.10 µm d_f), from J&W (Agilent, Little Falls, DE, USA). The first 0.80 m of the ^{2}D column, connected in series to the 1D column by a silTite µ-union (Trajan Scientific and Medical, Ringwood, Victoria, Australia), were wrapped in the modulator slit and used as loop-capillary for cryogenic

modulation. The carrier gas was helium at a constant flow of 1.6 mL/min. The oven temperature program was from 75°C (1 min) to 290°C (15 min) at 4°C/min.

For primary metabolites profiling, 2.0 µL of the derivatized solution (section *2.3.3*) was analyzed under the following conditions: split/splitless injector in split mode, split ratio 1:20, injector temperature 290°C. The *n*-alkanes liquid sample solution for I^T determination was analyzed under the following conditions: split/splitless injector in split mode, split ratio 1:50, injector temperature 290°C, and injection volume 1 µL.

2.7 GC-MS configuration and settings for volatiles profiling

GC-MS profiling of volatiles from lab-scale roasted samples was by a system configured as follows: Agilent 7890 GC unit coupled to an Agilent 5977B MS detector (Agilent, Little Falls, DE. USA) with High Efficiency Source (HES) operating in EI mode at 70 eV. The transfer line was set to 270°C, HES Tune was applied and the scan range was set to m/z 35-350 with a scanning rate of 2,500 *amu*/s.

A Solgel-Wax™ capillary column was used (100% polyethylene glycol, 30 m × 0.25 mm dc, 0.25 µm df) Trajan Scientific and Medical (Ringwood, Victoria, Australia).

The analytes sampled by HS-SPME were thermally desorbed from the fiber for 5 min, directly into the GC injector, under the following conditions: split mode, split ratio 1/20, injector temperature 250°C. The carrier gas was helium, at a constant flow rate of 1.0 mL/min. The oven temperature program was 40°C (1 min) to 180°C at 3°C/min and to 240°C at 15°C/min (5 min). MS source was set at 250°C while MS quadrupole was set at 180°C.

2.8 Method performance parameters: retention times and response repeatability

Method validation was run on a three-weeks basis and aimed at the evaluation of repeatability of retention times and *UT* peaks response precision [30]. Retention times in both chromatographic dimensions (^1t_R and ^2t_R) were collected from the *UT* peak-regions at 70 eV of analyzed samples (3 cultivar/origins × 5 extraction/derivatization batches × 2 analytical replicates) for a total of 30 analytical runs processed. Results are reported as relative standard deviation (RSD) in **Supplementary Table 1 – ST1** with quite good retention-times stability with an average RSD of 4.78E-03 for ^1t_R and 1.80E-02 for ^2t_R. Response repeatability was calculated on all UT peak-regions at 70 eV and was based on normalized responses from Piedmont samples (5 extraction/derivatization batches × 2 analytical replicates). As reported in in **Supplementary Table 1 – ST1,** repeatability RSD averaged of 0.10, with a minimal of 0.03 for feature (35) and a maximum of 0.20 for (238), both near the end of the run.

2.9 Data acquisition and 2D data processing

GC×GC data were acquired by TOF-DS software (Markes International, Llantrisant, UK), 1D-GC-MS data were acquired and processed by Enhanced MassHunter (Agilent, Little Falls, DE, USA). GC×GC-TOF MS data were processed using GC Image GC×GC Edition, ver 2.8 (GC Image, LLC, Lincoln NE, USA). Data elaboration and results visualization were by XL-Stat (Addinsoft Inc, New York, USA) and by open source Gene-E (Broadinstitute.org).

3. Results and Discussion

In this study, hazelnut primary metabolome fingerprints are, for the first time, explored by GC×GC-TOF MS featuring tandem ionization and the information potential embedded in 2D-patterns evaluated in terms of both discrimination power between cultivars and origins and aroma precursors distribution. In the panorama of existing studies aimed at delineating hazelnuts nutritional quality, there is a lack of systematic investigations on non-volatile aroma precursors distribution studied for their correlation to potent odorants formed after technological roasting. Alasalvar *et al.* [31] studied the proximate composition, minerals, vitamins, dietary fiber, amino acids, and taste active components (free amino acids, sugars, and organic acids) distribution in *Tombul* hazelnuts. They applied a multi-platform analytical approach to profile major and essential nutrients, writing that the *Tombul* hazelnut would "*serve as a good source of vital nutrients and taste active components*" [31]. The study then was extended to other cultivars/origins [32,33] and also to other tree nuts [34] by also integrating the analytical profiling with bio-active phytochemicals [35].

The first attempt to correlate hazelnut metabolome and its volatile composition was by Kiefl [26], who, based on available literature, reviewed the most relevant non-volatile precursors responsible of the formation of key-aroma compounds and the reaction pathways in hazelnuts after dry-roasting. However, the study mostly focused on the volatile metabolome, lacking experimental evidence of the expected strong correlations between precursors and reaction products.

In this study, we designed an analytical approach based on a single analytical platform, to map the characteristic distribution (quali-quantitative) of informative analytes selected from those playing a major role in defining hazelnuts aroma potential and sensory quality. Reactions, promoted by dry-roasting, include carbohydrates caramelization, consisting of dehydration and isomerization to form α-dicarbonyl compounds (2,3-butanedione and 2,3-pentanedione), and furanones (5-hydroxymethyfurfural and 4-hydroxy-2,5-dimethyl-3(2H)-furanone). 2-/3-Methylbutanal and phenylacetaldehyde are formed by Strecker degradation of amino acids in presence of α-dicarbonyls from Ile, Leu and Phe respectively [36]. Moreover, α-amino carbonyl compounds, resulting from Strecker reaction, are pyrazine precursors; their dimerization leads to the formation of dihydropyrazines that can either oxidize into pyrazines or react

with aldehydes to generate substituted pyrazines [37]. Additionally, reactions between Ala, Arg, Lys, Pro and Orn, and sugar degradation products (i.e., deoxyosones) forms 2-acetyl and 2-propyonilpyrroline that can be further oxidized to pyrrole derivatives [38]. Finally, thermal degradation of primary metabolites also can affect the lipid fraction, promoting autoxidation and formation of secondary products such as saturated and unsaturated aldehydes (hexanal, octanal, (*Z*)-2-octenal, (*E*)-2-octenal, (*Z*)-2-nonenal, (*Z*)-2-decenal and (*E,E*)-2,4-nonadienal, etc.).

The analytical strategy includes multi-analyte profiling and fingerprinting by GC×GC-TOF MS after oximation/silylation of non-volatile precursors. Tandem ionization is evaluated by combining 70 and 12 eV ionization energies to improve the confidence on analytes identification and to validate the comparative capacity of 2D-fingerprints. The high energy (70eV) enables confident identification based on the similarity of characteristic fragmentation patterns; the lower energy (12eV) provides complementary information and maximizes the relative intensity of structural informative fragments [16–18].

Experimental data is based on a selection of three relevant hazelnuts cultivar/origins (*Roman, Piedmont,* and *Ordu*) from harvest year 2015 analyzed at time zero (i.e. after in-field post-harvest drying). Five different subsets of samples which were independently defatted/extracted/derivatized and two analytical replicates were run for each.

The next sections include: (*a*) a discussion of the informative power of the approach and on the coverage of the non-volatile precursors, e.g., primary metabolites, of interest; (*b*) an evaluation of the complementary nature of tandem signals and their synergy in revealing compositional differences between samples; (*c*) an evaluation of existing correlations between primary metabolites and volatiles generated by lab-scale roasting with some insights on specific compounds generating key-aromas.

3.1 Fingerprinting information power

In the development of the analytical protocol, reference studies based on GC-MS and GC×GC-MS [13,14,39–41] were considered for their potential to cover the chemical dimensions [22] of interest. Extraction and derivatization were therefore validated for their exhaustiveness and efficacy toward aminoacids, sugars and polyalcohols and some low-molecular weight acids. Exhaustiveness was evaluated by analysing each aliquot of the ten successive extractions until none of the targeted primary metabolites (including all silylated forms) were detected. Results are illustrated visually in **Supplementary Figure 2 - SF2** as normalized responses (Normalized 2D-Peak Volume) while some considerations about derivatization issues are commented and supported by dedicated references.

B

Experimental results indicate that AA are almost 100% recovered after four extractions while for most of the acids, the exhaustive recovery is after 5 extractions. Exceptions are phosphoric acid, malic acid, lactic acid, and citric acid. According to literature, sugars and polyalcohols are major metabolites (excluding the fat components) and, for most of them, a dedicated procedure would be necessary to achieve their complete extraction. However, the first four aliquots were collected and unified before oximation/silylation to obtain more representative extracts.

The application of the combined untargeted/targeted (*UT*) fingerprinting procedure [20,42] on the set of the 30 2D-patterns acquired at 70 eV (3 cultivars/origins × 5 extraction/derivatization batches × 2 analytical replicates) revealed 138 *UT* peak-regions included in a UT *feature* template. They contained 108 targeted peaks putatively identified through their 70 eV EI-MS fragmentation pattern, by comparing them to those collected in commercial and in-house databases and for which a Direct Match Factor (DMF) ≥ 900 was obtained by applying the NIST Similarity match algorithm [43] on 2D-peak spectrum. The latter is the average spectrum obtained from the highest modulation within the 2D-peak region. Positive identifications were accepted for candidates with an experimental ^{1}D I^{T} ± 15 units tolerance. Within this list of 110 known analytes, 25 were confirmed by reference standards (see *section 2.1*). The identities of 30 analytes could not be confirmed and are tagged with unique numerical identifiers [(#)] Their re-alignment across the 30 2D-patterns was by template matching and by applying retention times and mass spectral similarity constraints ($^{1}t_R$ and $^{2}t_R$ ± 15% and DMF ≥ 800) with affine transformation [44,45]. The complete list of *UT* peak-regions together with their average retention times ($^{1}t_R$ min, $^{2}t_R$ sec), experimentally determined ^{1}D I^{T} values and tabulated ones (NIST database [43]), is reported in the **Supplementary Table 1** – **ST1**.

The next section discusses the complementary nature of tandem signals acquired at 70 and 12 eV and their information power in the chemical fingerprinting of hazelnut primary metabolome.

3.2 Tandem ionization: the complementary nature of tandem signals

The impact of different ionization voltages on spectral profiles and, thereby, on total ion response can be highlighted by applying datapoint features fingerprinting [46]. The approach compares the response, or the relative response, between image pairs based on a pixel-basis and promptly evidences those pattern regions in which detector responses greatly varies between the two data streams. **Figure 1** reports the comparative visualization, rendered as the *colorized fuzzy ratio,* between the 12 eV data stream (*analyzed* image) and the 70 eV signal (*reference* image) from a *Piedmont* hazelnut extract. Analyte relative response in the two detector signals is highlighted by color coding (green, red and light-grey).

Regions colored in green indicate analytes for which the relative intensity (Signal-to-Noise) was higher in the *analyzed* image (12 eV) and those colored in red indicate analytes with a higher relative intensity in the *reference* image (70 eV).

Insert Figure 1 here

UT peak-regions corresponding to some monosaccharides (glucose and fructose derivatives), sucrose and polyalcohols (xylitol and myo-inositol), and AA, showed a higher absolute and relative response on the 12 eV signal. This suggests that the lower ionization energy would have a great impact on analyte fragmentations. This evidence has been confirmed by calculating DMF and Reverse Match Factors (RMF) values on a selection of analytes at 12 eV and 70 eV. They are listed in **Table 1** together with base peak (BP) information at the two ionization energies and SNR values.

The DMF values averaged 608 with minimal values for Valine 2TMS derivative – 222, Serine 3TMS – 253, Valine TMS derivative – 254, and Alanine 2TMS – 317. The RMF averaged 657, slightly higher because the similarity computation does not include the analyzed spectra (12 eV) fragments not present in the reference at 70 eV. Smaller RMF values were for Valine 2TMS derivative – 228, Serine 3TMS – 374, Alanine 2TMS – 376, and Glutaric acid 2TMS – 389. Compared to average values registered in a previous study focused on the cocoa volatilome [16], for which DMF averaged 779 and RMF 787, spectral dissimilarity is higher emphasizing the complementary nature of high and low ionization energies. This evidence suggests that a differential impact on sensitivity has to be expected (see below).

An additional parameter considered in this comparative process was the base peak (BP) *m/z* value. *UT*-peaks spectra were manually inspected, background subtraction and deconvolution were applied, when necessary, to isolate a clean reference spectrum from which BP could be defined. Results are reported in **Table 1**. For 53 of 72 *2D*-peaks (74%), the BP at 70 eV had lower *m/z* value when compared to that recorded at 12 eV (shown as negative values in the *BP Diff* column). This result is quite interesting; it indicates that structure-informing fragments are more prevalent at lower ionization, thus providing an additional basis to identify analytes and, furthermore, additional specificity could be explored when the two data streams are combined. To note, just 3 analytes (Glucopyranose 5TMS, Glycine 2TMS, and Serine 3TMS) presented a larger intensity BP at 70 eV (shown as a difference positive).

To further evaluate the complementary nature of tandem signals, relative sensitivity was considered through the ratio of the experimental SNR values registered at 12 and 70 eV data streams (data is the average value between data streams and replicates). Results indicate that for 38 of 138 *UT*-peaks (27%) listed in **Supplementary Table 1 – ST1**, the lower ionization energy signal has a larger relative intensity including several targeted analytes of interest for their role as aroma precursors. They include

sucrose 8TMS (SNR ratio 13), fructose 5TMS *syn* (SNR ratio 5), glucose 5TMS (SNR ratio 4), xylitol, 5TMS derivative (SNR ratio 4), malic acid 3TMS (SNR ratio 4), myo-inositol 6TMS derivative (SNR ratio 3), glycerol 3TMS (SNR ratio 3), glycine 3TMS (SNR ratio 2), ornithine 4TMS derivative (SNR ratio 2), and citric acid 4TMS derivative (SNR ratio 2).

Figure 2 reports the spectra of four analytes showing dissimilar fragmentation patterns: glucose in its cyclic form (i.e., glucopyranose 5TMS), Gly 3TMS, citric acid 4TMS, and succinic acid 2TMS. In all cases, the molecular ion was detected at both 70 and 12 eV. In the glucopyranose spectrum at lower ionization energy, the molecular ion (i.e, 540 m/z) was slightly larger (1.0E-02 *vs.* 9.8E-04 percent intensity). In general, spectra at 12 eV are dominated by structurally informing fragments whereas those at 70 eV are dominated by the fragments of silylating agent (i.e., TMS – 73 *m/z*) with low informing power. In the case of glucose and citric acid derivatives, the base peak (BP) between 12 (red) and 70 (blue) eV spectra is different. These results confirm the complementary nature of tandem ionization signals, a characteristic that is very useful for the accurate fingerprinting of complex samples. For analytes showing larger SNR values at 12 eV, the relative sensitivity of the method increases at lower energies and additional benefits could be also expected on the linearity range. At lower energies the amount of ionized molecules is reduced, compared to 70 eV, and the risk of detector saturation is minimized. Experimental evidence on the extended linearity range of responses for volatile compounds is discussed in an application of interest for the flavor and fragrance field [47].

Insert Figure 2 here

The *UT feature template* was applied to the 30 2D-patterns obtained at 70 eV and re-aligned to the 12 eV data streams, then response information from all *UT* peak-regions was extracted to exploit the complementary information from tandem signals in terms of fingerprinting information potential. The final data matrix was 60×144 (60 2D-patterns [3 cultivar/origins × 5 extraction/derivatization batches × 2 ionization energies × 2 analytical replicates] × 144 UT peak-regions). Results are visualized as heat-map in **Figure 3. Figure 3A** shows the 70 eV data stream and represents, in a color scale from blue to red, the normalized *UT* peak-regions volume distribution; **Figure 3B** shows, in a color scale from green to red, the 12 eV *UT* peak-regions. Hierarchical clustering (HC) is based on Euclidean distances and was applied after *Z*-score normalization of the data (i.e., subtract mean and divide by standard deviation).

Insert Figure 3 here

Primary metabolites fingerprints can independently cluster samples based on their cultivar/origin. The two data streams show a high consistency in their discrimination potential: both data streams closely cluster *Piedmont* and *Roman* samples although informative UT peaks contributing to this classification are

different (see rows/UT peaks HC clusters). In particular, *Ordu* samples are connoted by a lower relative concentration of primary metabolites (confirmed by both data streams). Only a few analytes show a higher relative abundance: Gly and tartaric acid within the known compounds set and (1199) and (1153) within the unknown compounds set. On the other hand, the discrimination between *Piedmont* and *Roman* samples is mainly driven by: *(a)* sugars that are more abundant in the *Piedmont* cultivar (e.g. glucose, galactose, maltose, and fructose (considering all derivatives); *(b)* AA like Trp, Orn, and Tyr more abundant in *Piedmont* and Leu, Ile, Met, Val, Phe, Pro, and pyroglutamic acid with a more intense response in the *Roman* samples; *(c)* acids such as lactic acid (monomer 2TMS and dimer 2TMS), glutaric acid, galacturonic acid, fumaric acid, tartaric acid, and oxalic acid which are more abundant in *Roman* samples. Univariate statistics, in box-plot diagrams, for a selection of *UT* peaks are reported in the **Supplementary Figure 3 – SF3**.

Tandem signals, when independently processed, enable cross validation of fingerprinting results. For some analytes, linear regression analysis between absolute responses recorded at 12 eV (dependent variable) and at 70 eV (independent variable) give good results ($R2 \geq 0.90$) as shown in **Figure 4** for some aroma precursors: glucose, Orn, Leu, and Ile. The coherent trend of analytes response validate the differential distribution of markers within sample set, while the different response factors between ionization channels, described by regression functions with slopes $\neq 1$, confirms their complementary nature.

The next section deals with the correlation between primary metabolites signatures and hazelnuts volatile metabolome, in particular with key-aroma compounds and marker analytes distribution after lab-scale roasting.

Insert Figure 4 here

3.3 Hazelnuts aroma potential and correlation with primary metabolites signatures

Samples were submitted to lab-scale roasting, designed in a previous study, to evaluate the aroma quality of roasted hazelnuts. The protocol includes a dry-roasting in a ventilated oven at 160°C for 15 minutes [28]. Roasted samples then were submitted to HS-SPME sampling to extract volatiles and GC-MS profiling to monitor quali-quantitative changes in volatile markers including potent odorants [26,29,48–50].

The list of targeted volatiles is provided in the **Supplementary Table 2 – ST2**, together with the experimental I^T, odor quality and odor threshold for markers. They include key-aroma compounds validated by sensomics [23]: 2- and 3-methylbutanal, with *malty* notes; 2,3-butanedione and 2,3-

pentanedione, with *buttery* odor quality; 3-methyl-4-heptanone, 5-methyl-(Z)-2-hepten-4-one and 5-methyl-(E)-2-hepten-4-one (i.e., filbertone), with characteristic *nutty* and *fruity* odors; the *earthy* pyrazines 2,3,5-trimethylpyrazine and 3,5-dimethyl-2-ethylpyrazine; acetic acid, responsible of the *sour* note; phenylacetaldehyde, with *honey* and *flowery* odors; 2-methylbutyric acid, with a *sweaty* note; and 4-hydroxy-2,5-dimethyl-3(2H)-furanone (i.e., furaneol), with the characteristic *sweet* and *caramel-like* note. Within the volatile fraction of roasted samples, several other compounds were targeted. They belong to the group of roasting indicators, i.e., compounds formed during thermal processing and for which a meaningful increase was registered during lab-scale roasting [29]. They include: carbonyl derivatives (2-methylpropanal, 2-butenal and 3-hydroxy-2-butanone), alkyl pyrazines (pyrazine, 2-methylpyrazine, 2,5-dimethylpyrazine, 2,6-dimethylpyrazine, ethylpyrazine, 2,3-dimethylpyrazine, 2-ethyl-6-methylpyrazine, 2-ethyl-5-methylpyrazine, 2,3,5-trimethylpyrazine and 3-ethyl-2,5-dimethylpyrazine), furanones (furfural, 5-methylfurfural, dihydro-2(3H)-furanone and furfuryl alcohol), and pyrroles (1(H)-pyrrole, 1-methyl-1(H)pyrrole and 1(H)-pyrrole-2-carboxaldehyde). The present discussion does not include secondary products of lipid oxidation that contribute to some odor qualities and inform about shelf-life stability, although a dedicated procedure to study the fat fraction has been developed.

The squared Pearson correlation coefficient (r) was calculated to evaluate the existence of positive correlations between primary metabolites and volatiles. The p-values computed for each coefficient afford testing the null hypothesis that r values are not significantly different from 0 and informs about the relevance of the correlation.

The data matrix included the normalized responses (i.e. normalized 2D peak-regions from GC×GC data and normalized chromatographic areas for 1D-GC data) for targeted primary metabolites and all volatiles respectively (see **ST1** and **ST2**). The resulting matrix was 30×162 features (30 2D-patterns [3 cultivar/origins × 5 extractions/derivatization batches × 1 ionization energy 70 eV × 2 analytical replicates] × 108 targeted peak-regions + 54 volatiles). Before processing, data were scaled by Pareto scaling [51] in order to reduce the relevance of large values while keeping data structure almost intact. The resulting Pearson correlation matrix, showing r values, is illustrated as heat-map in **Figure 5**. The color scale, from blue to red, has been adapted to emphasize (red color) r values ≥ 0.8, arbitrarily considered meaningful in this specific data-set for notable correlations. HC is based on "one minus Pearson correlation" value as metrics.

Insert Figure 5 here

Results show: (*a*) high *r* values (accompanied by meaningful *p*-values – α=0.05) within primary metabolites (**Figure 5 – box 1**) confirming fingerprinting results where *Piedmont* and *Roman* samples showed a higher relative abundance for most of the non-volatile precursors and metabolites compared to *Ordu* origin (see **Fig. 3A** and **SF3**); (*b*) high correlation values ($r \geq 0.8$ - **Figure 5 - box 2**) within volatiles including 3-methyl-4-heptanone, 2,3-pentanedione, (*E)*-3-penten-2-one, 3-hydroxy-2-butanone, benzenemethanol, furaneol, furfuryl alcohol, 3-methylbutanal, 1H-pyrrole, 2,3-dimethylpyrazine, ethylpyrazine, methylpyrazine, and trimethyl pyrazine; and (c) between primary metabolites (and/or non-volatile precursors) and volatiles with a high informing power in terms of aroma quality. Further details on these last correlations are discussed below.

These correlations were tested for their significance (*p*-value) and linearity; the coefficients of determination (R^2) of the regression were calculated, considering the precursor(s) as independent variable (*x*) and key-volatiles as dependent variable (*y*). **Figure 6A** reports the regression function between 3-methylbutanal and Leu (R^2 0.9577), **Fig. 6B** between 3-methylbutanal and Ile (R^2 0.9284), **Fig. 6C-D** between 2,3-butanedione and 2,3-pentanedione and the sum of fructose (Fructose 5TMS *syn*- and *anti*-forms) and glucose (Glucopyranose 5TMS and Glucose 5TMS) derivatives (R^2 0.8543 and 0.8860), **Fig. 6E** between 2,5-dimethylpyrazine and Ala (R^2 0.8822) and pyrroles (1H-pyrrole, 3-methyl-1H-pyrrole and 1H-pyrrole-2-carboxaldehyde) and the sum of Orn and Ala derivatives (R^2 0.8604).

Insert Figure 6 here

These results are interesting, and evidence a solid foundation for adoption of the proposed approach for a comprehensive primary metabolome fingerprinting as an informative tool to characterize aroma potential of hazelnuts. However, the approach needs validation through a more extensive sampling that includes further variables influencing primary metabolites distribution in hazelnuts: cultivar and origins should be accompanied by a selection of multiple harvest years, shelf-life, and storage conditions that also are relevant because of their known impact on the aroma potential [28].

4. Conclusions

The present study is focused on variable EI energy TOF MS adopted to extend the analytical dimensions of GC×GC for hazelnut primary metabolites profiling. By combining standard 70 eV and 12 eV, tandem data streams are generated and spectra show a complementary nature. This characteristic can be exploited in identification and structural elucidation studies. At low energy, structure informing fragments prevail while overall fragmentation is reduced, achieving, for some analytes, a higher relative sensitivity that results in a wider dynamic range of the method. Tandem data responses from *UT*-peaks in

complex chemical signatures are strongly correlated but exhibit different response factors that open interesting perspectives for quantitative studies.

Response data recorded from primary metabolites fingerprints in different cultivars/origins show good correlation to volatiles formed during dry-roasting and validate empirical observations from previous experiments [28]. Linear regressions with meaningful R^2 give solid foundations to the causal relationship between metabolome and sensobolome signatures. The proposed approach, once validated over a wider sampling design, would introduce a new concept in hazelnuts quality assessment while opening new perspectives for breeding studies and shelf-life quality evaluation.

Funding

The research was carried out thanks to the financial support of Soremartec Italia Srl, Alba (CN, Italy).

Compliance with ethical standards Notes

Prof. Stephen E. Reichenbach has a financial interest in GC Image, LLC.

Melanie Charron, Federica Manini, Roberto Menta and Mauro Fontana are employees of Soremartec Italia Srl, Alba (CN, Italy).

501 **References**

502 [1] C. Cordero, J. Kiefl, S.E. Reichenbach, C. Bicchi, Characterization of odorant patterns by
503 comprehensive two-dimensional gas chromatography: A challenge in omic studies, TrAC - Trends
504 Anal. Chem. 113 (2019) 364–378. doi:10.1016/j.trac.2018.06.005.

505 [2] M. Adahchour, J. Beens, R.J.J. Vreuls, U.A.T. Brinkman, Recent developments in comprehensive
506 two-dimensional gas chromatography (GC x GC). IV. Further applications, conclusions and
507 perspectives, TrAC - Trends Anal. Chem. 25 (2006) 821–840. doi:10.1016/j.trac.2006.03.003.

508 [3] M. Adahchour, L.L.P. van Stee, J. Beens, R.J.J. Vreuls, M.A. Batenburg, U.A.T. Brinkman,
509 Comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric
510 detection for the trace analysis of flavour compounds in food, J. Chromatogr. A. 1019 (2003) 157–
511 172. isi:000186503600014.

512 [4] C. Cordero, J. Kiefl, P. Schieberle, S.E. Reichenbach, C. Bicchi, Comprehensive two-dimensional gas
513 chromatography and food sensory properties: Potential and challenges, Anal. Bioanal. Chem. 407
514 (2015) 169–191. doi:10.1007/s00216-014-8248-z.

515 [5] P.Q. Tranchida, P. Donato, F. Cacciola, M. Beccaria, P. Dugo, L. Mondello, Potential of
516 comprehensive chromatography in food analysis, TrAC - Trends Anal. Chem. 52 (2013) 186–205.
517 doi:10.1016/j.trac.2013.07.008.

518 [6] P.Q. Tranchida, F.A. Franchina, P. Dugo, L. Mondello, Comprehensive two-dimensional gas
519 chromatography-mass spectrometry: Recent evolution and current trends, Mass Spectrom. Rev.
520 35 (2016) 524–534. doi:10.1002/mas.21443.

521 [7] F. Magagna, E. Liberto, S.E. Reichenbach, Q. Tao, A. Carretta, L. Cobelli, M. Giardina, C. Bicchi, C.
522 Cordero, Advanced fingerprinting of high-quality cocoa: Challenges in transferring methods from
523 thermal to differential-flow modulated comprehensive two dimensional gas chromatography, J.
524 Chromatogr. A. 1535 (2018) 122–136. doi:10.1016/j.chroma.2017.07.014.

525 [8] G. Purcaro, L. Barp, M. Beccaria, L.S. Conte, Fingerprinting of vegetable oil minor components by
526 multidimensional comprehensive gas chromatography with dual detection, Anal. Bioanal. Chem.
527 407 (2015) 309–319. doi:10.1007/s00216-014-8140-x.

528 [9] G. Purcaro, L. Barp, M. Beccaria, L.S. Conte, Characterisation of minor components in vegetable oil
529 by comprehensive gas chromatography with dual detection, Food Chem. 212 (2016) 730–738.
530 doi:10.1016/j.foodchem.2016.06.048.

531 [10] P.Q. Tranchida, A. Giannino, M. Mondello, D. Sciarrone, P. Dugo, G. Dugo, L. Mondello, Elucidation
532 of fatty acid profiles in vegetable oils exploiting group-type patterning and enhanced sensitivity of

comprehensive two-dimensional gas chromatography, J. Sep. Sci. 31 (2008) 1797–1802. doi:10.1002/jssc.200800002.

[11] P.Q. Tranchida, S. Salivo, I. Bonaccorsi, A. Rotondo, P. Dugo, L. Mondello, Analysis of the unsaponifiable fraction of lipids belonging to various milk-types by using comprehensive two-dimensional gas chromatography with dual mass spectrometry/flame ionization detection and with the support of high resolution time-of-flight mass , J. Chromatogr. A. 1313 (2013) 194–201. doi:10.1016/j.chroma.2013.07.089.

[12] P.Q. Tranchida, S. Salivo, F.A. Franchina, I. Bonaccorsi, P. Dugo, L. Mondello, Qualitative and quantitative analysis of the unsaponifiable fraction of vegetable oils by using comprehensive 2D GC with dual MS/FID detection, Anal. Bioanal. Chem. 405 (2013) 4655–4663. doi:10.1007/s00216-013-6704-9.

[13] C.H. Weinert, B. Egert, S.E. Kulling, On the applicability of comprehensive two-dimensional gas chromatography combined with a fast-scanning quadrupole mass spectrometer for untargeted large-scale metabolomics, J. Chromatogr. A. 1405 (2015) 156–167. doi:10.1016/j.chroma.2015.04.011.

[14] C. Mack, D. Wefers, P. Schuster, C.H. Weinert, B. Egert, S. Bliedung, B. Trierweiler, C. Muhle-Goll, M. Bunzel, B. Luy, S.E. Kulling, Untargeted multi-platform analysis of the metabolome and the non-starch polysaccharides of kiwifruit during postharvest ripening, Postharvest Biol. Technol. 125 (2017) 65–76. doi:10.1016/j.postharvbio.2016.10.011.

[15] Markes International, Select-eV: The next generation of ion source technology, Tech. Note. Applicatio (2016).

[16] C. Cordero, A. Guglielmetti, C. Bicchi, E. Liberto, L. Baroux, P. Merle, Q. Tao, S.E. Reichenbach, Comprehensive two-dimensional gas chromatography coupled with time of flight mass spectrometry featuring tandem ionization: challenges and opportunities for accurate fingerprinting studies, J. Chromatogr. A. (2019). doi:10.1016/j.chroma.2019.03.025.

[17] M.S. Alam, C. Stark, R.M. Harrison, Using Variable Ionization Energy Time-of-Flight Mass Spectrometry with Comprehensive GC??GC to Identify Isomeric Species, Anal. Chem. 88 (2016) 4211–4220. doi:10.1021/acs.analchem.5b03122.

[18] L.M. Dubois, K.A. Perrault, P.H. Stefanuto, S. Koschinski, M. Edwards, L. McGregor, J.F. Focant, Thermal desorption comprehensive two-dimensional gas chromatography coupled to variable-energy electron ionization time-of-flight mass spectrometry for monitoring subtle changes in volatile organic compound profiles of human blood, J. Chromatogr. A. 1501 (2017) 117–127.

565 doi:10.1016/j.chroma.2017.04.026.
566 [19] C.E. Freye, N.R. Moore, R.E. Synovec, Enhancing the chemical selectivity in discovery-based analysis
567 with tandem ionization time-of-flight mass spectrometry detection for comprehensive two-
568 dimensional gas chromatography, J. Chromatogr. A. 1537 (2018) 99–108.
569 doi:10.1016/j.chroma.2018.01.008.
570 [20] F. Magagna, L. Valverde-Som, C. Ruíz-Samblás, L. Cuadros-Rodríguez, S.E. Reichenbach, C. Bicchi,
571 C. Cordero, Combined untargeted and targeted fingerprinting with comprehensive two-
572 dimensional chromatography for volatiles and ripening indicators in olive oil, Anal. Chim. Acta. 936
573 (2016) 245–258. doi:10.1016/j.aca.2016.07.005.
574 [21] D. Bressanello, E. Liberto, M. Collino, F. Chiazza, R. Mastrocola, S.E. Reichenbach, C. Bicchi, C.
575 Cordero, Combined untargeted and targeted fingerprinting by comprehensive two-dimensional
576 gas chromatography: revealing fructose-induced changes in mice urinary metabolic signatures,
577 Anal. Bioanal. Chem. 410 (2018) 2723–2737. doi:10.1007/s00216-018-0950-9.
578 [22] J.C. Giddings, Sample dimensionality: A predictor of order-disorder in component peak distribution
579 in multidimensional separation, J. Chromatogr. A. 703 (1995) 3–15. doi:10.1016/0021-
580 9673(95)00249-M.
581 [23] J. Kiefl, P. Schieberle, Evaluation of process parameters governing the aroma generation in three
582 hazelnut cultivars (Corylus avellana L.) by correlating quantitative key odorant profiling with
583 sensory evaluation, J. Agric. Food Chem. 61 (2013) 5236–5244. doi:10.1021/jf4008086.
584 [24] J. Kiefl, G. Pollner, P. Schieberle, Supporting Information Sensomics Analysis of Key Hazelnut
585 Odorants (Corylus avellana L ., ' Tonda Gentile ') Using Comprehensive Two-Dimensional Gas
586 Chromatography in Combination with Time-of- Flight-Mass Spectrometry (GC × GC / TOF-MS), J.
587 Agric. Food Chem. 4 (2013) 1–20.
588 [25] J. Kiefl, G. Pollner, P. Schieberle, Sensomics analysis of key hazelnut odorants (Corylus avellana L.
589 'Tonda Gentile') using comprehensive two-dimensional gas chromatography in combination with
590 time-of-flight mass spectrometry (GC×GC-TOF-MS), J. Agric. Food Chem. 61 (2013) 5226–5235.
591 doi:10.1021/jf400807w.
592 [26] J. Kiefl, Differentiation of Hazelnut Cultivars (Corylus avellana L.) by Metabolomics and Sensomics
593 Approaches Using Comprehensive Two-dimensional Gas Chromatography Time-Of-Flight Mass
594 Spectrometry (GC×GC/TOF-MS), 2013. doi:10.1017/CBO9781107415324.004.
595 [27] J. Kiefl, Differentiation of Hazelnut Cultivars (Corylus avellana L.) by Metabolomics and Sensomics
596 Approaches Using Comprehensive Two-dimensional Gas Chromatography Time-Of-Flight Mass

Spectrometry (GC×GC/TOF-MS), Verlag Deutsche Forschungsanstalt für Lebensmittelchemie (DFA), Freising, 2013.

[28] M. Cialiè Rosso, E. Liberto, N. Spigolon, M. Fontana, M. Somenzi, C. Bicchi, C. Cordero, Evolution of potent odorants within the volatile metabolome of high-quality hazelnuts (Corylus avellana L.): evaluation by comprehensive two-dimensional gas chromatography coupled with mass spectrometry, Anal. Bioanal. Chem. 410 (2018) 3491–3506. doi:10.1007/s00216-017-0832-6.

[29] L. Nicolotti, C. Cordero, C. Bicchi, P. Rubiolo, B. Sgorbini, E. Liberto, Volatile profiling of high quality hazelnuts (Corylus avellana L.): Chemical indices of roasting, Food Chem. 138 (2013) 1723–1733. doi:10.1016/j.foodchem.2012.11.086.

[30] Eurachem, Eurachem Guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics., 2014. doi:978-91-87461-59-0.

[31] C. Alasalvar, F. Shahidi, C.M. Liyanapathirana, T. Ohshima, Turkish Tombul Hazelnut (Corylus avellana L.). 1. Compositional Characteristics , J. Agric. Food Chem. 51 (2007) 3790–3796. doi:10.1021/jf0212385.

[32] F. Seyhan, G. Ozay, S. Saklar, E. Ertaş, G. Satir, C. Alasalvar, Chemical changes of three native Turkish hazelnut varieties (Corylus avellana L.) during fruit development, Food Chem. 105 (2007) 590–596. doi:10.1016/j.foodchem.2007.04.016.

[33] S.K. Chang, C. Alasalvar, B.W. Bolling, F. Shahidi, Nuts and their co-products: The impact of processing (roasting) on phenolics, bioavailability, and health benefits - A comprehensive review, J. Funct. Foods. 26 (2016) 88–122. doi:10.1016/j.jff.2016.06.029.

[34] C. Alasalvar, F. Shahidi, Tree Nuts: Composition, Phytochemicals, and Health Effects, Chromatographia. 72 (2010) 589–589. doi:10.1365/s10337-010-1619-5.

[35] C. Alasalvar, B.W. Bolling, Review of nut phytochemicals, fat-soluble bioactives, antioxidant components and health effects, Br. J. Nutr. 113 (2015) S68–S78. doi:10.1017/S0007114514003729.

[36] N. Göncüoğlu Taş, V. Gökmen, Maillard reaction and caramelization during hazelnut roasting: A multiresponse kinetic study, Food Chem. 221 (2017) 1911–1922. doi:10.1016/j.foodchem.2016.11.159.

[37] P.V. Guerra, V.A. Yaylayan, Dimerization of azomethine ylides: An alternate route to pyrazine formation in the maillard reaction, J. Agric. Food Chem. 58 (2010) 12523–12529. doi:10.1021/jf103194k.

[38] A. Burdack-Freitag, P. Schieberle, Changes in the key odorants of Italian hazelnuts (Coryllus avellana L. Var. Tonda Romana) induced by roasting, J. Agric. Food Chem. 58 (2010) 6351–6359.

629 doi:10.1021/jf100692k.
630 [39] E.C.Y. Chan, K.K. Pasikanti, J.K. Nicholson, Global urinary metabolic profiling procedures using gas
631 chromatography-mass spectrometry, Nat. Protoc. 6 (2011) 1483–1499.
632 doi:10.1038/nprot.2011.375.
633 [40] M.M. Ulaszewska, C.H. Weinert, A. Trimigno, R. Portmann, C. Andres Lacueva, R. Badertscher, L.
634 Brennan, C. Brunius, A. Bub, F. Capozzi, M. Cialiè Rosso, C.E. Cordero, H. Daniel, S. Durand, B. Egert,
635 P.G. Ferrario, E.J.M. Feskens, P. Franceschi, M. Garcia-Aloy, F. Giacomoni, P. Giesbertz, R. González-
636 Domínguez, K. Hanhineva, L.Y. Hemeryck, J. Kopka, S.E. Kulling, R. Llorach, C. Manach, F. Mattivi,
637 C. Migné, L.H. Münger, B. Ott, G. Picone, G. Pimentel, E. Pujos-Guillot, S. Riccadonna, M.J. Rist, C.
638 Rombouts, J. Rubert, T. Skurk, P.S.C. Sri Harsha, L. Van Meulebroek, L. Vanhaecke, R. Vázquez-
639 Fresno, D. Wishart, G. Vergères, Nutrimetabolomics: An Integrative Action for Metabolomic
640 Analyses in Human Nutritional Studies, Mol. Nutr. Food Res. 63 (2019) 1–38.
641 doi:10.1002/mnfr.201800384.
642 [41] D. Chelius, K. Jing, A. Lueras, D.S. Rehder, T.M. Dillon, A. Vizel, R.S. Rajan, T. Li, M.J. Treuheit, P. V.
643 Bondarenko, Formation of pyroglutamic acid from N-terminal glutamic acid in immunoglobulin
644 gamma antibodies, Anal. Chem. 78 (2006) 2370–2376. doi:10.1021/ac051827k.
645 [42] S.E. Reichenbach, C.A. Zini, K.P. Nicolli, J.E. Welke, C. Cordero, Q. Tao, Benchmarking Machine
646 Learning Methods for Comprehensive Chemical Fingerprinting and Pattern Recognition, J.
647 Chromatogr. A. (2019). doi:10.1016/j.chroma.2019.02.027.
648 [43] NIST/EPA/NIH Mass Spectral Library with Search Program Data Version: NIST v17, (n.d.).
649 [44] S.E. Reichenbach, D.W. Rempe, Q. Tao, D. Bressanello, E. Liberto, C. Bicchi, S. Balducci, C. Cordero,
650 Alignment for Comprehensive Two-Dimensional Gas Chromatography with Dual Secondary
651 Columns and Detectors, Anal. Chem. 87 (2015) 10056–10063. doi:10.1021/acs.analchem.5b02718.
652 [45] R.D. W., R.S. E., T. Qingping, C. Chiara, R.W. E., Z.C. Alcaraz, Effectiveness of Global, Low-Degree
653 Polynomial Transformations for GCxGC Data Alignment, Anal. Chem. 88 (2016) 10028–10035.
654 doi:10.1021/acs.analchem.6b02254.
655 [46] S.E. Reichenbach, X. Tian, C. Cordero, Q. Tao, Features for non-targeted cross-sample analysis with
656 comprehensive two-dimensional chromatography, J. Chromatogr. A. 1226 (2012) 140–148.
657 doi:10.1016/j.chroma.2011.07.046.
658 [47] C. Cordero, E. Gabetti, C. Bicchi, P. Merle, E. Belhassen, GC×GC with parallel detection by FID and
659 TOF MS featuring tandem ionization: extra-dimensions for great flexibility in fragrance allergens
660 profiling, in: D.W. Armstrong, K.A. Schug (Eds.), 43rd Int. Symp. Capill. Chromatogr. 16th GCxGC

661 Symp., Fort Worth, Texas, USA, 2019: p. 1.
662 [48] C. Cordero, E. Liberto, C. Bicchi, P. Rubiolo, P. Schieberle, S.E. Reichenbach, Q. Tao, Profiling food
663 volatiles by comprehensive two-dimensional ga schromatography coupled with mass
664 spectrometry: Advanced fingerprinting approaches for comparative analysis of the volatile fraction
665 of roasted hazelnuts (Corylus avellana L.) from different ori, J. Chromatogr. A. 1217 (2010) 5848–
666 5858. doi:10.1016/j.chroma.2010.07.006.
667 [49] C. Cordero, C. Bicchi, P. Rubiolo, Group-type and fingerprint analysis of roasted food matrices
668 (coffee and hazelnut samples) by comprehensive two-dimensional gas chromatography, J. Agric.
669 Food Chem. 56 (2008) 7655–7666. doi:10.1021/jf801001z.
670 [50] A. Burdack-freitag, P. Schieberle, Characterization of the Key Odorants in Raw Italian Hazelnuts, J.
671 Agric. Food Chem. 60 (2012) 5057–5064.
672 [51] R.A. van den Berg, H.C.J. Hoefsloot, J.A. Westerhuis, A.K. Smilde, M.J. van der Werf, Centering,
673 scaling, and transformations: Improving the biological information content of metabolomics data,
674 BMC Genomics. 7 (2006) 1–15. doi:10.1186/1471-2164-7-142.
675

Figure Captions

Figure 1: comparative visualization, rendered as the *colorized fuzzy ratio,* between the 12 eV data stream (*analyzed* image) and the 70 eV signal (*reference* image) from a *Piedmont* hazelnut extract. Green highlighted areas indicate peak-regions where the differential response between channels was higher on 12 eV ionization energy.

Figure 2: fragmentation patterns recorded for a selection of analytes at 12 eV (red trace) and 70 eV (blue trace) ionization energy.

Figure 3: heat-map visualization of the *UT* peak-regions normalized response from 70 eV (Fig. 3A) and 12 eV (Fig. 3B) data streams. Analytical replicates were averaged resulting in a (15+15) × 138 data matrix. Hierarchical clustering (HC) is based Euclidean distances - Z-score normalization of the data (i.e., subtract median and divide by standard deviation).

Figure 4: regression analysis between absolute responses recorded at 12 eV (dependent variable) and at 70 eV (independent variable) for aroma precursors: glucose, Orn, Leu and Ile.

Figure 5: heat-map illustrating Pearson correlation matrix results referred as *r* values. Color scale, from blue-to-red, is adapted to emphasize (red color) *r* values ≥ 0.8; HC is based on "one minus Pearson correlation and green squares indicates analytes pairs commented in the text.

Figure 6: regression functions between 3-methylbutanal and Leu (R^2 0.9577), 3-methylbutanal and Ile (R^2 0.9284), 2,3-butanedione and 2,3-pentanedione and the sum of fructose (Fructose 5TMS *syn*- and *anti*-forms) and glucose (Glucopyranose 5TMS and Glucose 5TMS) derivatives (R^2 0.8543 and 0.8860), 2,5-dimethylpyrazine and Ala (R^2 0.8822) and pyrroles (1H-pyrrole, 3-methyl-1H-pyrrole and 1H-pyrrole-2-carboxaldehyde) and the sum of Orn and Ala derivatives (R^2 0.8604). Graph legend (—) linear regression function, (---) confidence interval mean 95%, (—) confidence interval observ. 95%

Table Captions:

Table 1: selected targeted *UT* peak-regions corresponding to informative primary metabolites listed together with their average retention times (^1t_R min, ^2t_R sec), experimental and tabulated (NIST database [21]) I^r, base peak (BP) m/z values recorded from the 2D-apex spectrum and their difference (BP diff), direct match factor (DMF) and reverse match factors (RMF) values calculated between 12 eV *vs.* 70 eV and SNR value between 12 and 70 eV channels (values in bold highlight analytes where SNR at 12 eV was higher compared to 70 eV).