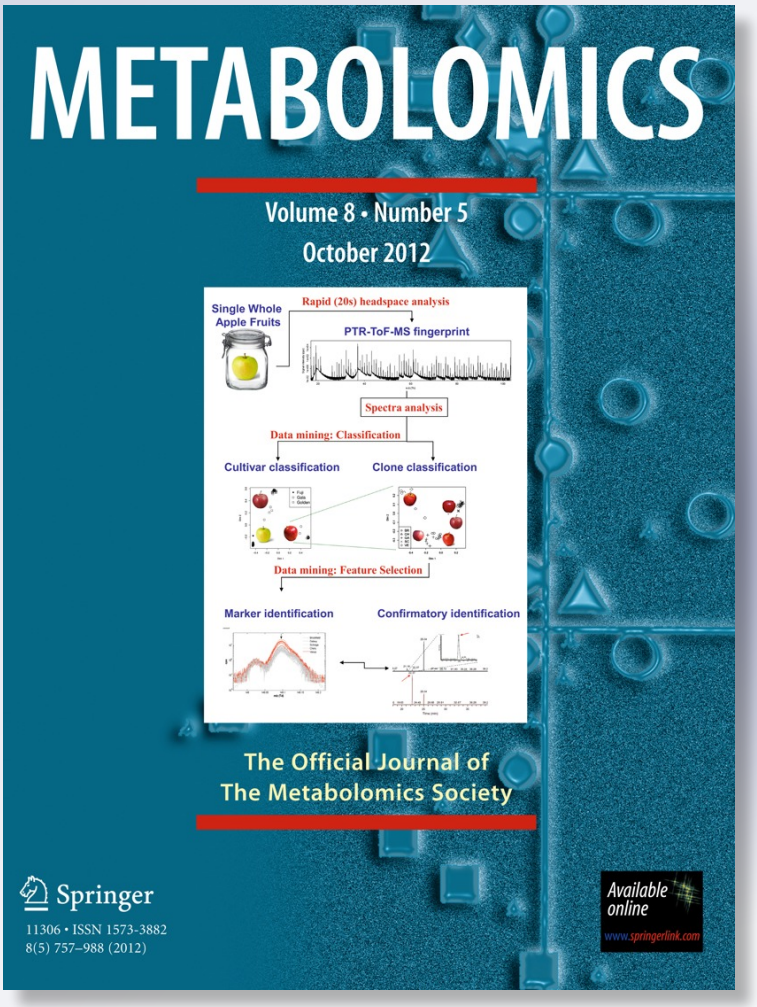
PTR-ToF-MS and data mining methods: a new tool for fruit metabolomics

Luca Cappellin, Christos Soukoulis, Eugenio Aprea, Pablo Granitto, Nicola Dallabetta, Fabrizio Costa, Roberto Viola, Tilmann D. Märk, et al.

Metabolomics
An Official Journal of the Metabolomics Society

ISSN 1573-3882
Volume 8
Number 5

Metabolomics (2012) 8:761-770
DOI 10.1007/s11306-012-0405-9



Springer

Metabolomics (2012) 8:761–770
DOI 10.1007/s11306-012-0405-9

ORIGINAL ARTICLE

PTR-ToF-MS and data mining methods: a new tool for fruit metabolomics

Luca Cappellin · Christos Soukoulis · Eugenio Aprea · Pablo Granitto · Nicola Dallabetta · Fabrizio Costa · Roberto Viola · Tilmann D. Märk · Flavia Gasperi · Franco Biasioli

Received: 5 August 2011 / Accepted: 1 February 2012 / Published online: 4 March 2012

Abstract Proton Transfer Reaction-Mass Spectrometry (PTR-MS) in its recently developed implementation based on a time-of-flight mass spectrometer (PTR-ToF-MS) has been evaluated as a possible tool for rapid non-destructive investigation of the volatile compounds present in the metabolome of apple cultivars and clones. Clone characterization is a cutting-edge problem in technical management and royalty application, not only for apple, aiming at unveiling real properties which differentiate the mutated individuals. We show that PTR-ToF-MS coupled with multivariate and data mining methods may successfully be employed to obtain accurate varietal and clonal apple fingerprint. In particular, we studied the VOC emission profile of five different clones belonging to three well known apple cultivars, such as 'Fuji', 'Golden Delicious' and 'Gala'. In all three cases it was possible to set classification models which can distinguish all cultivars and some of the clones considered in this study. Furthermore, in the case of 'Gala' we also identified estragole and hexyl 2-methyl butanoate contributing to such clone characterization. Beside its applied relevance, no data on the volatile profiling of apple clones are available so far, our study indicates the general viability of a metabolomic approach for volatile compounds in fruit based on rapid PTR-ToF-MS fingerprinting.

Keywords Proton transfer reaction-mass spectrometry · Apple (*Malus domestica*) · Cultivars · Clones · Chemometrics · Data mining · Marker identification

L. Cappellin · C. Soukoulis · E. Aprea · F. Costa · R. Viola · F. Gasperi · F. Biasioli (✉)
IASMA Research and Innovation Centre, Fondazione Edmund Mach, Via E. Mach, 1, 38010 S. Michele a/A, Italy
e-mail: franco.biasioli@iasma.it

L. Cappellin · T. D. Märk
Institut für Ionenphysik und Angewandte Physik, Leopold-Franzens Universität Innsbruck, Technikerstr. 25, 6020 Innsbruck, Austria

P. Granitto
CIFASIS, French Argentina International Center for Information and Systems Sciences, UPCAM (France)/UNR-CONICET (Argentina), Bv 27 de Febrero 210 Bis, 2000 Rosario, Argentina

N. Dallabetta
IASMA Consulting and Services Centre, Fondazione Edmund Mach, Via E. Mach, 1, 38010 S. Michele a/A, Italy

1 Introduction

Comprehensive metabolomics approaches are mostly based on hyphenated methods where chromatographic separation is followed by mass spectrometry (Dunn and Ellis 2005). The need, intrinsic to metabolomics, of high-throughput methods is however driving the development and application of more rapid, chromatography-free techniques that, on one side, allow screening larger sample-sets and, on the other, reduce the possible artefacts related to the extraction and concentration procedures (Han et al. 2009). Among the possible alternative methods, direct infusion mass spectrometry is one of the most widely investigated in the case of liquid samples and has been used with different ionisation methods and mass analysers: see for instance (Favé et al. 2011; Højer-Pedersen et al. 2008; Mattoli et al. 2010; McDougall et al. 2008). The main drawback of this approach is the so called "ion suppression" (Annesley 2003; Sterner et al. 2000): the most abundant ions generated by the ionization methods used are those exhibiting the highest yield and this necessarily leads to a suppression effect for the other less abundant molecular species thereby

providing a not completely reliable metabolite profiling (Mattoli et al. 2010).

Among the different classes of secondary metabolites, volatile compounds are a particularly interesting group. In fact, they can be measured without previous derivatisation and thus provide a rapid and non-invasive tool for metabolomic investigation, an advantage that may be even necessary in research fields as for instance breath analysis (Herbig et al. 2009; Španěl and Smith 2011) or plant physiology (Tholl et al. 2006). Moreover they have a high economic and fundamental relevance in fields as diverse as fruit appreciation by consumers (Taylor 2000), breath marker of important diseases (Greenwald et al. 2010) or atmospheric chemistry because, e.g., of plant emissions (Peñuelas and Staudt 2010).

For these and other reasons, i.e. monitoring of fast dynamic processes, direct injection mass spectrometry (DIMS) for volatile compound detection and quantification has recently being investigated and several methods have been proposed and applied in different fields like environmental monitoring, health sciences, food science and technology (Biasioli et al. 2011b) but little has been done in metabolomics and this only very recently. See for instance (Cajka et al. 2010; Gu et al. 2011). A particularly promising DIMS method for volatile compound detection is proton transfer reaction-mass spectrometry (PTR-MS) and in particular its recent version based on a Time-of-Flight (ToF) mass analyser (PTR-ToF-MS) that, while preserving ultrahigh sensitivity (parts per trillion by volume), increases rapidity and analytical information: a single spectrum can be obtained in a split second and in most cases the sum formula of the observed peaks can be determined (Cappellin et al. 2010a). Compared to standard GC/MS, PTR-MS allows to reduce of approximately 100 times the measurement time required to characterize a sample headspace. PTR-MS has been successfully applied for characterization of fresh and processed foods (Aprea et al. 2009; Aprea et al. 2006; Biasioli et al. 2006; Fabris et al. 2010) and identification of origin (Aprea et al. 2007; Araghipour et al. 2008) and has been reviewed by (Blake et al. 2009; De Gouw and Warneke 2007) and for food applications by (Biasioli et al. 2011a). Thanks to its time resolution it has been used also for the on-line monitoring of VOC headspace in several model and real food systems (Soukoulis et al. 2010). The basic idea of PTR-MS is the chemical ionization of VOCs having proton affinity higher than water by means of reaction with hydronium ions (H_3O^+). PTR-MS is characterized by a large dynamic range, being sensitive from the low pptv region (parts per trillion by volume) up to several ppmv (Cappellin et al. 2011b): this is a very important aspect for applications in metabolomics dealing with metabolites whose abundance can vary by many orders of magnitude (Dunn et al. 2005). PTR-MS precision and accuracy have been evaluated in several works for a large number of compounds; typical values are lower than 2–5%, respectively (De Gouw et al. 2003). The limitations of the quadrupole version of PTR-MS (PTR-QUAD-MS), which is characterized by a unit mass resolution and a relatively slow spectra acquisition have been recently overcome by coupling PTR-MS with a Time-of-Flight (ToF) mass analyser (Jordan et al. 2009). This offers several advantages including higher mass resolution ($m/\Delta m$ up to 8,000) and higher time resolution (0.1 s). The very first applications of PTR-ToF-MS in environmental sciences (Müller et al. 2010), in food science and technology (Fabris et al. 2010; Soukoulis et al. 2010) and in health sciences (Herbig et al. 2009) were recently published.

Recent literature addresses also the challenges related to PTR-ToF-MS applications, mostly in data handling and analysis (Cappellin et al. 2011a), and indicates the usefulness of the application of multivariate and data mining methods to PTR-ToF-MS rapid fingerprinting. Moreover we have indication that the PTR-QUAD-MS fingerprint, can be used to efficiently discriminate fruits of different cultivars over several years (Granitto et al. 2007a) and that can be related to molecular information (Zini et al. 2005).

All these developments and results suggest that PTR-ToF-MS can be a valuable tool for fruit metabolomic. The present work aims at demonstrating this by investigating, as a relevant case study, the possibility of applying a metabolomic approach based on the rapid and non invasive PTR-ToF-MS fingerprinting for the classification of apple cultivars and clones.

In standard apple nursery management, trees are vegetatively propagated, enabling the mitosis as the sole process to replicate their genetic material. This is accomplished for two main reasons, the first is to produce a set of identical individuals (clone), otherwise impossible to obtain by crossing due to the heterozygous nature of the apple genome, and the second is to reduce the juvenile unproductive phase. However, some differences have been observed in specific cultivars because of their attitude to generate clonal variation caused by mutation events which are able to induce stable genetic changes (Forneck 2005). Such mutations have been largely identified specifically for vegetative habit or fruit colour. These two phenomena (easy to detect visually) have also a great impact in the productive system, affecting plant canopy management and fruit quality properties. However, the specific characterization of this variation is normally very difficult because it might depend on a change in a single nucleotide within the entire genome, and most molecular marker techniques which are available nowadays are not efficient and cost-effective for an exhaustive clonal fingerprinting (Venturi et al. 2005). Because it is difficult to determine if an

observed difference can be attributed to a true genetic mutation event (clone variability) or just to an environmental effect, there is a considerable interest in this field, especially for apple where some of the most cultivated varieties such as 'Gala' (White 1991), 'Braeburn' and 'Fuji', have generated important clones. Grapevine (*Vitis* ssp. L.) is another example were clone characterization plays a major role in improving genetic variation. Apple and grape, both relying on vegetative propagation, are thus two excellent case of study for clonal variation (Forneck 2005). The possibility of employing a high resolution technology to detect physiological changes among different clones will represent a great methodological improvement to support their characterization and better define and control their features.

Published results on fingerprinting of apple clones are mainly based on biochemical composition (Sedov and Makarkina 2008) and genetic polymorphism (e.g. employing sequence specific amplifying polymorphism (Venturi et al. 2005)), but, to the best of our knowledge, there are not published studies considering the differences of volatile emission profiles. We consider thus the description of apple cultivars and clones by rapid and non invasive volatile compounds phenotyping an interesting benchmark application.

In this study the analytical capabilities provided by PTR-ToF-MS are initially employed to rapidly and non-invasively investigate the volatile compound emission profile of three commercial apple cultivars ('Gala', 'Fuji' and 'Golden Delicious'), and then to investigate the possibility of classifying 5 clones for each cultivar. SPME-GC/MS was used to support and confirm identification of the compounds.

2 Materials and methods

2.1 Samples

The fruits used in this study were collected from trees grown in the same plot located in the experimental orchard of the Fondazione Edmund Mach (Trento, Italy). Plants were maintained following regular agronomical practice of pruning, thinning and chemical treatment to prevent fungal disease and insect attack.

We considered 5 clones of three different cultivars: 'Gala', 'Golden Delicious' and 'Fuji' (Table 1). For each sample (thought as a clone) 10 fruits from three plants of the same clone were harvested.

Fruit collection was carried out based on the commercial harvest decided on the optimal ripening stage established by evaluating colour change, fruit firmness, total sugar content and starch index. The ten fruits per each clone, without any visible damage, were selected in order to have homogeneous shape and colour.

Table 1 List of considered clones

	'Gala'	'Fuji'	'Golden Delicious'
1.	Brookfield	Fujiko	Clone B
2.	Cherry	Fubrox	Golden 2000
3.	Galaxy	Kiku 8	Smoothee 764
4.	Schniga	Aztec	Quemoni
5.	Venus	Spike	Leratess

Prior to analysis, fruit were kept at room temperature ($\sim$20°C) for 7 days in order to perform the volatile assessment during the climacteric phase, coincident with the ethylene burst.

It is in fact known that in commercially harvested apples the production of ethylene is completed during the post-harvest ripening (Costa et al. 2010a, b) and, recently, a genomic approach has revealed that the final steps of the biochemical pathways involved in aroma production in apple are regulated by the amount of the hormone ethylene (Schaffer et al. 2007).

All samples have been measured by PTR-ToF-MS while only selected samples have been evaluated by GC–MS in order to support compounds identification.

2.2 Proton transfer reaction time of flight mass spectrometry (PTR-ToF-MS) analysis

For the analysis of volatile compounds, each single fruit was placed in glass jars (1,000 mL, 30°C) provided with two Teflon/silicone septa on opposite sides. To standardise the measurements all samples were equilibrated at 37°C for 30 min in a water bath prior to analysis. VOCs were then measured by direct injection of the head space mixture into the PTR-ToF-MS drift tube via a heated (110°C) peek inlet for 30 s, allowing the acquisition of 30 average spectra. Measurements were carried out following the procedure described in previous works for other food samples (Fabris et al. 2010; Soukoulis et al. 2010) using a commercial PTR-ToF-MS 8000 apparatus (Ionicon Analytik GmbH, Innsbruck, Austria), in its standard configuration (V mode). The sampling time per channel of ToF acquisition is 0.1 ns, amounting to 350,000 channels for a mass spectrum ranging up to $m/z = 400$, with the following conditions in the drift tube: drift voltage 600 V, temperature 110°C and pressure 2.25 mbar. Every single spectrum is the sum of 28,600 acquisitions lasting for 35 μs each.

2.3 Spectra analysis

The external calibration automatically done by the acquisition program provided a poor mass accuracy, thus internal calibration of ToF spectra was performed off-line

(Cappellin et al. 2010a). Signal losses caused by the detector dead time and duty cycle were corrected for (Cappellin et al. 2011b). Data pre-processing on ToF spectra was carried out to remove the baseline and noise reduction was achieved by averaging over the 30 consequent ToF spectra corresponding to the same sample, thereby allowing the improvement of the signal-to-noise ratio by about five times. Peak identification and area extraction then followed the procedure described in details by (Cappellin et al. 2011a). Throughout this paper we report experimental *m/z* values up to the third decimal and the expected exact *m/z* values up to the fourth, VOC concentration is expressed in ppbv (part per billion by volume) and has been calculated from peak areas according to the formula described by (Lindinger et al. 1998), using a constant value for the reaction rate coefficient ($k_R = 2 \times 10^{-9}$ cm^3/s). This introduces a systematic error for the absolute concentration for each compound that in most cases is below 30% and can be accounted for if the actual rate constant is available (Cappellin et al. 2010b). The obtained concentration data for each apple sample were then normalized by the total emission of that apple (Aprea et al. 2006; Granitto et al. 2007a).

2.4 SPME/GC–MS analysis

For GC analysis the same procedure used for PTR-Tof-MS analysis was adopted. Each selected fruit was placed in glass jars (1,000 mL, 30°C) provided with two Teflon/silicone septa on opposite sides and kept at room temperature for 30 min prior volatile compounds collection. Headspace volatile compounds were extracted and concentrated on a 2 cm Solid Phase Microextration fibre coated with divinylbenzene/carboxen/polydimethylsiloxane 50/30 μm (DBV/CAR/PDMS, Supelco, Bellefonte, PA, USA) using a manual holder (Supelco, Bellefonte, PA, USA). The fibre was exposed to the apple headspace for 30 min. Volatile compounds adsorbed on the SPME fibre were desorbed at 250°C in the injector port of a GC interfaced with a mass detector which operates in electron ionization mode (EI, internal ionization source; 70 eV) with a scan range from *m/z* 35–300 (GC Clarus 500, PerkinElmer, Norwalk CT, USA). Separation was achieved on a HP-Innowax fused-silica capillary column (30 m, 0.32 mm ID, 0.5 μm film thickness; Agilent Technologies, Palo Alto, CA, USA). The GC oven temperature program consisted in 40°C for 3 min, then 40–220°C at 4°C min^{-1}, stable at 220°C for 1 min, and then 220–250 at 10°C min^{-1}, and finally 250°C for 1 min. Helium was used as carrier gas with a constant column flow rate of 2 mL min^{-1}. Compounds identification was based on mass spectra matching with the standard NIST05/Wiley98 libraries and retention indices (RI) of authentic reference standards.

2.5 Statistical analysis

Our dataset was organized in a matrix with 150 rows, corresponding to the 150 analysed apple samples (3 cultivars × 5 clones × 10 biological replicates), and 727 columns, each corresponding to an identified PTR-ToF-MS peak, containing the normalized intensity of the peaks.

Data analysis follows the procedure explained in details in previous studies (Cappellin et al. 2011a) and will only be briefly reviewed here. Principal component analysis (Jolliffe 2002) was used as a graphical tool to have a first insight on the data. Supervised classification methods were employed to actually assess the separability of the classes. Random Forest (RF) (Breiman 2001), Penalized Discriminant Analysis (PDA) (Wold et al. 2001), Discriminant Partial Least Squares (dPLS) (Wold et al. 2001) and Support Vector Machines (SVM) (Vapnik 1995), were applied according to (Granitto et al. 2007a) and its recent implementation for PTR-ToF-MS data. A leave-one-out (LOO) procedure and confusion matrices were used to evaluate the results of the classification methods (Aprea et al. 2011; Westerhuis et al. 2008). In order to complement PCA, RF was also used to analyse the data set in a graphical way (Granitto et al. 2007b). The most relevant peaks for the classification problem were obtained via an appropriate feature selection method as described in (Cappellin et al. 2011a), but with a simple modification: instead of the RF-RFE method (Granitto et al. 2006), we applied the PDA-RFE selection method, in which we used a PDA classifier in order to rank peaks into the RFE loop. We used PDA-RFE because, as we show in the next section, PDA usually produces the best discrimination results for the data sets considered.

3 Results and discussion

3.1 Classification of apple cultivars

A first insight into the data is provided by the PCA in Fig. 1a. The first and the second principal components explain 23.9 and 20.1% of the total variance, respectively. It is clear that the three cultivars are well separated, with a very marked discrimination between 'Gala' and the other two cultivars. This preliminary result confirms that PTR-ToF-MS is capable of distinguishing the different VOC profile in the headspace of diverse apple cultivars as it was already pointed out (Cappellin et al. 2011a). The use of a supervised classification method such as Random Forest stresses the discrimination between the considered cultivars (Fig. 1b). Leave-One-Out cross validation confirms that the cultivars can be unambiguously classified (data not shown). Figure 1a shows the capacity of PTR-ToF–MS to

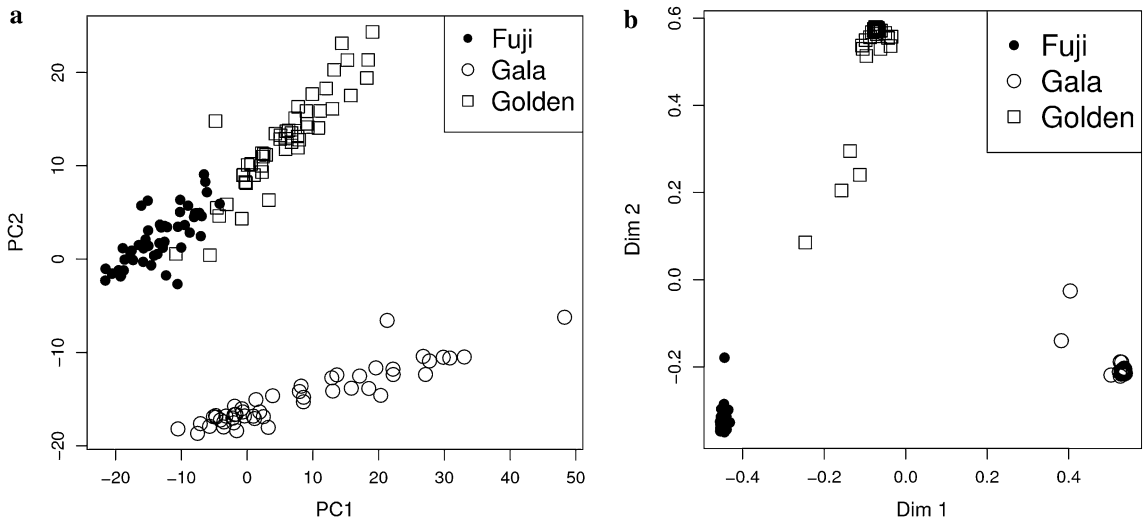


Fig. 1 **a** First and second component of the PCA analysis of the PTR-TOF-MS spectral data of all samples. **b** Random Forest graphical output for the discriminant analysis of the PTR-ToF-MS spectral data of all samples

Table 2 Classification errors for the clones of 'Gala', 'Golden Delicious' and 'Fuji'

	'Gala'	'Fuji'	'Golden Delicious'
RF	0.52	0.6	0.82
PDAA	0.44	0.82	0.78
SVM	0.48	0.88	0.76
dPLS	0.56	0.76	0.82

characterize different cultivars based on their volatile emission rather than their genetic relationship. In fact, 'Gala' and 'Golden Delicious' are well clustered in the plot, even if they are more genetically related than 'Fuji'.

3.2 Classification of apple clones

The possibility of discriminating different cultivars by rapid and non-invasive analysis is interesting but somehow trivial and will not be discussed further. In fact, the present work addresses a more challenging question: can PTR-ToF-MS analysis unveil significant differences in the VOC emission profiles of the considered clones of the same cultivar. In terms of classification capability this would mean that an accurate classification of the clones would be possible using the PTR-ToF-MS fingerprints. To address this issue we focus on each individual cultivar separately.

Table 2 reports the classification error of the clones by the supervised methods RF, PDA, SVM and dPLS. For 'Golden Delicious' and 'Fuji' the classification errors are very high and similar to the random choice level (0.80) for all methods, suggesting that PTR-ToF-MS fingerprints cannot discriminate the five clones at the same time. 'Gala' clones, on the other side, displays a relatively lower classification error for all discriminant methods and in particular for PDA.

A deeper insight in the performances of the classification models is provided by the confusion matrices reported in Table 3. For brevity only the confusion matrices corresponding to PDA are reported. In the case of 'Golden Delicious' and 'Fuji' all clones are confused and no evidence of class separation appears.

The confusion matrix of 'Gala' clones indicates that 'Gala' Venus samples are correctly classified in 9 cases out of 10 cases, indicating that this clone has a significant and robust difference in the VOC profile compared to the other 'Gala' clones. A standard and unsupervised PCA analysis graphically suggests such a separation and the RF graph shows it clearly (see Fig. 2a). For other clones the separation is less clear but evidently better than random choice (Tables 3, 4). Table 4 compares the classification results of the different methods, showing that Venus samples are in general correctly classified independently of the method employed.

The graphical representation of the 'Gala' clone classification by RF highlights once more the separation of 'Gala' Venus. Moreover, both graphical analyses suggest the presence of three groups, represented by Venus, Schniga and Galaxy. We therefore refine our analysis by only considering these three clones.

The confusion matrices obtained with the prediction models that consider only Galaxy, Schniga and Venus are reported in Table 5. The presence of three classes is well confirmed; in particular PDA and SVM provide very good predictions with only 3 misclassified samples (13%). As expected from Fig. 2, there is some overlap between Galaxy and Schniga clones that is evident for all discriminant methods.

Table 3 Confusion matrices for the classification by PDA of considered apple cultivar clones

'Gala'	*1*	*2*	*3*	*4*	*5*	*'Golden Delicious'*	*1*	*2*	*3*	*4*	*5*	*'Fuji'*	*1*	*2*	*3*	*4*	*5*
1	**4**	**1**	**3**	**1**	**1**	*1*	**1**	**4**	**2**	0	**3**	*1*	**2**	**2**	0	**1**	**5**
2	**2**	**3**	**2**	**3**	0	*2*	**1**	**2**	**1**	**4**	**2**	*2*	**1**	**2**	**1**	**4**	**2**
3	**3**	**1**	**5**	**1**	0	*3*	**1**	**1**	**5**	**2**	**1**	*3*	**1**	**1**	**1**	**6**	**1**
4	**1**	**2**	**1**	**6**	0	*4*	0	**5**	0	**2**	**3**	*4*	**1**	**2**	**3**	**1**	**3**
5	**1**	0	0	0	**9**	*5*	**1**	**5**	0	**3**	**1**	*5*	**4**	**1**	**1**	**2**	**2**

Rows correspond to the true classes and columns to the predicted ones

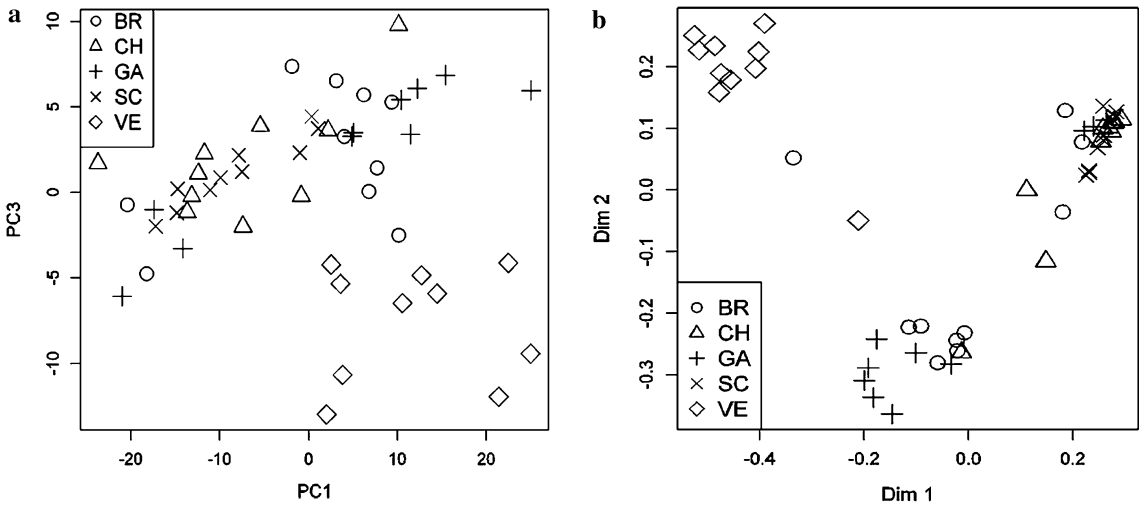


Fig. 2 **a**. First and third component of the PCA analysis of the PTR-ToF-MS spectral data of all 'Gala' samples. The components explain 22.3–3.7% of the total variance, respectively. **b** Random Forest graphical output for the discriminant analysis of the PTR-ToF-MS spectral data of all 'Gala' samples

Table 4 Confusion matrices for the classification by RF, SVM and dPLS (see Table 3 for PDA) of considered 'Gala' clones

RF	*1*	*2*	*3*	*4*	*5*	*SVM*	*1*	*2*	*3*	*4*	*5*	*dPLS*	*1*	*2*	*3*	*4*	*5*
1	**6**	**1**	0	**2**	**1**	*1*	**3**	**2**	**3**	**1**	**1**	*1*	**2**	**1**	**4**	**1**	**2**
2	**2**	0	**1**	**7**	0	*2*	**2**	**2**	**1**	**5**	0	*2*	**3**	**3**	**2**	**2**	0
3	0	**1**	**6**	**2**	1	*3*	**1**	0	**7**	**2**	0	*3*	**2**	**1**	**5**	**1**	**1**
4	**1**	**6**	0	**3**	0	*4*	**1**	**2**	**1**	**6**	0	*4*	**2**	**2**	**2**	**4**	0
5	**2**	0	0	0	**8**	*5*	**2**	0	0	0	**8**	*5*	**1**	0	0	0	**9**

Concerning the other two cultivars it is interesting to point out that a PCA/RF analysis of 'Fuji' clones, similar to Fig. 2 (not shown), suggests that three of the clones should be distinguishable, namely Aztec, Fubrox and Fujiko. The confusion matrices (not shown) for our four discriminant methods on this reduced problem show that RF have the best performance in this problem with 8 errors over the 30 samples (27%) evenly distributed among the three classes. The other discriminant methods have lower performances in this problem. In the case of 'Golden Delicious' clones, only two classes (Golden 2000 and Quemoni) can be discriminated with some accuracy. The RF classifier in this case leads to 5 errors over 20 samples (25%).

3.3 Feature selection

Returning to the 'Gala' case, the selection of the peaks that are more relevant to separate the clone classes is done in two steps. First, the behaviour of the mean discrimination error of PDA as a function of the number of masses in the

Table 5 Confusion matrices for the classification by RF, PDA, PLS and SVM of three 'Gala' clones (30 samples)

RF	*3*	*4*	*5*	*PDA*	*3*	*4*	*5*	*SVM*	*3*	*4*	*5*	*dPLS*	*3*	*4*	*5*
3	**6**	**3**	**1**	*3*	**8**	**2**	0	*3*	**8**	**2**	0	*3*	**5**	**5**	0
4	**1**	**9**	0	*4*	**1**	**9**	0	*4*	**1**	**9**	0	*4*	**1**	**9**	0
5	0	0	**10**	*5*	**1**	0	**9**	*5*	**1**	0	**9**	*5*	**1**	0	**9**

Classification errors are 0.17 (RF), 0.13 (PDA), 0.13 (SVM), 0.23 (PLS)

3 Galaxy, *4* Schniga, *5* Venus

problem is analysed (Fig. 3). In our case a good trade-off is provided by four masses. In a second step, we asses which are the most important peaks for models constructed by PDA using only 4 masses. Figure 4 reports how often the listed masses are selected among the 4 most relevant peaks by PDA, over the 30 LOO replicated PDA experiments. Two peaks, namely $m/z = 149.098$ and $m/z = 187.170$, turn out to be particularly relevant, being selected more than 90% of the times and can be considered as potential markers for clone differentiation. The achieved mass accuracy (Cappellin et al. 2010a) allows us to classify these peaks as corresponding to the sum formulas $C_{10}H_{13}O^+$ and $C_{11}H_{23}O_2^+$, respectively. In order to identify the actual compounds, the headspace of the three samples for Venus 'Gala' clone were analysed by SPME GC–MS. Among the acquired chromatographic peaks, the only peak that can possibly contribute to the signal at m/z 149.098 is the one eluting at 22.37 min (see Fig. 5). The peak at 25.00 min is α-Farnesene. The mass spectrum of peak at 22.37 min matches the estragole spectra reported in the NIST05/Wiley98 MS libraries. Estragole has been reported as one of the compounds contributing to the apple aroma and is thought to be synthesized from the phenylpropanoid pathway (Schaffer et al. 2007).

For m/z 187.170, and thus for a compound having a molecular weight of 186 amu, we identified three peaks matching with hexyl 2-methyl butanoate (rt 15.74 min), butyl heptanoate (rt 18.27 min) and propyl octanoate (rt 18.49 min) respectively. ((Young et al. 2004) and reference therein) report 6 compounds with a nominal mass of 186 but only the three mentioned above have a relatively large signal at m/z 186. Butyl heptanoate and propyl octanoate, when present, are reported at low concentration ((Young et al. 2004) and reference therein). Furthermore, hexyl 2-methyl butanoate gives the highest signal at m/z 186 (Fig. 6, lower panel) thus we conclude that the signal m/z 187.170 corresponds mostly to this compound that has a typical apple-like aroma and is reported to improve the quality of 'Golden Delicious' apples (Paillard 1990). Red Delicious is also reported for the capacity to synthesize 2-Methylbutanoate esters, including hexyl 2-methylbutanoate (Rowan et al. 1996).

4 Conclusions

We investigated the possibility to apply PTR-ToF-MS as a metabolomic tool for the rapid and non invasive analysis of

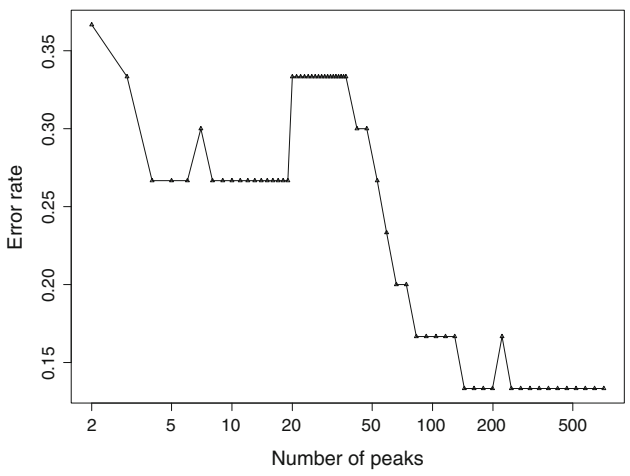


Fig. 3 Mean classification error rate as a function of the number of peaks selected by the PDA-RFE method for the 3 'Gala' clone classes (30 samples) problem

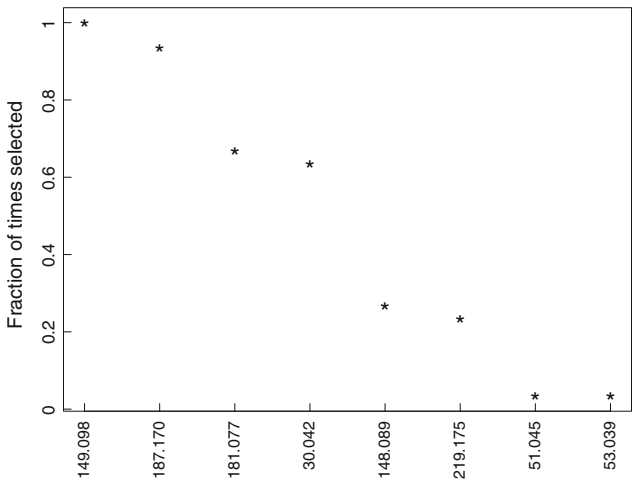


Fig. 4 Fraction of times that each peak was selected among the 4 more discriminant features on PDA over the 30 LOO replicated experiments, for the 3 'Gala' clones classes problem

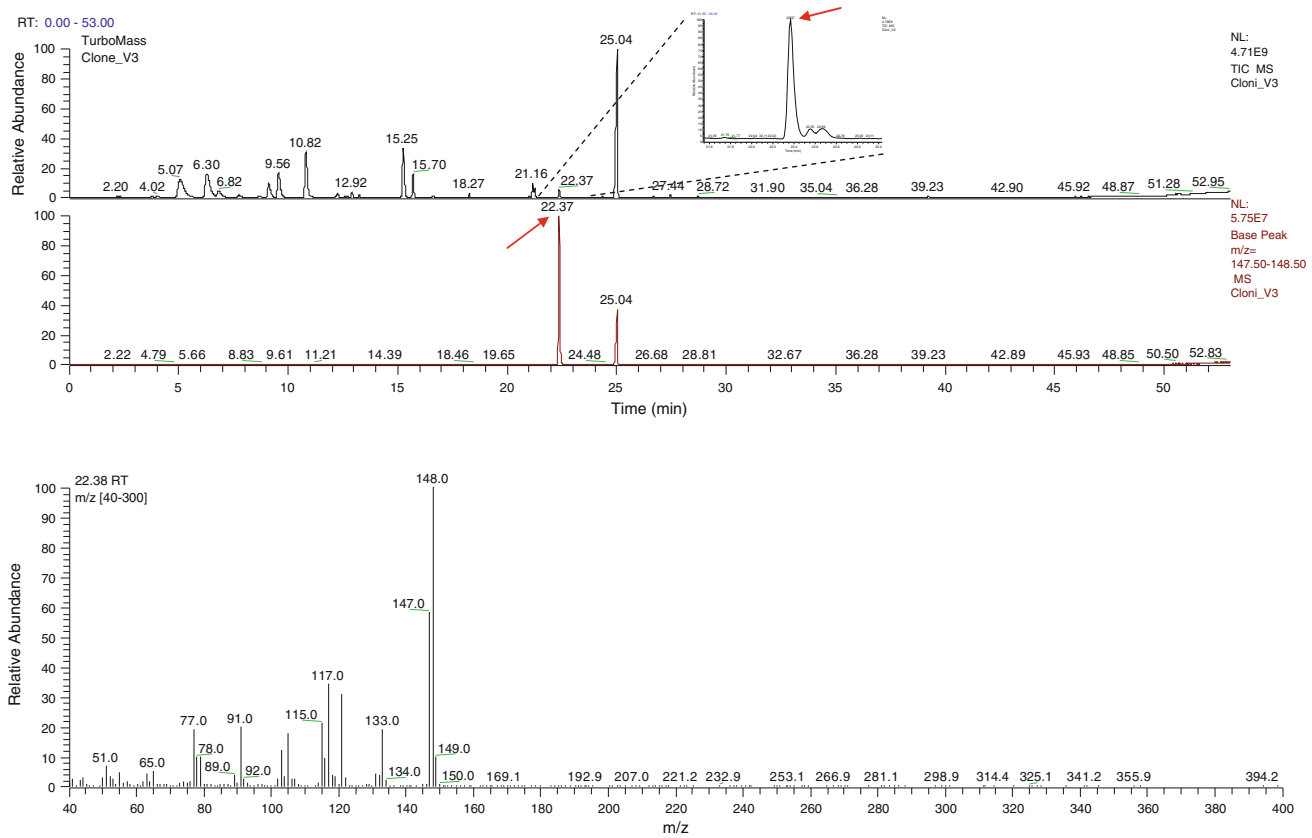


Fig. 5 GC/MS. Chromatogram of a Venus 'Gala' apple. In the upper panel are reported the total ion current and the single ion (*m/z* 148) traces. In the lower panel is reported the ion spectra of the peak at 22.37 min of retention time. This spectrum match the estragole spectrum reported in NIST05/Wiley98 libraries

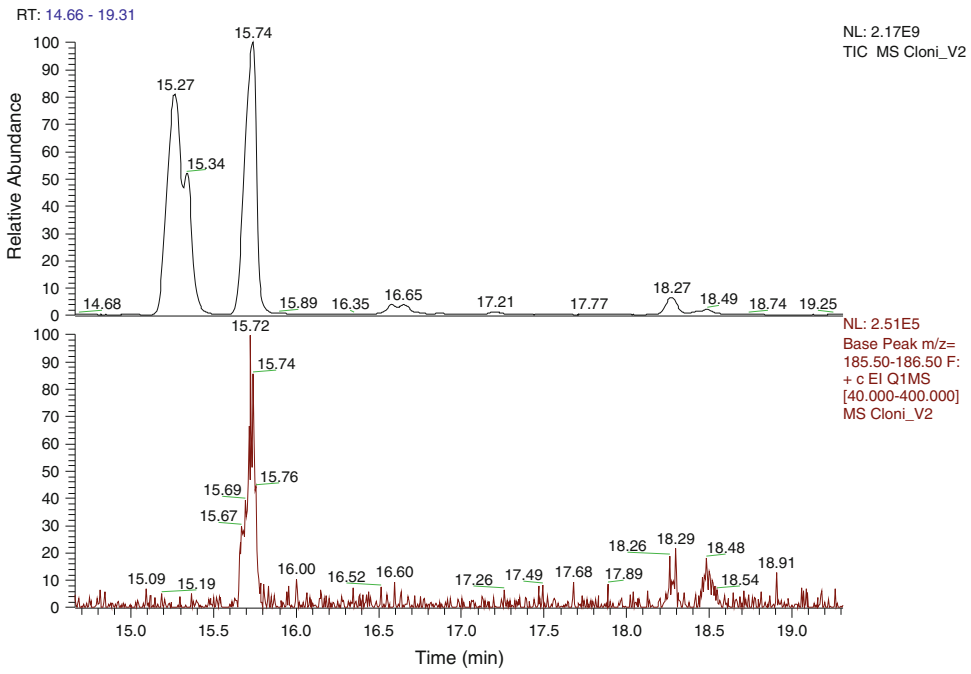


Fig. 6 GC/MS. Chromatogram of a Venus 'Gala' apple. Total ion current and single ion (*m/z* 186) traces are reported

apple cultivars and clones. Our methodology can easily differentiate among different apple cultivars. The differentiation and characterization of clones is generally a more difficult problem, given their extremely similar genetic profile. In fact, standard genetic analyses based on molecular markers usually do not differentiate among them

and the use of less rapid and cost effective methods, such as complete genome re-sequencing, could be required. In the present study we investigated whether differences in volatile compounds released from 5 different clones of 'Golden Delicious', 'Gala' and 'Fuji' could be rapidly and non-invasively detected by the newly available technique PTR-ToF-MS. Beside that, this novel technique was also applied in order to analyze the different volatile compounds pattern for the clones under investigation. Relevant differences were found for the considered clones of 'Gala', 'Golden Delicious' and 'Fuji'. In particular, three 'Fuji' clones and two 'Golden Delicious' clones can be discriminated with good accuracy. In the case of 'Gala' the clones clustered in three different classes, with Venus remarkably showing a clear separation from the other four 'Gala' clones. The employed feature selection method identified two mass peaks as primarily important for the construction of the multivariate calibration models highlighting the separation. Further SPME-GC/MS analysis on few selected samples supported the identification of the compounds related to these peaks: estragole and hexyl 2-methyl butanoate. Estragole in particular is a very important constituent of apple flavour profile.

In conclusion PTR-ToF-MS provides a new tool for metabolomics studies that can rapidly and non-invasively fingerprint the volatile compounds profile of single apple fruits. Data mining methods allow setting classification models that easily distinguish different cultivars. In some cases even clones of the same cultivar can be correctly identified. Moreover the characteristics of PTR-ToF-MS allow gathering some analytical information that has been confirmed or supported by GC/MS headspace analysis allowing the full identification of the selected markers.

Acknowledgements Work partially supported by PAT (AP 2009-2011). PMG acknowledges partial support from ANPCyT (grant PICT 237/08).

References

Annesley, T. M. (2003). Ion suppression in mass spectrometry. *Clinical Chemistry, 49*, 1041–1044. doi:10.1373/49.7.1041.

Aprea, E., Biasioli, F., Sani, G., et al. (2006). Proton transfer reaction-mass spectrometry (PTR-MS) headspace analysis for rapid detection of oxidative alteration of olive oil. *Journal of Agriculture and Food Chemistry, 54*, 7635–7640.

Aprea, E., Biasioli, F., Carlin, S., et al. (2007). Rapid white truffle headspace analysis by proton transfer reaction mass spectrometry and comparison with solid-phase microextraction coupled with gas chromatography/mass spectrometry. *Rapid Communications in Mass Spectrometry, 21*, 2564–2572.

Aprea, E., Biasioli, F., Carlin, S., et al. (2009). Investigation of Volatile compounds in two raspberry cultivars by two headspace techniques: Solid-phase microextraction/gas chromatography–mass spectrometry (SPME/GC–MS) and proton-transfer reaction–mass spectrometry (PTR–MS). *Journal of Agriculture and Food Chemistry, 57*, 4011–4018. doi:10.1021/jf803998c.

Aprea, E., Gika, H., Carlin, S., et al. (2011). Metabolite profiling on apple volatile content based on solid phase microextraction and gas-chromatography time of flight mass spectrometry. *Journal of Chromatography,*. doi:10.1016/j.chroma.2011.05.019.

Araghipour, N., Colineau, J., Koot, A., et al. (2008). Geographical origin classification of olive oils by PTR-MS. *Food Chemistry, 108*, 374–383. doi:10.1016/j.foodchem.2007.10.056.

Biasioli, F., Gasperi, F., Aprea, E., et al. (2006). Correlation of PTR-MS spectral fingerprints with sensory characterisation of flavour and odour profile of "Trentingrana" cheese. *Food Quality and Preference, 17*, 63–75. doi:10.1016/j.foodqual.2005.06.004.

Biasioli, F., Yeretzian, C., Gasperi, F., & Märk, T. D. (2011a). PTR-MS monitoring of VOCs and BVOCs in food science and technology. *TRAC-Trends in Analytical Chemistry, 30*, 968–977. doi:10.1016/j.trac.2011.03.009.

Biasioli, F., Yeretzian, C., Märk, T. D., et al. (2011b). Direct-injection mass spectrometry adds the time dimension to (B)VOC analysis. *TRAC-Trends in Analytical Chemistry, 30*, 1003–1017. doi:10.1016/j.trac.2011.04.005.

Blake, R., Monks, P., & Ellis, A. (2009). Proton-transfer reaction mass spectrometry. *Chemical Reviews, 109*, 861–896. doi:10.1021/cr800364q.

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.

Cajka, T., Riddellova, K., Tomaniova, M., & Hajslova, J. (2010). Ambient mass spectrometry employing a DART ion source for metabolomic fingerprinting/profiling: a powerful tool for beer origin recognition. *Metabolomics,*. doi:10.1007/s11306-010-0266-z.

Cappellin, L., Biasioli, F., Fabris, A., et al. (2010a). Improved mass accuracy in PTR-TOF-MS: Another step towards better compound identification in PTR-MS. *International Journal of Mass Spectrometry and Ion Physics, 290*, 60–63. doi:10.1016/j.ijms.2009.11.007.

Cappellin, L., Probst, M., Limtrakul, J., et al. (2010b). Proton transfer reaction rate coefficients between H3O + and some sulphur compounds. *International Journal of Mass Spectrometry and Ion Physics, 295*, 43–48. doi:10.1016/j.ijms.2010.06.023.

Cappellin, L., Biasioli, F., Granitto, P., et al. (2011a). On data analysis in PTR-TOF-MS: From raw spectra to data mining. *Sensor and Actuators B-Chemical, 155*, 183–190. doi:10.1016/j.snb.2010.11.044.

Cappellin, L., Biasioli, F., Schuhfried, E., et al. (2011b). Extending the dynamic range of proton transfer reaction time-of-flight mass spectrometers by a novel dead time correction. *Rapid Communications in Mass Spectrometry, 25*, 179–183.

Costa, F., Peace, C. P., Stella, S., et al. (2010a). QTL dynamics for fruit firmness and softening around an ethylene-dependent polygalacturonase gene in apple (Malusxdomestica Borkh.). *Journal of Experimental Botany, 61*, 3029–3039. doi:10.1093/jxb/erq130.

Costa, F., Alba, R., Schouten, H., et al. (2010b). Use of homologous and heterologous gene expression profiling tools to characterize transcription dynamics during apple fruit maturation and ripening. *BMC Plant Biology, 10*, 229. doi:10.1186/1471-2229-10-229.

De Gouw, J., & Warneke, C. (2007). Measurements of volatile organic compounds in the earth's atmosphere using proton-transfer-reaction mass spectrometry. *Mass Spectrometry Reviews, 26*, 223–257. doi:10.1002/mas.20119.

De Gouw J, Goldan P, Warneke C, et al. (2003) Validation of proton transfer reaction-mass spectrometry (PTR-MS) measurements of gas-phase organic compounds in the atmosphere during the New England air quality study (NEAQS) in 2002. Journal of Geophysics Research Atmosphere 108. doi:10.1029/2003JD003863.

Dunn, W. B., & Ellis, D. I. (2005). Metabolomics: Current analytical platforms and methodologies. *TrAC-Trends in Analytical Chemistry, 24*, 285–294. doi:10.1016/j.trac.2004.11.021.

Dunn, W. B., Bailey, N. J. C., & Johnson, H. E. (2005). Measuring the metabolome: Current analytical technologies. *Analyst, 130*, 606. doi:10.1039/b418288j.

Fabris, A., Biasioli, F., Granitto, P., et al. (2010). PTR-TOF-MS and data-mining methods for rapid characterisation of agro-industrial samples: influence of milk storage conditions on the volatile compounds profile of Trentingrana cheese. *Journal of Mass Spectrometry and Ion Physics, 45*, 1065–1074.

Favé, G., Beckmann, M., Lloyd, A. J., et al. (2011). Development and validation of a standardized protocol to monitor human dietary exposure by metabolite fingerprinting of urine samples. *Metabolomics,*. doi:10.1007/s11306-011-0289-0.

Forneck, A. (2005). Plant breeding: Clonality—a concept for stability and variability during vegetative propagation. In K. Esser, U. Lüttge, W. Beyschlag, & J. Murata (Eds.), *Progress in botany* (pp. 164–183). Berlin: Springer.

Granitto, P., Furlanello, C., Biasioli, F., & Gasperi, F. (2006). Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. *Chemometrics and Intelligent Laboratory Systems, 83*, 83–90. doi:10.1016/j.chemolab.2006.01.007.

Granitto, P., Biasioli, F., Aprea, E., et al. (2007a). Rapid and non-destructive identification of strawberry cultivars by direct PTR-MS headspace analysis and data mining techniques. *Sensor Actuators B-Chemistry, 121*, 379–385. doi:10.1016/j.snb.2006.03.047.

Granitto, P., Gasperi, F., Biasioli, F., et al. (2007b). Modern data mining tools in descriptive sensory analysis: A case study with a random forest approach. *Food Quality and Preferences, 18*, 681–689. doi:10.1016/j.foodqual.2006.11.001.

Greenwald, R., Fitzpatrick, A. M., Gaston, B., et al. (2010). Breath formate is a marker of airway s-nitrosothiol depletion in severe asthma. *PLoS ONE, 5*, e11919. doi:10.1371/journal.pone.0011919.

Gu, H., Pan, Z., Xi, B., et al. (2011). Principal component directed partial least squares analysis for combining nuclear magnetic resonance and mass spectrometry data in metabolomics: Application to the detection of breast cancer. *Analytica Chimica Acta, 686*, 57–63. doi:10.1016/j.aca.2010.11.040.

Han, J., Datla, R., Chan, S., & Borchers, C. H. (2009). Mass spectrometry-based technologies for high-throughput metabolomics. *Bioanalysis, 1*, 1665–1684. doi:10.4155/bio.09.158.

Herbig, J., Müller, M., Schallhart, S., et al. (2009). On-line breath analysis with PTR-TOF. *Journal of Breath Research, 3*, 027004. doi:10.1088/1752-7155/3/2/027004.

Højer-Pedersen, J., Smedsgaard, J., & Nielsen, J. (2008). The yeast metabolome addressed by electrospray ionization mass spectrometry: Initiation of a mass spectral library and its applications for metabolic footprinting by direct infusion mass spectrometry. *Metabolomics, 4*, 393–405. doi:10.1007/s11306-008-0132-4.

Jolliffe, I. (2002). *Principal component analysis*. New York: Springer.

Jordan, A., Haidacher, S., Hanel, G., et al. (2009). A high resolution and high sensitivity proton-transfer-reaction time-of-flight mass spectrometer (PTR-TOF-MS). *International Journal of Mass Spectrometry and Ion Physics, 286*, 122–128. doi:10.1016/j.ijms.2009.07.005.

Lindinger, W., Hansel, A., & Jordan, A. (1998). On-line monitoring of volatile organic compounds at pptv levels by means of proton-transfer-reaction mass spectrometry (PTR-MS)—Medical applications, food control and environmental research. *International Journal of Mass Spectrometry and Ion Physics, 173*, 191–241.

Mattoli, L., Cangi, F., Ghiara, C., et al. (2010). A metabolite fingerprinting for the characterization of commercial botanical dietary supplements. *Metabolomics,*. doi:10.1007/s11306-010-0268-x.

McDougall, G., Martinussen, I., & Stewart, D. (2008). Towards fruitful metabolomics: High throughput analyses of polyphenol composition in berries using direct infusion mass spectrometry☆. *Journal of Chromatography B, 871*, 362–369. doi:10.1016/j.jchromb.2008.06.032.

Müller, M., Graus, M., Ruuskanen, T. M., et al. (2010). First eddy covariance flux measurements by PTR-TOF. *Atmospheric Measurement Techniques, 3*, 387–395. doi:10.5194/amt-3-387-2010.

Paillard, N. M. M. (1990). The flavour of apples, pears and quinces. In I. D. Morton & A. J. Macleod (Eds.), *Food flavours part c the flavour of fruit* (pp. 1–41). Amsterdam: Elsevier.

Peñuelas, J., & Staudt, M. (2010). BVOCs and global change. *Trends in Plant Science, 15*, 133–144. doi:10.1016/j.tplants.2009.12.005.

Rowan, D. D., Lane, H. P., Allen, J. M., et al. (1996). Biosynthesis of 2-methylbutyl, 2-methyl-2-butenyl and 2-methylbutanoate esters in Red Delicious and Granny Smith apples using deuterium-labeled substrates. *Journal of Agriculture and Food Chemistry, 44*, 3276–3285.

Schaffer, R. J., Friel, E. N., Souleyre, E. J. F., et al. (2007). A genomics approach reveals that aroma production in apple is controlled by ethylene predominantly at the final step in each biosynthetic pathway. *Plant Physiology, 144*, 1899–1912. doi:10.1104/pp.106.093765.

Sedov, E. N., & Makarkina, M. A. (2008). Biochemical composition of fruit of apple cultivar clones and tetraploid forms. *Russian Agricultural Science, 34*, 71–73. doi:10.3103/S1068367408020018.

Soukoulis, C., Aprea, E., Biasioli, F., et al. (2010). Proton transfer reaction time-of-flight mass spectrometry monitoring of the evolution of volatile compounds during lactic acid fermentation of milk. *Rapid Communications in Mass Spectrometry, 24*, 2127–3134. doi:10.1002/rcm.4617.

Španěl, P., & Smith, D. (2011). Progress in SIFT-MS: Breath analysis and other applications. *Mass Spectrometry Reviews, 30*, 236–267. doi:10.1002/mas.20303.

Sterner, J. L., Johnston, M. V., Nicol, G. R., & Ridge, D. P. (2000). Signal suppression in electrospray ionization Fourier transform mass spectrometry of multi-component samples. *Journal of Mass Spectrometry and Ion Physics, 35*, 385–391. doi:10.1002/(SICI)1096-9888(200003)35:3<385:AID-JMS947>3.0.CO;2-O.

Taylor, A. (2000). Atmospheric pressure chemical ionisation mass spectrometry for in vivo analysis of volatile flavour release. *Food Chemistry, 71*, 327–338. doi:10.1016/S0308-8146(00)00182-5.

Tholl, D., Boland, W., Hansel, A., et al. (2006). Practical approaches to plant volatile analysis. *Plant Journal, 45*, 540–560. doi:10.1111/j.1365-313X.2005.02612.x.

Vapnik, V. (1995). *The nature of statistical learning theory*. New York: Springer.

Venturi, S., Dondini, L., Donini, P., & Sansavini, S. (2005). Retrotransposon characterisation and fingerprinting of apple clones by S-SAP markers. *Theoretical and Applied Genetics, 112*, 440–444. doi:10.1007/s00122-005-0143-8.

Westerhuis, J. A., Hoefsloot, H. C. J., Smit, S., et al. (2008). Assessment of PLSDA cross validation. *Metabolomics, 4*, 81–89. doi:10.1007/s11306-007-0099-6.

White, A. (1991). The Gala apple. *Fruit Varieties Journal, 45*, 2–3.

Wold, S., Sjostrom, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory System, 58*, 109–130.

Young, J. C., Chu, C. L. G., Lu, X., & Zhu, H. (2004). Ester variability in apple varieties as determined by solid-phase microextraction and gas chromatography–mass spectrometry. *Journal of Agriculture and Food Chemistry, 52*, 8086–8093. doi:10.1021/jf049364r.

Zini, E., Biasioli, F., Gasperi, F., et al. (2005). QTL mapping of volatile compounds in ripe apples detected by proton transfer reaction-mass spectrometry. *Euphytica, 145*, 269–279. doi:10.1007/s10681-005-1645-9.