Journal of Chromatography A, 1217 (2010) 565–574

Contents lists available at ScienceDirect

Journal of Chromatography A

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

Profiling analysis of volatile compounds from fruits using comprehensive two-dimensional gas chromatography and image processing techniques

Hans-Georg Schmarr[a,*], Jörg Bernhardt[b]

[a] Dienstleistungszentrum Ländlicher Raum (DLR) Rheinpfalz, Viticulture and Enology Group, Breitenweg 71, D-67435 Neustadt a.d. Weinstraße, Germany
[b] Ernst-Moritz-Arndt-University Greifswald, Institute of Microbiology, Friedrich Ludwig Jahn Straße 15, D-17487 Greifswald, Germany

ARTICLE INFO

Article history:
Received 26 August 2009
Received in revised form
13 November 2009
Accepted 18 November 2009
Available online 24 November 2009

Keywords:
Profiling analysis
HS-SPME-GC × GC-qMS
Comprehensive two-dimensional GC
Contour plots
Image processing
Pattern comparison

ABSTRACT

An image processing approach originating from the proteomics field has been transferred successfully to the processing of data obtained with comprehensive two-dimensional gas chromatographic separations data. The approach described here has proven to be a useful analytical tool for unbiased pattern comparison or profiling analyses, as demonstrated with the differentiation of volatile patterns ("aroma") from fruits such as apples, pears, and quince fruit. These volatile patterns were generated by headspace solid phase microextraction coupled to comprehensive two-dimensional gas chromatography (HS-SPME-GC × GC). The data obtained from GC × GC chromatograms were used as contour plots which were then converted to gray-scale images and analyzed utilizing a workflow derived from 2D gel-based proteomics. Run-to-run variations between GC × GC chromatograms, respectively their contour plots, have been compensated by image warping. The GC × GC images were then merged into a fusion image yielding a defined and project-wide spot (peak) consensus pattern. Within detected spot boundaries of this consensus pattern, relative quantities of the volatiles from each GC × GC image have been calculated, resulting in more than 700 gap free volatile profiles over all samples. These profiles have been used for multivariate statistical analysis and allowed clustering of comparable sample origins and prediction of unknown samples. At present state of development, the advantage of using mass spectrometric detection can only be realized by data processing off-line from the identified software packages. However, such information provides a substantial basis for identification of statistically relevant compounds or for a targeted analysis.

1. Introduction

Since its first instrumental implementation in the early 1990s [1,2], the advantages of comprehensive two-dimensional (2D) chromatographic separations for the analysis of many complex samples have been well demonstrated and applications from various fields have been recently reviewed [3–6]. Improved resolution, higher peak capacities and structured separation space are some of the important benefits obtained with comprehensive 2D separations. The dimensionality of the analytical data compared to classical (1D) analyses (time versus signal intensity) increases dramatically by adding multivariate spectrometric detection, such as mass spectrometry, especially in the case of comprehensive 2D separations. This could be shown *e.g.* for comprehensive 2D gas chromatography coupled to time-of-flight mass spectrometry (GC × GC-TOFMS) [7,8]. The amount of information and structure of such 2D data sets has encouraged many research groups to look into chemometric methods for more efficient treatment of the analytical information obtained after comprehensive 2D separations. Citing only a subset of many articles published, chemometric approaches have been applied to method development [9], peak alignment or warping [10–13], background removal [14,15], deconvolution of overlapping peaks [16–20], and pattern recognition [21–27] for clustering or "profiling" analysis. An overview about the still evolving chemometric techniques applied may be found in reviews published by several authors, with focus on comprehensive 2D separations [28,29]. Despite considerable improvements in analytical instrumentation, computer technology and software development over the last decade, there is still a demand for solutions to process comprehensive 2D separation data as the recent publications from many research groups reflect [27,30–37].

In the work described here, we were exploring commercially available software solutions which could be used for a global as well as for cluster analysis of the data obtained by comprehensive 2D gas chromatographic analysis. Inspired from the field of proteomics, specifically the separation of proteins in two-dimensional gel electrophoresis [38]—which is often abbreviated as "2-D" (discussion on this peculiarity in the editorial of J. Sep. Sci. (2006) 479)—we looked into the data analysis methods commonly applied in this area. Among "omics" technologies, 2D gel electrophoresis is one

* Corresponding author. Tel.: +49 6321 671 289; fax: +49 6321 671 375.
E-mail address: hans-georg.schmarr@dlr.rlp.de (H.-G. Schmarr).

0021-9673/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2009.11.063

of the most prominent examples for decomposing complex mixtures of analytes, proteins in this case, by the application of two orthogonal separations. This may result in patterns of more than 10,000 protein spots using a 2D gel of $30 \times 40\,cm^2$ in size [39], which are recorded in image files with sufficient *x*; *y*-resolution in size and signal resolution in depth. The comparison of patterns originating from different states of cell cultures, tissue samples or microbial fermentation batches provides us with insights about the expression behavior of any expressed protein, and allows a better understanding of physiological phenomena resulting from varying experimental conditions. A series of expression data for a protein is called an expression profile and commonly reflects the accumulated amount of this biomolecule *e.g.* within a defined time course or in different cell types. The methods for data analysis in these experiments face a variety of problems which can be found also when processing contour plot images from GC × GC chromatograms — among them distortions of the separation pattern, spot detection and separation problems, background removal, and more. While 2D gel electrophoresis separates non-volatile protein analytes according to their pI (isoelectric point) and MW (molecular weight) within electric fields, comprehensive 2D gas chromatography (GC × GC) decomposes a complex mixture of volatile analytes according to their retention time in two preferably orthogonal chromatographic separations. The resulting data sets are structurally related to 2D gel separations, with chromatographic peaks as "2D spots" (the latter being gel electrophoresis terminology), and peaks representing a cluster of pixels with larger intensity values than the surrounding background in comprehensive 2D chromatograms [40]. The 2D spots possess a defined *x*-; *y*-coordinate and a signal intensity representing *z*-coordinate, often declared as image or grey level depth, and in GC × GC software packages often presented in color mode for better visualization (potentially applying some data processing prior to visualization [41]). If we assume a sufficient sample to sample reproducibility of GC × GC separations and signal intensities which are proportional to the analyte amounts, volatile profiles over several samples should be extractable. These volatile profiles give us information about the sample content of specific amounts of different volatiles and, if we assume a time course analysis, also a time response. In this way, analytical techniques common in "omics" studies can, at least in part, be applied or transferred to experiments with chromatographic analyses, especially for the purpose of non-targeted profiling analyses of complex mixtures. Besides classical chromatographic analytical approaches, such as target component or group-type analysis, "fingerprinting" analysis becomes more and more relevant, especially for complex mixtures and in combination with multivariate statistical methods (chemometrics) [42].

Since this study will demonstrate how an image processing technique known from 2D gel electrophoresis can be applied for the analysis of GC × GC chromatograms, it is reasonable to first explain some general aspects of the techniques involved. With the ongoing development in computer hardware, powerful image processing techniques have been developed in the last years and are available for desktop computers. Today, image processing is a well established method not only for gel-based proteomics [43], but also for gel-free workflows, such as liquid chromatography coupled with mass spectrometry (LC–MS or LC–MS/MS) [44]. Similarly to gas chromatographic separations (certain aspects of which will be discussed later), although fundamentally different, the 2D gel images present comparable and often highly complex spot patterns after 2D separations of proteins. Such gels or their images suffer from distortion effects, thus shifts in the isoelectric focusing or polyacrylamide SDS electrophoretical separation (PAGE), which then have to be aligned or "warped" prior to further analytical treatment. Applicable correction methods and algorithms derive from astronomical photography [45], and reviews on correction methods have been published recently [46–48]. Since the early beginning of 2D gel electrophoresis, an improved workflow has developed, based on aligning (also referred to as "registering" or "warping") of individual images to the coordinates of a defined reference (gel) image first, then creating a consensus image using a fusion algorithm [49]. On the basis of this fusion image, a consensus spot pattern is generated, and finally, the consensus spot pattern is propagated to all gel images for quantification. The main advantage of this workflow is that a resulting (spot) table is achieved without missing values, the latter being a pre-requisite for statistical analysis. Finally, in 2D gel-based experiments so-called expression profiles can be achieved. For further reading, the present state-of-the-art in this field has been summarized recently [50,51].

In the work described hereafter, a state-of-the-art software package from 2D gel electrophoresis will be evaluated for its application to the analysis of GC × GC data with particular emphasis on the potential for image alignment (or peak alignment), peak recognition and profiling analysis, using statistical methods. The differentiation of patterns obtained from volatile components derived from fruit varieties ("aroma profiling analysis") will be described as an example.

2. Experimental

2.1. Samples

Apples (*Malus domestica* varieties "Braeburn", "Cox-Orange", "Elstar", "Fuji", "Gala", and "Pinova"), pear fruits (*Pyrus communis L.* "Alexander Lucas", "Conference", and "Williams Christ"), as well as a quince fruit (*Cydonia oblonga Mill.*) were obtained from the Pomology Department of the DLR Rheinpfalz, Neustadt, Germany. The ripe fruit samples were stored under normal household conditions and cut into small pieces (except core) prior to analysis.

2.2. Instrumental analysis

A GC × GC system equipped with a programmed temperature vaporizing injector and a double cryo jet modulator (Trace GC × GC Ultra) was coupled to a DSQ quadrupole mass spectrometer (both ThermoFisher Scientific, Dreieich, Germany). Data acquisition and control of the GC × GC-qMS system was achieved with Xcalibur software version 1.4, while HyperChrom software version 2.4.1 was used to process the two-dimensional GC × GC-qMS data (both from ThermoFisher Scientific). Automated headspace solid phase microextraction (HS–SPME) was performed with a TriPlus autosampler including the SPME option (ThermoFisher Scientific).

2.3. Analytical conditions

Analyte extraction was achieved after incubation of 5–7 g of cut fruit pieces for 10 min at 45 °C in a 20 ml headspace vial, crimped with teflon coated silicon rubber septa. A 50/30 μm divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS, Sigma–Aldrich, Steinheim, Germany) fiber coating was selected for extraction. Desorption of the SPME fiber was performed at 230 °C in the vaporizing injector, using a dedicated small bore SPME liner in splitless mode (BGB Analytik, Böckten, Switzerland). A pressure surge of 450 kPa was maintained for the first 3 min and splitless conditions were held for 2 min, after which the split valve was opened to allow a split flow of 15 ml/min. Carrier gas used for chromatographic separation was helium at a constant flow of 1.2 ml/min. The analytical column system consisted of a 30 m × 0.25 mm i.d. fused silica first dimension separation column, coated with 0.25 μm SolGel–Wax (a polyethylene glycol phase) from SGE, Griesheim, Germany and coupled to the second dimension column set, consisting of a 0.15 m × 0.15 mm i.d.

fused silica column, coated in-house with a 0.15 μm film thickness of OV–1701–vi (a 14% cyanopropylphenyl–86% methyl polysiloxane phase from Supelco (Sigma–Aldrich, Steinheim, Germany)); and 2 m × 0.15 mm i.d. of BPX–5 (SGE; a 5% phenyl polydimethylsiloxane phase with 0.25 μm film thickness). Column connections were made via press-fit connectors (BGB). Oven temperature was programmed from 50 °C (5 min isothermal) at 6 °C/min to 235 °C (condition (i)), or at 5 °C/min to 225 °C (condition (ii)). A modulation period of 7 s was used with a time delay of 5 min. Cryo-modulation took place on the OV-1701 second dimension column. Mass spectra were recorded with a time delay of 6 min in the electron impact (EI) positive ion mode applying an electron energy of 70 eV, having the source temperature set to 240 °C and transfer line temperature set to 240 °C. The total ion monitoring (TIC) conditions were *m*/*z* 40–210 (4–25 min), and *m*/*z* 40–250 (final segment) at 18 and 16 Hz data acquisition rates, respectively. With respect to the relatively low data acquisition rate available with our (scanning) quadrupole MS instrument, GC × GC parameters had been set to reduced overall chromatographic performance, *e.g.* with respect to the modulation ratio [52]. However, the applied conditions provided more than sufficient separation efficiency for clustering the aroma compounds, as well as having the mass spectral information available. In-between two analytical runs, a column and fiber conditioning run, programmed from 100 °C at 30 °C/min to 250 °C (10 min isothermal) was applied (split flow at 16 ml/min). Additionally, a SPME fiber with different fiber coating (70 μm carbowax/divinylbenzene (CW/DVB)) was tested (condition (iii)); other conditions for (iii) were as described for (i).

2.4. Conversion of GC × GC-qMS data

Original data from GC × GC-qMS analysis was exported from the HyperChrom software as a character separated value (csv) matrix. Using Microsoft Excel software (version 2004 for Macintosh), comma separators were replaced by tabs and the data were converted to 32-bit gray-scale images. Then *x* and *y* coordinates were scaled by a factor of 4, the image was inverted, rotated by 90°, and saved in a tagged image file format using ImageJ (Macintosh version 1.41o, National Institute of Health, USA). For image processing and data analysis, Delta2D version 4.02 (Decodon, Greifswald, Germany) was used.

2.5. Image analysis workflow

32-bit gray-scale images (Figure S-1) were imported into Delta2D and a project workflow was set-up according to the following scheme.

Replicate separations were organized in analysis groups (Fig. 1a). For best possible image warping, a manual warp graph connecting all images was constructed (Fig. 1b), where each line of the warp graph represents an image transformation bringing the GC × GC images from both ends into congruency. For this purpose,

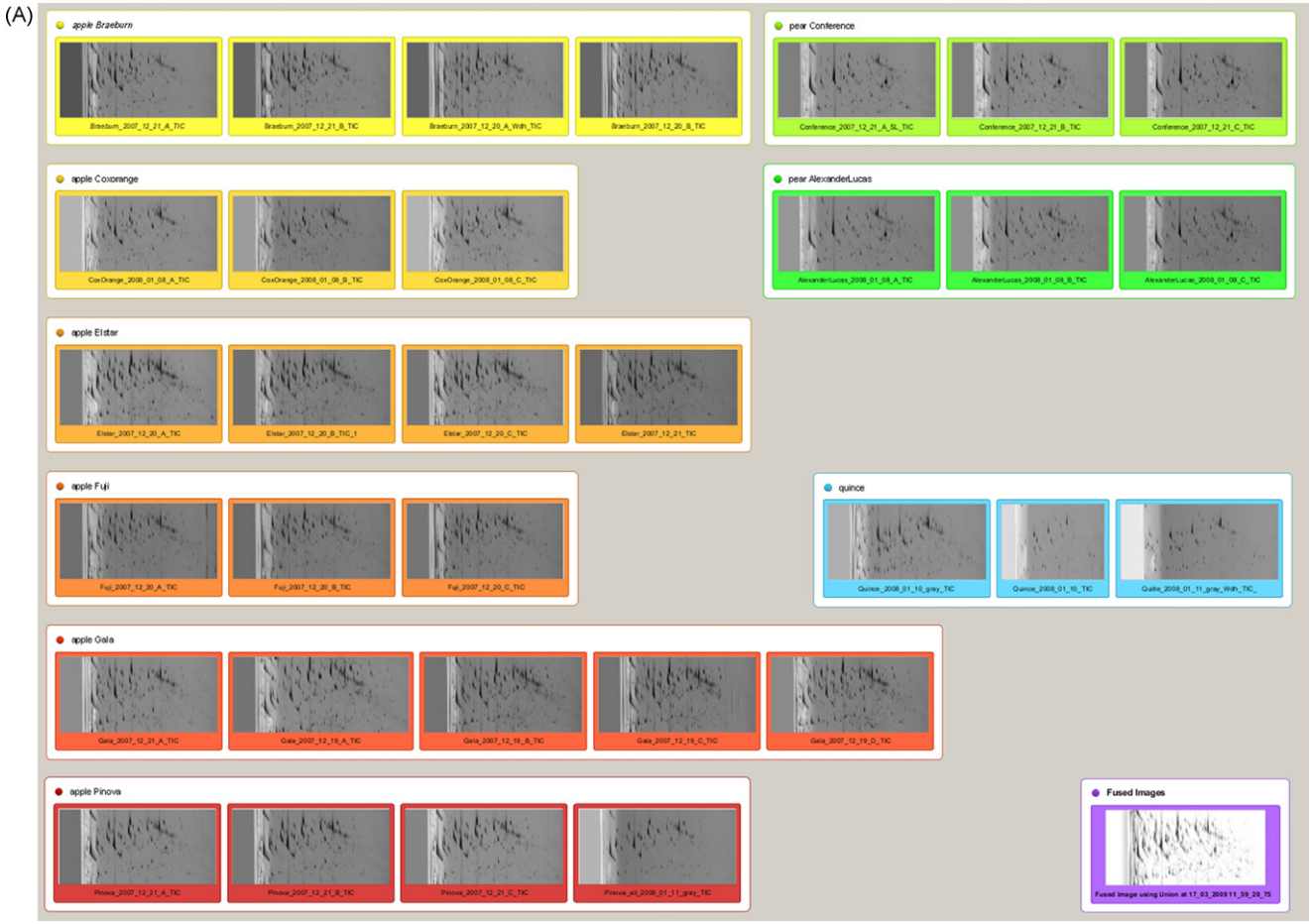

Fig. 1. (a) In Delta2D fruit sample replicates have been organized in replicate groups according to the following color code: reddish and yellowish color shades represent apple samples (yellow – "Braeburn", light orange – "Cox-Orange", orange – "Elstar", orange red – "Fuji", red – "Gala", dark red – "Pinova"), greenish hues indicate pear samples (light green – "Conference", green – "Alexander Lucas") and the blue colored group contains quince fruit samples. The violet group highlights the consensus image that does not contain any background signal by definition. (b) The warp graph clearly highlights that all images have been connected within a warp graph with an apple "Braeburn" sample in the center that simultaneously served as the fusion image (violet) reference.

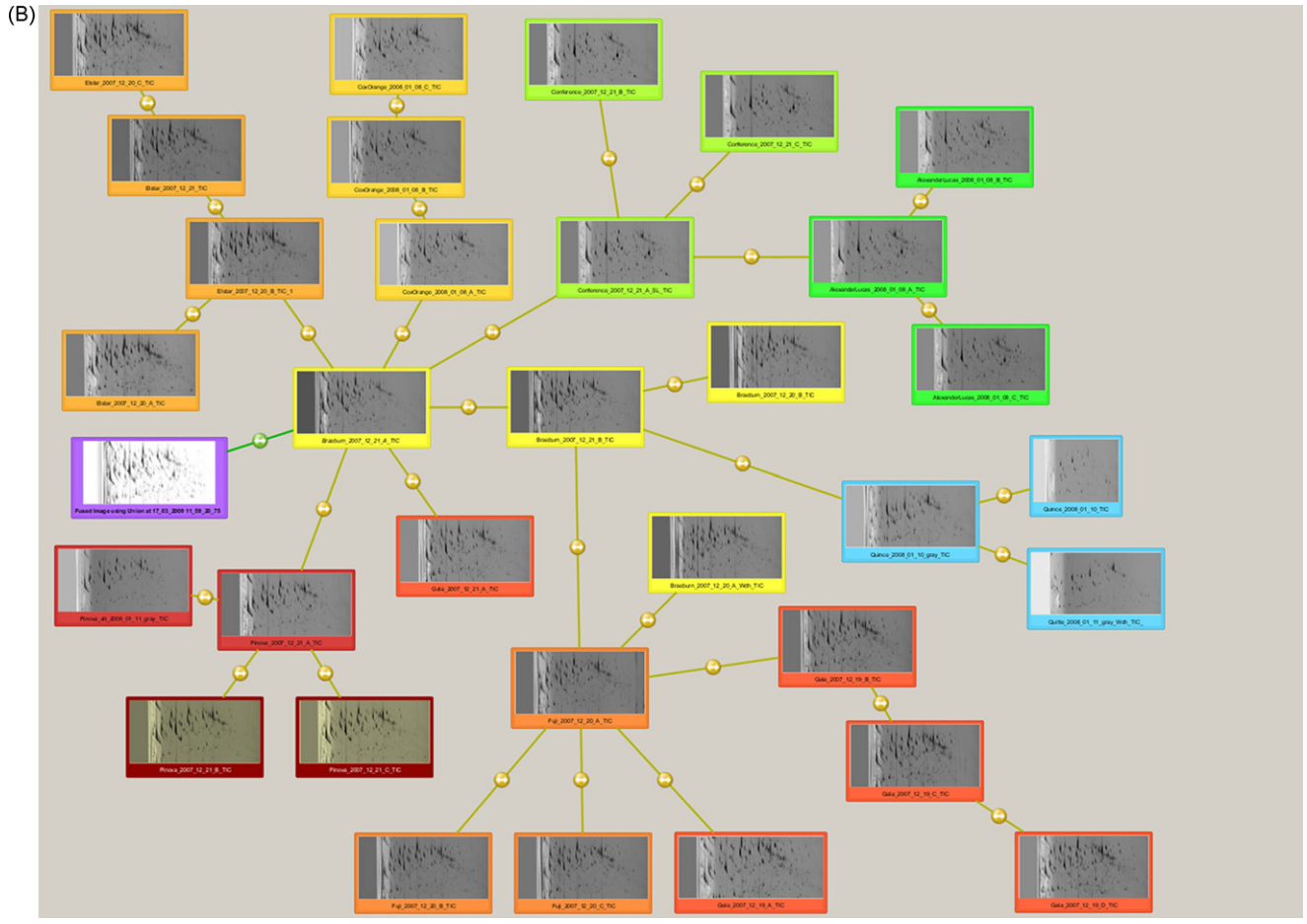

Fig. 1. (*Continued*).

Delta2D automatically searches for similarities within the GC × GC images and generates a set of warp vectors linking corresponding image regions and bringing them onto each other by applying a warp transform (see Fig. 2). Delta2D allows the user to manually optimize these automatically detected warp vectors. The pair wise warp transforms can be combined in a way that also brings more distant images within the warp graph into congruency.

After defining the warp transforms, all GC × GC images are in positional congruency to one of the "Braeburn" images (yellow). By also using the "union fusion" approach described by Luhn et al. [49], rarely occurring spots are preserved over the whole project, and a realistic looking volatile map representing all spots (peaks) appearing in the whole experiment is created. This volatile map serves as basis for the generation of the spot consensus. Because Delta2D's spot detection algorithm has been optimized for 2D gel separated protein spots, rather than for chromatographic peaks, which are often suffering from fronting, tailing or overloading effects, a manual "spot edit marker" based approach has been used instead of the automatic spot detection method. In this respect, the "spot edit marker" has been applied within a spot, allowing automatic detection of every spot boundary within the volatile map (Fig. 3).

After creation of the spot consensus pattern, the spot boundaries are transferred according to the warp transforms to all GC × GC images of the analysis project. This guarantees a reliable quantification because grey level integration occurs within the transferred spot boundaries on the unwarped (original) images. Because signals from inhomogeneous backgrounds interfere with spot quantities on GC × GC images, we applied Delta2D's background removal approach that has been shown useful for 2D gels and that functions similar to the rolling ball method [53,54] as illustrated in Fig. 4.

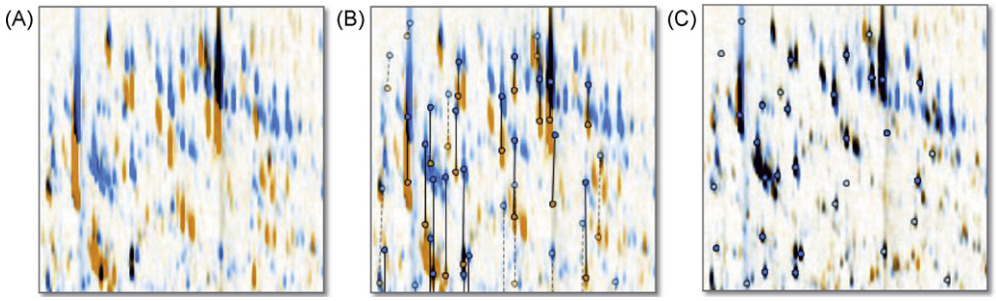

Fig. 2. The warp graph (see Fig. 1 panel B) is composed of $n-1$ (n; number of GC × GC images) pair wise image warp transforms. Here, one pair wise image warping is illustrated in panel A by using a dual channel image (one image is false colored in orange, the second one in blue). Delta2D automatically generates warp vectors for linking corresponding image areas (panel B). The calculated warp transform brings both images into congruency (panel C). Differentially occurring volatiles from the first GC × GC image appear in orange while volatiles with higher quantities on the second GC × GC image appear in blue. Volatiles appearing in comparable amounts are indicated by black.

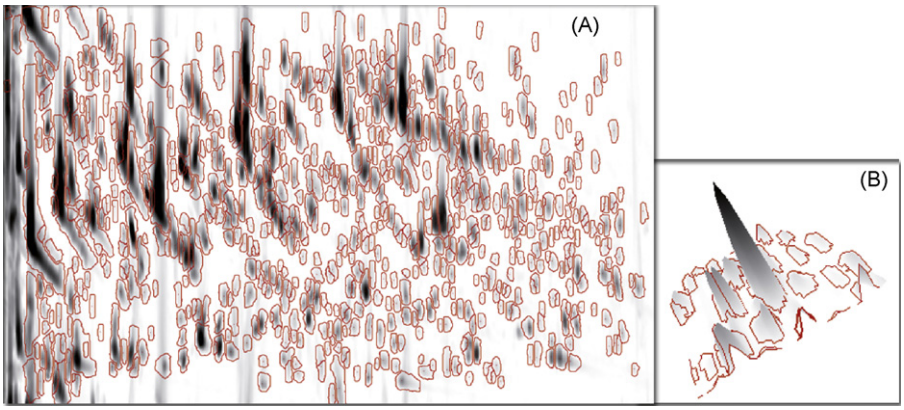

Fig. 3. Image fusion ("union fusion approach") in Delta2D results in a realistically looking GC × GC volatile map which represents all volatiles ever appearing in the GC × GC images which have been analyzed within the same analysis project. The volatile consensus pattern consists of more than 700 detected spots (panel A). For a closer inspection of the spot detection quality an integrated 3D spot viewer is available (panel B).

Besides other artifacts, baseline drifts caused by column bleed are of general concern in data processing of chromatographic separations (especially in our case, as a polar first dimension column has been used), and can be efficiently corrected as shown in this example. In principal, the rolling ball method defines the background quantity which has to be subtracted from the summarized grey levels within any spot boundary.

Because each GC × GC image receives the same consensus pattern of spot boundaries we achieve project-wide sets of quantities for any detected volatile spot. Once the processing is completed, this data set serves as the basis for further statistical analysis.

2.6. Data normalization and statistical analysis

The global detectable amount of volatiles for each sample was set as 100%. The spot quantities are given as part from the whole and follow a distribution illustrated in Fig. 5. For the statistical analysis, different methods which are known from DNA array analysis have been implemented into the software, among them analysis of variance (ANOVA), which can be applied to find significantly changed volatiles over different cultivars and fruit species, principal component analysis (PCA), hierarchical clustering for quality control of sample preparation, and template matching for finding volatiles with a predefined occurrence behavior. Among others, these methods have been implemented by the TIGR Institute of Genome Research within TMEV (http://www.tm4.org) [55], wherefrom a subset is included in the Delta2D software for 2D pattern based data analysis. For a more detailed description of the statistical analyses please see in Figure S-2 and Fig. 7.

3. Results and discussion

3.1. Analysis of aroma compounds

Volatile components from different cultivars of apples, pears and quince fruit were analyzed by headspace solid phase micro

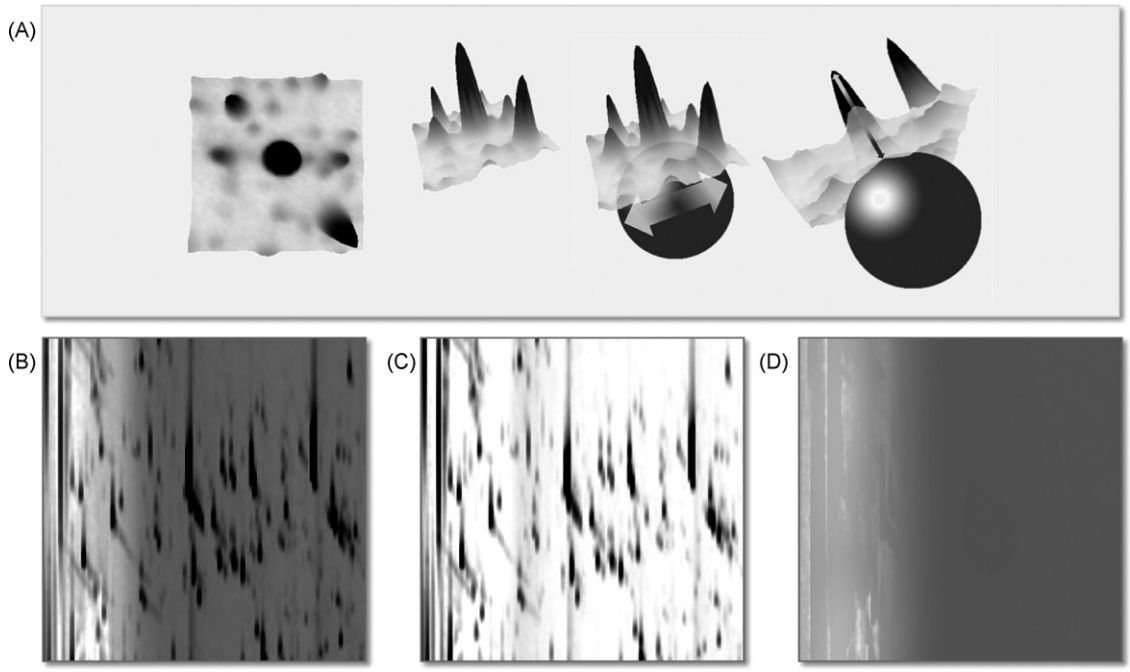

Fig. 4. Delta2D's background removal algorithm is based on the rolling ball method, which is able to reliably recognize and remove artificial signals from inhomogeneous background without truncating spot quantities. Panel A clearly illustrates how this works with 2D gel-based protein images. The same algorithm with an adapted diameter (3-fold of the largest spot) has been applied to a GC × GC image (panel B) and resulted in a spot component (panel C) and a component which corresponds to the background model (panel D).

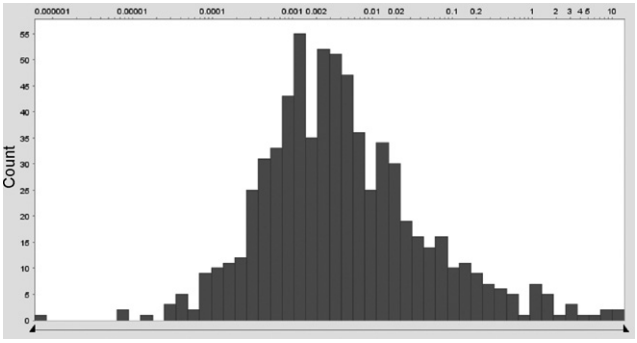

Fig. 5. The quantity frequency plot of detected volatile spots clearly illustrates that the majority of spots have low or very low quantities while the high quantity spots appear only in a small amount. Note, that the *x*-axis (relative quantity) on top is log scaled.

extraction (HS-SPME) coupled to comprehensive gas chromatography and quadrupole mass spectrometry (GC × GC-qMS). The GC × GC analysis not only allowed an extensive analysis of the chemical composition of the volatiles, but also provided the basis for a rapid comparative analytical approach of the resulting pattern of the volatile components. Reliable generation of volatile component patterns presupposes standardized and reproducible analytical conditions. With respect to the analysis of volatile components, HS-SPME is a commonly applied extraction method today, involving only minimum manual sample preparation steps [56]. However, repeatability is often critical as extraction equilibria conditions of analytes between the different phases involved greatly influence the analyte recovery. In our case, repeatable extraction conditions were best realized using an automated SPME device, minimizing sample to sample variation for extracted compounds or generation of potential artifacts due to sample preparation. The chromatographic conditions for GC × GC-qMS analysis were optimized for the desired separation, using the combination of coupling a polar to an apolar stationary phase (1st dimension column to 2nd dimension column), as is often described for aroma analysis in literature. We also considered the consumption of liquid carbon dioxide in the modulation system used in our instrument with respect to the number of analyses, and furthermore, the setting of the scan parameters for the quadrupole mass spectrometer. The conditions usually applied (*i*), involved a relatively fast oven temperature program rate (6 °C/min), a modulation period of 7 s, and a reduced overall duration of modulation time per run of about 30 min. The resulting 2D chromatograms still provided the necessary resolution and good utilization of the separation space as well as acceptable mass spectra for further identification purposes. A more detailed discussion on these aspects has been published recently [57].

Examples for the analysis of volatile substances from the three fruit varieties and their resulting patterns are shown with their contour plots in Figure S-1. Fruit aroma, such as apple or pear, are among the most investigated flavors due to their economic importance. The composition of the volatile substances is primarily dominated by a large amount of homologous esters, such as ethyl butanoate, ethyl 2-methylbutanoate, 2-methylbutyl acetate, butyl acetate, hexyl acetate, or alcohols, such as butanol, 2-methylbutanol, hexanol, or (*E*)-2-hexenol. Furthermore, carbonyls such as hexanal, (*Z*)-3-hexenal, (*E*)-2-octenal, (*E/Z*)-2-nonenal, 1-octen-3-one, and others, like (*E*)-β-damascenone, limonene, farnesene, linalool, eugenol, or butyric acid have been described as main aroma compounds [58]. As can be seen in the chromatograms presented in Figure S-2, the structured separation space available in this "reversed" column combination set-up is sufficiently utilized and shows the high sample dimensionality [59] of such fruit aromas, which also cover wide concentration ranges over more than 5 orders of magnitude (see Fig. 5) from several milligram to single nanogram per kilogram amounts. The divinylbenzene/carboxen/polydimethylsiloxane SPME fiber was used for the experiments, as this phase is well known for its ability to extract both polar and apolar substances. A first visual inspection of the GC × GC data (see Figure S-1) shows closer similarity between apples and pears, but differences with quince fruit.

3.2. Data analysis with image processing software

During application of the Delta2D software package we found that the proven 2D gel workflow approach is also suitable for an analysis project of 2D GC × GC chromatograms (Fig. 6) and allows a reliable processing in adequate time. In particular, the workflow can be summarized as follows. To start, gray-scale images of the individual chromatographic experiments had to be created by exporting text files from the chromatographic software with conversion to images in tiff file format, a step which is equivalent with scanning gels in a gel-based application. This step took about 1 h hands-on time for all experiments here. Then, setting up the Delta2D analysis project took about half an hour, including the 2D GC × GC image import (32 images) and classification into 9 replicate groups (6 sorts of apples with 3–6 replicates, 2 sorts of pears with 3 replicates and 1 sort of quince fruit with 3 replicates). For image registration, a manually determined warp graph (warp strategy) has been applied (Fig. 1b). As a result, 31 warp transforms have been built automatically by Delta2D, then manual correction and improvements of Delta2D generated warp vectors took about 2.5 h hands-on time. The resulting warp transforms of Delta2D merged all 2D GC × GC images into a project-wide 2D GC × GC volatile map by using the "union fusion approach". Because the automatic spot detection algorithm of Delta2D has been optimized for 2D gels, all spots from the volatile map have been detected manually by using so called "spot edit markers". This took about 1 h hands-on time and resulted in more than 700 volatile spots (see Fig. 3). The application of the spot consensus pattern to each 2D GC × GC image and quantitation of the spots was initiated manually and automatically processed within 5 min. Because all achieved quantities rely on spot boundaries of a project-wide spot consensus pattern, any detected spot appears as a complete volatile profile within the extracted quantitation table. At this point for 32 samples, before we started statistical data analysis, we spent about 5 h of hands-on time for processing and data extraction at the computer (this is some 10 min per sample).

3.3. Statistical analysis of sample data by using the Delta2D integrated TMEV module:

Hierarchical clustering (Figure S-2) is a fast and powerful unsupervised clustering approach useful for rapidly getting an overview of the data structure in these experiments. On the basis of standardized intensity values samples clearly cluster according to their species (apples, pears and quince cluster distant from each other) and in some cases also according to their cultivars. ("Alexander Lucas" and "Conference" distinctly cluster among pears. "Braeburn" and "Fuji" do not separate but "Elstar", "Gala", "Pinova" can be distinguished among apples. Why one "Elstar" and one "Gala" sample do not cluster according the other "Elstar" and "Gala" samples we cannot explain, but believe that ripening state may play a certain role in the release of volatile compounds.) Spots cluster according their expression behaviour. Rows of cluster clearly highlight substances which are typical for cultivars or species.

The PCA plot obtained for the comparative study of the fruit volatiles pattern (Fig. 7) shows a clear interspecies, and to a certain extent also an interracial, sample clustering. The apple samples are represented in the middle and "Fuji" and "Elstar" seem to be

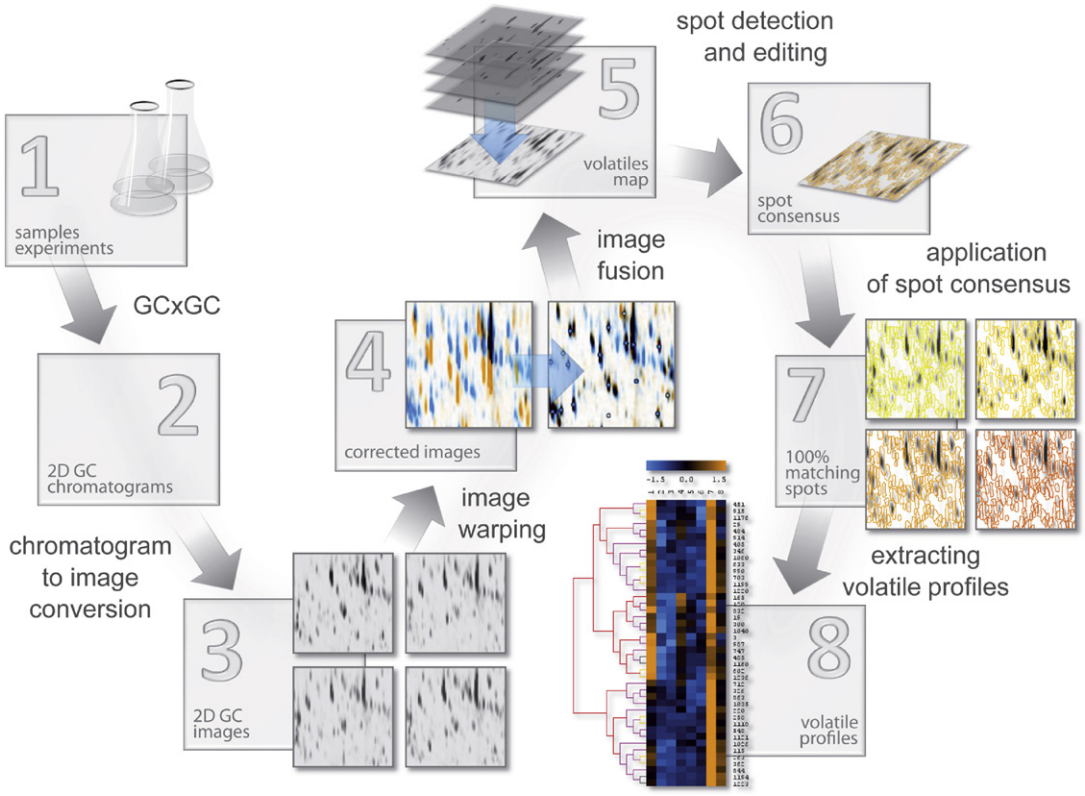

Fig. 6. Workflow of the described method: (1) Samples have been prepared and analyzed by HS-SPME-GC × GC-qMS; (2) 2D GC chromatograms have been transformed into 32-bit images; (3) 2D GC images were stored in Delta2D; (4) Positional correction (warp vectors) resulted in image congruency (dual channel overlay color code: blue = image1, orange = image2 and black = overlap); (5) Volatiles map as a result of project-wide 2D GC image fusion; (6) Detected spot consensus; (7) Spot consensus boundaries were applied to all 2D GC images for gray level integration; (8) Gray level integration results in quantitative data which can be summarized in volatile profiles (blue – low amount, black – average amount, orange – large amount of volatile).

close relatives, followed by "Pinova". The other apples are somewhat more different in their aroma profile, especially "Cox-Orange", which clusters in the upper area. In the family tree, no ancestors for "Cox-Orange" could be found, however it is crossbred into "Braeburn". "Gala" is crossbred from "Golden Delicious" and "Kidds Orange" (the latter containing "Cox-Orange" and "Golden Delicious") – in principle, about 75% "Golden Delicious" and 25% "Cox-Orange". The relationships within the species cannot be sufficiently differentiated by the profiles of their volatile compounds. Interspecies relations are well depicted. Pears are closely related to apples, and quince is further distant from the others. We also analyzed one sample of "Williams Christ" (chromatogram C in Figure S-1) as an unknown sample, which clustered well within the group of pear fruits (data not shown), indicating the potential for identification of unknowns.

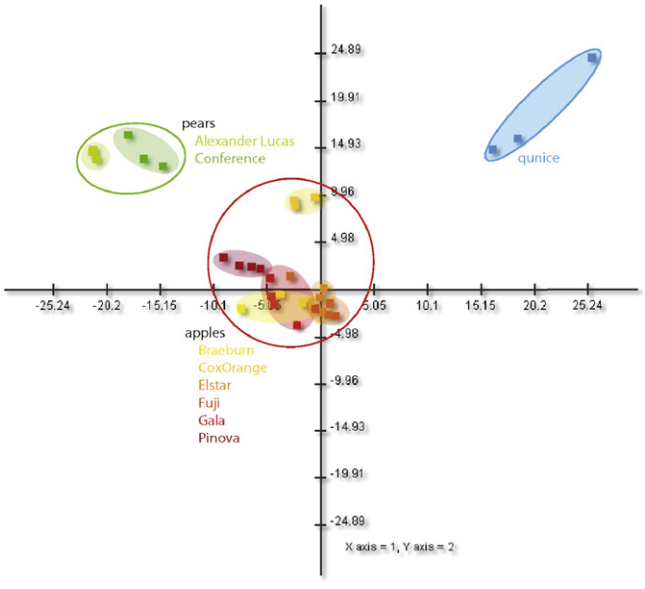

Fig. 7. PCA analysis: In the first/second principal component plot (panel A), except for "Cox-Orange" and with much lower distance "Pinova", all apples (reddish and yellowish color shades) are projected into the center. Pears, which are encoded by green color hue, appear on the upper left, while "Alexander Lucas" and "Conference" are clearly distinguishable. The group of quince fruit samples appears at largest distance to the other samples on the upper right.

3.4. Identification of individual aroma compounds

The Delta2D software allows labelling of those spots which are statistically relevant for differentiation (Fig. 8). At present, in our case, access to the mass spectrometric data for further spot (compound) identification is only possible off-line in the original software from the instrument manufacturer (HyperChrom or Xcalibur software, respectively). This is done via pattern comparison or retention time, respectively. Delta2D allows an export of spot labels in table format which can be further used for identification purposes via the retention time information. As an example, the identification of the spot labelled by the Delta2D software as "quince 12" (Fig. 8C) is presented. Within HyperChrom software, this spot could be identified via its mass spectrometric and retention time (index) information as the compound "marmelo oxide", which is known to be a characteristic substance in quince fruit aroma [60,61]. Furthermore, using Xcalibur or HyperChrom software, the available mass spectrometric data can also be used directly to identify (target) compounds or even compound classes,

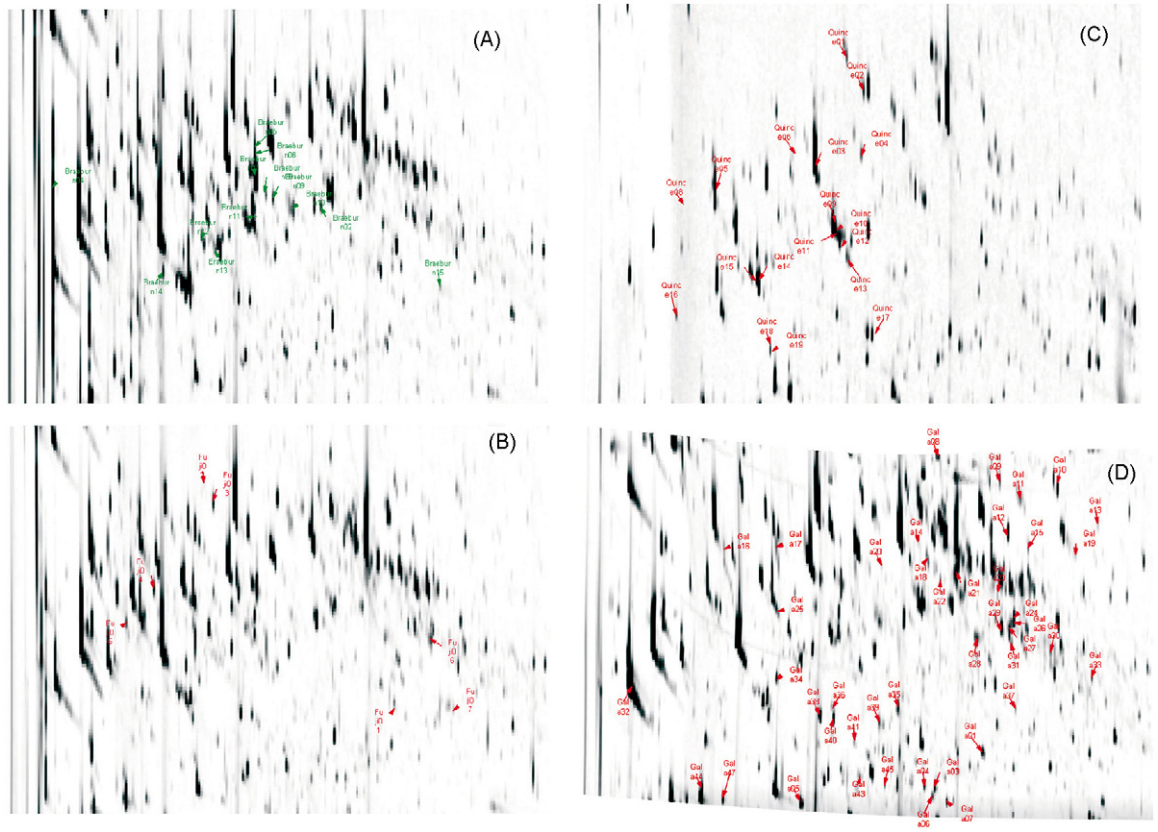

Fig. 8. Spot (spot) maps of HS-SPME-GC × GC-qMS analysis of fruits A: apple ("Braeburn"), condition (i); B: apple ("Fuji)", condition (i); C: quince fruit, condition (iii), and D: apple ("Gala"), condition (ii). All spots indicated and labeled with number represent statistically relevant components.

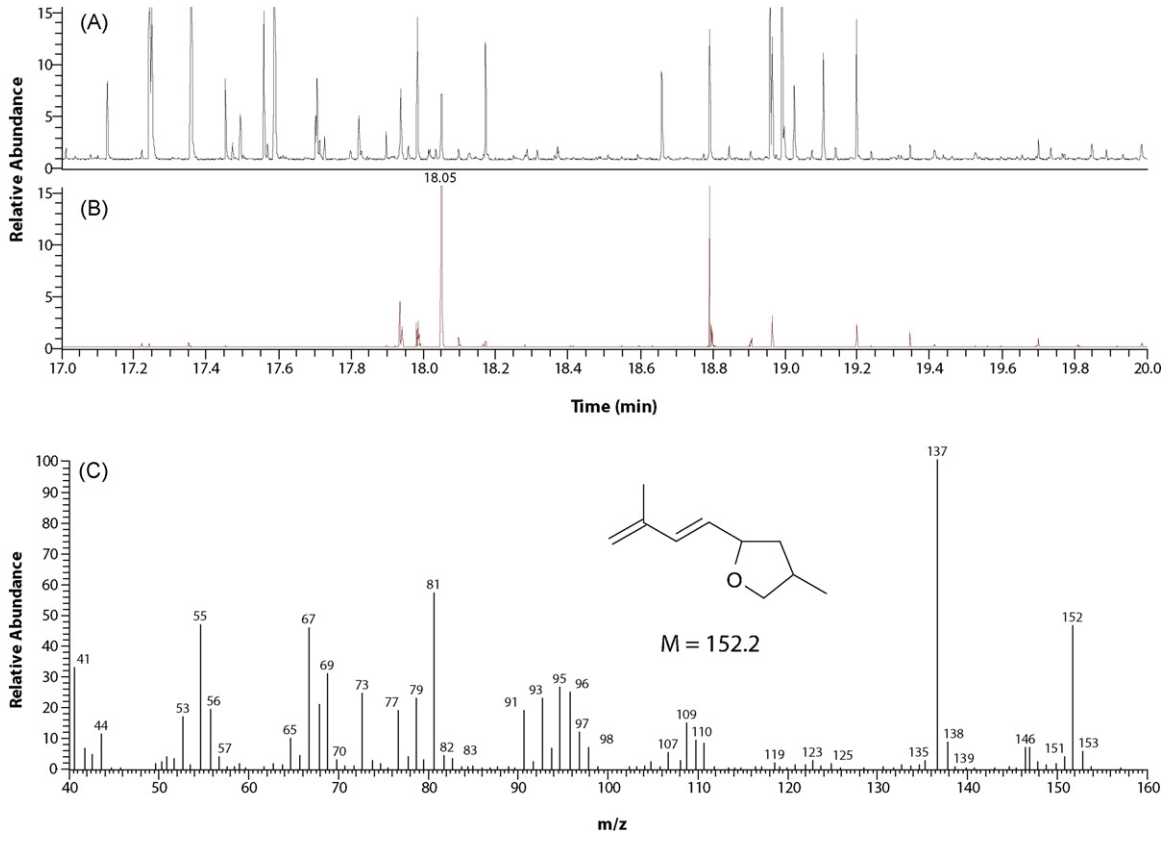

Fig. 9. Identification of compound marmelo oxide with Xcalibur software: A: Modulated chromatogram (TIC) of quince fruit condition (i); B: extracted ion chromatogram of *m/z* 137 (base peak of marmelo oxide); and C: mass spectrum of compound eluting at 18.05 min. showing identical spectrum to the one described for marmelo oxide by Tsuneya et al. [61].

using extracted ion chromatograms as described *e.g.* for group-type analysis [62]. This way, the selective detection of the marmelo oxide can also be achieved by an extracted ion chromatogram of *m*/*z* 137 or 152, characteristic ions in the spectra of marmelo oxides (Fig. 9). Another example of the selective detection of known target fruit aroma compounds could be (*E,Z*)- and (*E,E*)-ethyl 2,4-decadienoate, which are characteristic esters in pear aroma, having fragment ion of *m*/*z* 151 in common (data not shown).

3.5. Robustness of peak warping and statistical analysis

The robustness of both peak warping and clustering algorithms can be seen with results obtained for GC × GC-qMS chromatograms, applying different analytical conditions. In Fig. 8, two chromatograms (apples; A and D) were obtained with somewhat varying oven temperature program rates. Although total run times and absolute retention times were considerably different between these two chromatograms, peak warping allowed reliable comparison of the peak patterns and correct clustering of these samples. The different run conditions also explain the "distorted" image of Fig. 8D, as a more drastic warp transform had to be applied. Another example is shown in Fig. 8C, as this chromatogram was generated using somewhat altered extraction conditions, using a different SPME fiber. In this case, different (absolute) extraction values for quince fruit aroma compounds due to the different chemistry of the SPME fibre have to be considered, but still, this chromatogram could be found within the quince cluster. Surely, the latter is favored due to the more distinct nature of quince fruit, with its own pattern of aroma compounds. But it seems to be noteworthy mentioning these results, as different retention times due to different run conditions could otherwise only be compensated by applying retention index systems, which is somewhat challenging in GC × GC applications [63,64].

4. Conclusions

In approaches to pattern recognition analysis, chromatographers often rely on peak based studies, which means in a first step, peaks have to be aligned, background subtracted, then integrated, and the resulting peak data further processed by statistical means such as ANOVA, PCA and others. Since the quality of results obtained in this way relies to a strong extent on the quality of both peak integration and peak purity (also considering algorithms used for aligning, background subtraction, and eventually peak deconvolution), additional work is often necessary and a qualified operator supervising the integration-based results is indispensable. Compared to background subtraction methods normally applied in chromatography, and particularly in GC × GC [14], the rolling ball approach used in the Delta2D software is a promising alternative for subtracting background noise from comprehensive 2D chromatographic data, as it considers the 3-dimensional "landscape" of the image. The set of methods incorporated for image processing in the Delta2D software package use sophisticated algorithms for image warping and spot detection, but should also be supervised by an experienced operator. The total workflow can be semi-automated and minimized in terms of hands-on time. Available methods for clustering and pattern comparison algorithms are well known from *e.g.* proteomics or DNA array analysis and could successfully be applied to GC × GC data when working on the contour plots. The described method guarantees that the total chromatographic information from all experiments is taken into consideration for (statistical) analysis. Compared to multitarget profiling — which can be seen as an extension of classic analytical chemistry by defining *e.g.* compound spectra, search retention index windows, quantification ions and spectra-similarity thresholds for multiple (known) metabolites — working with the total available information is an important issue and provides a robust approach for unbiased (non-targeted) profiling studies [65]. Unbiased profiling analyses require powerful data processing tools capable of detecting unidentified and potentially novel biomarkers, which drives an increasing interest in the field of metabolic fingerprinting [66,67] and visionary ideas, such as "chromatonomics" [68,69] where external calibration of the chromatographic information (profile) with sample information is targeted (an idea originating from *e.g.* FT-IR analysis [70]). Approaches based on comparative visualization, similar to the one described in this work or by others *e.g.* Hollingsworth et al. [37], can therefore be regarded as essential and promising tools when using GC × GC datasets for unbiased profiling analysis. In this respect, another promising step in this direction has recently been published by Gröger et al. [33] as "pixel-based analysis", using pixel-based chemometric processing methods.

As a final observation, the software evaluated here provides sophisticated tools to process 2D chromatographic data for peak aligning (warping), peak detection and (to some extent) peak quantification. The workflow from 2D gel electrophoresis analysis, which is common in proteomics studies, could be successfully transferred to a profiling analysis based on GC × GC data and the incorporated features for statistical analysis could be used for pattern recognition approaches. Besides metabolomics, unbiased pattern comparison techniques can help to explain complex questions or correlations, such as differences amongst cultivars, geographic origins, or technology treatments, without the need of finding specific target compounds. Furthermore, when mass spectrometry has been applied for detection, as in this study, the individual structural information is also available. So, with appropriate quantification methods, the classical target component analysis is possible, which provides fundamental information that is still pre-requisite for understanding analytic results on a molecular basis.

At present, however, the third-order-advantage [71,72] available with spectroscopic detectors is not available per se in Delta2D, due to its origin from 2D gel electrophoresis. Compared to other software packages for GC × GC data processing, such as ChromaTOF (LECO; St. Joseph, MI, USA) or GCImage (GCImage, LLC; Lincoln, NE, USA), this is a clear disadvantage. Although this data can still be accessed off-line and then used for peak identification, *e.g.* of statistically relevant substances, a future milestone in the development of the software applied in this study should be the incorporation of this feature. Then, aligning and deconvolution techniques using algorithms such as PARAFAC [18,29,71,73,74] or PARAFAC2 [75,76] could be incorporated as well. Still, using univariate detector information, as obtained with traditional flame ionization detection (FID) or with the MS information used as an "image" in our example, the features already available within Delta2D allow pattern recognition (profiling) analyses within reasonable hands-on time for data processing, using the workflow described here.

With increasing demand from metabolic and other "omic" areas, fingerprinting analysis using comprehensive 2D techniques will become more routine and will call for appropriate tools in data treatment. At the present time, important efforts put into further development of algorithms and software in this field can be derived from the many recent publications by other groups like Gröger et al. [33,77], Skov and co-workers [30,76], Synovec and co-workers [29,74,78,79], Hyoetylaeinen and co-workers [32], Vial et al. [27], and others. In this respect, we expect auspicious developments for the data treatment of comprehensive 2D chromatographic separations. and eventually hope to see the incorporation of some accomplishments from the related proteomics field.

Acknowledgements

The authors are grateful to Dr. Daniela Cavagnino from ThermoFisher Scientific (Italy) for helpful discussions and support as well as Matthias Berth and Frank Michael Moser from Decodon (Germany) for fine-tuning some parts within Delta2D software, especially for the modified spot detection algorithm. JB thanks Prof. Michael Hecker (University Greifswald, Germany) for the opportunity to follow this interesting topic off-site his daily research business. HGS thanks Prof. Ulrich Fischer (DLR Rheinpfalz, Germany) for his faith in this work and the great amount of time spent on this project. The authors are particularly thankful to Dr. Walter Weiss from the proteomics group of Prof. Angelika Görg (TU München, Germany) for initiating this personal co-operation between a chromatographic and proteomics research field. We are also indebted to Dr. Alan C. Heyvaert (DRI, Reno, NV, USA) for proofreading of the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.chroma.2009.11.063.

References

[1] Z. Liu, J.B. Phillips, J. Chromatogr. Sci 29 (1991) 227.
[2] M.M. Bushey, J.W. Jorgenson, Anal. Chem. 62 (1990) 161.
[3] H.J. Cortes, B. Winniford, J. Luong, M. Pursch, J. Sep. Sci. 32 (2009) 883.
[4] M. Adahchour, J. Beens, U.A.T. Brinkman, J. Chromatogr. A 1186 (2008) 67.
[5] P. Dugo, F. Cacciola, T. Kumm, G. Dugo, L. Mondello, J. Chromatogr. A 1184 (2008) 353.
[6] J. Dallüge, J. Beens, U.A.T. Brinkman, J. Chromatogr. A 1000 (2003) 69.
[7] R. Shellie, P. Marriott, P. Morrison, Anal. Chem. 73 (2001) 1336.
[8] J. Dallüge, R.J.J. Vreuls, J. Beens, U.A.T. Brinkman, J. Sep. Sci. 25 (2002) 201.
[9] C. Vendeuvre, F. Bertoncini, D. Thiebaut, M. Martin, M.-C. Hennion, J. Sep. Sci. 28 (2005) 1129.
[10] C.G. Fraga, B.J. Prazen, R.E. Synovec, Anal. Chem. 73 (2001) 5833.
[11] C.G. Fraga, B.J. Prazen, R.E. Synovec, Anal. Chem. 72 (2000) 4154.
[12] D. Zhang, X. Huang, F.E. Regnier, M. Zhang, Anal. Chem. 80 (2008) 2664.
[13] T. Skov, F. van den Berg, G. Tomasi, R. Bro, J. Chemom. 20 (2007) 484.
[14] S.E. Reichenbach, M. Ni, D. Zhang, E.B. Ledford, J. Chromatogr. A 985 (2003) 47.
[15] Y. Zhang, H.-L. Wu, A.L. Xia, L.-H. Hu, H.-F. Zou, R.-Q. Yu, J. Chromatogr. A 1167 (2007) 178.
[16] A.E. Sinha, J.L. Hope, B.J. Prazen, E.J. Nilsson, R.M. Jack, R.E. Synovec, J. Chromatogr. A 1058 (2004) 209.
[17] G. Vivo-Truyols, J.R. Torres-Lapasio, M.C. Garcia-Alvarez-Coque, P.J. Schoenmakers, J. Chromatogr. A 1158 (2007) 258.
[18] R.E. Mohler, K.M. Dombek, J.C. Hoggard, K.M. Pierce, E.T. Young, R.E. Synovec, Analyst 132 (2007) 756.
[19] J.C. Hoggard, R.E. Synovec, Anal. Chem. 79 (2007) 1611.
[20] B.J. Prazen, R.E. Synovec, B.R. Kowalski, Anal. Chem. 70 (1998) 218.
[21] B.R. Kowalski, C.F. Bender, J. Am. Chem. Soc. 94 (1972) 5632.
[22] K.J. Johnson, R.E. Synovec, Chemom. Intell. Lab. Syst. 60 (2002) 225.
[23] A.G. González, J. Chromatogr. A 1158 (2007) 215.
[24] L.A. Berrueta, R.M. Alonso-Salces, K. Héberger, J. Chromatogr. A 1158 (2007) 196.
[25] P. Jonsson, A.I. Johansson, J. Gullberg, J. Trygg, J.A.B. Grung, S. Marklund, M. Sjöström, H. Antti, T. Moritz, Anal. Chem. 77 (2005) 5635.
[26] L.T. Vaz-Freire, M.D.R. Gomes da Silva, A.M.C. Freitas, Anal. Chim. Acta 633 (2009) 263.
[27] J. Vial, H. Nocairi, P. Sassiat, S. Mallipatu, G. Cognon, D. Thiébaut, B. Teillet, D.N. Rutledge, J. Chromatogr. A 1216 (2009) 2866.
[28] O. Amador-Munoz, P.J. Marriott, J. Chromatogr. A 1184 (2008) 323.
[29] K.M. Pierce, J.C. Hoggard, R.E. Mohler, R.E. Synovec, J. Chromatogr. A 1184 (2008) 341.
[30] T. Skov, J.C. Hoggard, R. Bro, R.E. Synovec, J. Chromatogr. A 1216 (2009) 4020.
[31] T. Skov, R. Bro, Anal. Bioanal. Chem. 390 (2008) 281.
[32] M. Kallio, M. Kivilompolo, S. Varjo, M. Jussila, T. Hyoetylaeinen, J. Chromatogr. A 1216 (2009) 2923.
[33] T. Gröger, M. Schäffer, M. Pütz, B. Ahrens, K. Drew, M. Eschner, R. Zimmermann, J. Chromatogr. A 1200 (2008) 8.
[34] S. Peters, E. van Velzen, H.-G. Janssen, Anal. Bioanal. Chem. 394 (2009) 1273.
[35] S. Peters, G. Vivo-Truyols, P.J. Marriott, P.J. Schoenmakers, J. Chromatogr. A 1156 (2007) 14.
[36] S.E. Reichenbach, M. Ni, V. Kottapalli, A. Visvanathan, Chemom. Intell. Lab. Syst. 71 (2004) 107.
[37] B.V. Hollingsworth, S.E. Reichenbach, Q. Tao, A. Visvanathan, J. Chromatogr. A 1105 (2006) 51.
[38] P.H. O'Farrell, J. Biol. Chem. 250 (1975) 4007.
[39] J. Klose, U. Kobalz, Electrophoresis 16 (1995) 1034.
[40] S.E. Reichenbach, V. Kottapalli, M. Ni, A. Visvanathan, J. Chromatogr. A 1071 (2005) 263.
[41] A. Visvanathan, S.E. Reichenbach, Q. Tao, J. Electron. Imaging 16 (2007) 033004.
[42] V.G. van Mispelaar, H.-G. Janssen, A.C. Tas, P.J. Schoenmakers, J. Chromatogr. A 1071 (2005) 229.
[43] M. Hecker, H. Antelmann, K. Büttner, J. Bernhardt, Proteomics 8 (2008) 4958.
[44] L. Linsen, J. Locherbach, M. Berth, D. Becher, J. Bernhardt, IEEE T. Vis. Comput. Gr. 12 (2006) 497.
[45] H.F. Grahn, P. Geladi, Techniques and Applications of Hyperspectral Image Analysis, John Wiley & Sons Ltd, Chichester, UK, 2007.
[46] T. Aittokallio, J. Salmi, T.A. Nyman, O.S. Nevalainen, J. Chromatogr. B 815 (2005) 25.
[47] A.W. Dowsey, M.J. Dunn, G.-Z. Yang, Proteomics 3 (2003) 1567.
[48] Z. Smilansky, Electrophoresis 22 (2001) 1616.
[49] S. Luhn, M. Berth, M. Hecker, J. Bernhardt, Proteomics 3 (2003) 1117.
[50] J.E. Bandow, J.D. Baker, M. Berth, C. Painter, O.J. Sepulveda, K.A. Clark, I. Kilty, R.A. VanBogelen, Proteomics 8 (2008) 3030.
[51] M. Berth, F.M. Moser, M. Kolbe, J. Bernhardt, Appl. Microbiol. Biotechnol. 76 (2007) 1223.
[52] W. Khummueng, J. Harynuk, P.J. Marriott, Anal. Chem. 78 (2006) 4578.
[53] S.R. Sternberg, Comput. Vision Graph 35 (1986) 333.
[54] S.R. Sternberg, Computer 16 (1983) 22.
[55] A.I. Saeed, V. Sharov, J. White, J. Li, W. Liang, N. Bhagabati, J. Braisted, M. Klapa, T. Currier, M. Thiagarajan, A. Sturn, M. Snuffin, A. Rezantsev, D. Popov, A. Ryltsov, E. Kostukovich, I. Borisovsky, Z. Liu, A. Vinsavich, V. Trush, J. Quackenbush, BioTechniques 34 (2003) 374.
[56] H. Kataoka, H.L. Lord, J. Pawliszyn, J. Chromatogr. A 880 (2000) 35.
[57] H.-G. Schmarr, W. Sang, S. Ganß, U. Fischer, B. Köpp, C. Schulz, T. Potouridis, J. Sep. Sci. 31 (2008) 3458.
[58] E. Fuhrmann, W. Grosch, Nahrung 46 (2002) 187.
[59] J.C. Giddings, J. Chromatogr. A 703 (1995) 3.
[60] Y. Nishida, Y. Fukushima, H. Ohrui, H. Meguro, Agric. Biol. Chem. 48 (1984) 1217.
[61] T. Tsuneya, M. Ishihara, H. Shiota, M. Shiga, Agric. Biol. Chem. 47 (1983) 2495.
[62] N. Ochiai, T. Ieda, K. Sasamoto, A. Fushimi, S. Hasegawa, K. Tanabe, S. Kobayashi, J. Chromatogr. A 1150 (2007) 13.
[63] S. Bieri, P.J. Marriott, Anal. Chem. 80 (2008) 760.
[64] S. Bieri, P.J. Marriott, Anal. Chem. 78 (2006) 8089.
[65] O. Fiehn, TrAC Trend. Anal. Chem. 27 (2008) 261.
[66] M.F. Almstetter, I.J. Appel, M.A. Gruber, C. Lottaz, B. Timischl, R. Spang, K. Dettmer, P.J. Oefner, Anal. Chem. 81 (2009) 5731.
[67] R.A. Shellie, W. Welthagen, J. Zrostlikova, J. Spranger, M. Ristow, O. Fiehn, R. Zimmermann, J. Chromatogr. A 1086 (2005) 83.
[68] H.-G. Schmarr, Proceedings of the 3rd "Anwendertreffen Weinanalytik", Neustadt an der Weinstraße, Germany, 2005, p.42.
[69] H.-G. Schmarr, J. Bernhardt, Proceedings of the 32nd International Symposium on Capillary Chromatography and 5th GC × GC Symposium, Riva del Garda, Italy, 2008; T. Sandra, P. Sandra (Eds.), I.O.P.M.S. vzw, Kennedypark 26, Kortrijk, Belgium, p. 97.
[70] C.-D. Patz, A. Blieke, R. Ristow, H. Dietrich, Anal. Chim. Acta 513 (2004) 81.
[71] D. Bylund, R. Danielsson, G. Malmquist, K.E. Markides, J. Chromatogr. A 961 (2002) 237.
[72] B.J. Prazen, C.A. Bruckner, R.E. Synovec, B.R. Kowalski, J. Microcol. Sep. 11 (1999) 97.
[73] R.E. Mohler, B.P. Tu, K.M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, J. Chromatogr. A 1186 (2008) 401.
[74] J.C. Hoggard, R.E. Synovec, Anal. Chem. 80 (2008) 6677.
[75] T. Skov, Mathematical Resolution of Complex Chromatographic Measurements, Ph.D. thesis, University of Copenhagen, 2008.
[76] J.M. Amigo, T. Skov, R. Bro, J. Coello, S. Maspoch, TrAC, Trend. Anal. Chem. 27 (2008) 714.
[77] T. Gröger, W. Welthagen, S. Mitschke, M. Schaffer, R. Zimmermann, J. Sep. Sci. 31 (2008) 3366.
[78] K.M. Pierce, B.W. Wright, R.E. Synovec, J. Chromatogr. A 1141 (2007) 106.
[79] K.M. Pierce, J.C. Hoggard, J.L. Hope, P.M. Rainey, A.N. Hoofnagle, R.M. Jack, B.W. Wright, R.E. Synovec, Anal. Chem. 78 (2006) 5068.