Journal of Chromatography A, 1218 (2011) 3247–3254

Contents lists available at ScienceDirect

Journal of Chromatography A

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

Evaluation of automated sample preparation, retention time locked gas chromatography–mass spectrometry and data analysis methods for the metabolomic study of *Arabidopsis* species

Qun Gu[a,b], Frank David[b], Frédéric Lynen[b], Klaus Rumpel[c], Jasper Dugardeyn[d], Dominique Van Der Straeten[d], Guowang Xu[a,*], Pat Sandra[b,**]

[a] *CAS Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, 457 Zhongshan Road, 116023 Dalian, China*
[b] *Pfizer Analytical Research Centre, Ghent University, Krijgslaan 281, S4-bis, B-9000 Ghent, Belgium*
[c] *Pfizer PharmaTherapeutics R& D, Sandwich Laboratories, Ramsgate Road, Sandwich, Kent CT13 9NJ, UK*
[d] *Laboratory of Functional Plant Biology, Department of Physiology, Ghent University, K.L. Ledeganckstraat 35, B-9000 Ghent, Belgium*

ARTICLE INFO

Article history:
Available online 18 January 2011

Keywords:
Metabolomics
Gas chromatography–mass spectrometry
Arabidopsis thaliana
Multivariate statistical analysis
Principal component analysis
Mass profiler professional

ABSTRACT

In this paper, automated sample preparation, retention time locked gas chromatography–mass spectrometry (GC–MS) and data analysis methods for the metabolomics study were evaluated. A miniaturized and automated derivatisation method using sequential oximation and silylation was applied to a polar extract of 4 types (2 types × 2 ages) of *Arabidopsis thaliana*, a popular model organism often used in plant sciences and genetics. Automation of the derivatisation process offers excellent repeatability, and the time between sample preparation and analysis was short and constant, reducing artifact formation. Retention time locked (RTL) gas chromatography–mass spectrometry was used, resulting in reproducible retention times and GC–MS profiles. Two approaches were used for data analysis. XCMS followed by principal component analysis (approach 1) and AMDIS deconvolution combined with a commercially available program (Mass Profiler Professional) followed by principal component analysis (approach 2) were compared. Several features that were up- or down-regulated in the different types were detected.

1. Introduction

Gas chromatography combined with mass spectrometry (GC–MS) is suitable for the analysis of small molecules such as amino acids, amines, sugars, organic acids, fatty acids, and sterols, in metabolomic studies [1–10]. For the analysis of these polar molecules, derivatisation is however needed and, to this, oximation combined with silylation is often selected [11,12]. Oximation is used to inhibit cyclization of sugars, reducing the number of peaks per solute, and at the same time α-ketoacids are protected against decarboxylation. For silylation, N-methyl-N-(trimethylsilyl)trifluoroacetamide (MSTFA) was found to be superior to alternative chemicals such as N,O-bis(trimethylsilyl)-N-trifluoroacetamide (BSTFA) regarding completeness of derivatisation of amines and amino acids while reducing unwanted side reactions [11]. These derivatisation reactions can be performed at relatively low reaction temperatures. This sequential derivatisation method is however prone to artifact formation. Trying to apply for the published protocols to various metabolomic studies in our laboratory at the real beginning, several problems were encountered in terms of repeatability, artifact formation and stability of derivatised samples. Method validation such as a critical evaluation and optimization was therefore necessary [13–17].

GC–MS analysis following sample preparation should be later optimized. A primary requirement for long term reproducibility is needed, since metabolomic studies imply the analysis of large numbers of samples. Instrument stability and reproducibility of retention times, even after inlet maintenance and column change, are thus important prerequisites.

In this paper, the metabolomics analytical workflow was evaluated including automated sample preparation, GC–MS and data analysis. *Arabidopsis thaliana*, a popular model organism in plant biology and genetics, was used for testing. Two approaches were used to identify the compounds responsible for the largest difference between the samples. XCMS followed by principal component analysis (PCA, approach 1) and AMDIS deconvolution combined with a commercially available program (Mass Profiler Professional, a software enabling to perform multivariate analysis) followed by

* Corresponding author. Tel.: +86 411 84379530; fax: +86 411 84379559.
** Corresponding author.
E-mail addresses: xugw@dicp.ac.cn, dicp402@mail.dlptt.ln.cn (G. Xu), pat.sandra@richrom.com (P. Sandra).

0021-9673/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2011.01.024

PCA (approach 2), were compared and validated using the literature data.

2. Experimental

2.1. Chemicals

Methoxyamine hydrochloride, myristic-d_{27} acid (internal standard), HPLC grade methanol and chloroform were from Sigma–Aldrich (Bornem, Belgium). N-Methyl-N-(trimethylsilyl) trifluoroacetamide with 1% trimethylchlorosilane (MSTFA + 1% TMCS) was from Fluka Sigma (Steinheim, Germany). Anhydrous pyridine was obtained from Thermo Fisher (Rockford, USA). Water was obtained from a Milli-Q water purification system (Millipore, Bedford, MA, USA).

2.2. Sample harvest and preparation

A. thaliana samples were supplied by the Laboratory of Functional Plant Biology (Department of Physiology, Ghent University, Belgium). Two ecotypes of *A. thaliana* named Columbia (Col) and Wassilewskija (Ws) were selected. Plants were grown in a conditioned room at 22 °C and 60% relative humidity, with a photo period of 16 h light and 8 h of darkness. Plants were harvested after 6 weeks. The rosette leaves were used for analysis. Leaves were divided in two groups according to their developmental stage, with the older ones (leaves No. 1–5) and the younger ones (leaves No. 6–10). As a result, four types of samples were available, i.e., Ws younger, Ws older, Col younger and Col older.

Plant samples were prepared according to the method described by Fiehn [12]. After harvesting, plant tissues were immediately frozen in liquid nitrogen in a precooled mortar and crushed into powder. An aliquot of 50 mg tissue sample was transferred into a screw-cap vial. One mL of freshly prepared chloroform:methanol:water (1:2.5:1, v/v/v) was added for extraction. Internal standard myristic-d_{27} acid (5 μL of 3 mg/mL stock solution) was added for normalization [12]. After centrifuging, the supernatant was collected and transferred into a new vial. 400 μL of Milli-Q water was added. The mixture was vortex mixed (30 s) and the vial was centrifuged at 3000 × *g* for 5 min. 500 μL of the upper phase containing the polar compounds was transferred to a high recovery (bottom tapered) vial (Part No. 5183-4497, Agilent Technologies, Wilmington, DE, USA) and dried under nitrogen.

Automated derivatisation was performed on an Agilent 7693A Automatic Liquid Sampler (Agilent Technologies, Wilmington, DE, USA), consisting of two injection towers (rear position equipped with a 250 μL syringe for liquid handling, front position equipped with a 10 μL syringe for injection) and a tray that accommodates vial racks, a vortex position and a temperature controlled heating position. Methoxyamination was performed by adding 70 μL of a 20 mg/mL solution of methoxyamine hydrochloride in pyridine, and heating at 30 °C for 60 min. Next, an aliquot of 70 μL N-methyl-N-(trimethylsilyl) trifluoroacetamide with 1% trimethylchlorosilane (MSTFA + 1% TMCS) was added and heated at 37 °C for 30 min. Injection was performed right after the derivatisation process. Blank samples, only including the internal standard, were processed in the same way. The sample preparation sequence programmed on the autosampler is listed in Supplemental Table 1.

2.3. Retention time locked gas chromatography mass spectrometry (GC–MS)

GC–MS analyses were performed on an Agilent 7890A GC (Agilent Technologies, Wilmington, DE, USA) equipped with a MSD detector (5975C inert XL MSD with Triple-Axis Detector, Agilent). Separation was performed on a DB-5MS 30 m × 0.25 mm × 0.25 μm with a 10 m Duraguard capillary column from Agilent Technologies (Folsom, CA, USA). Injections (1 μL) were performed at 250 °C in split mode (split ratio 10:1). The oven temperature program was 60 °C for 1 min, 10 °C/min to 325 °C (10 min). Helium was used as carrier gas in constant flow mode at a linear velocity of about 35 cm/s (0.92 mL/min, about 50 kPa at 60 °C). Retention time was locked to myristic-d_{27} acid at 16.727 min. These conditions are based on the library developed by Kind et al. [18]. Electron ionization (EI) was used and MS was performed in scan mode (*m*/*z* 50–600) with the MS quadrupole at 150 °C and MS ion source at 250 °C. Data were processed by the MSD ChemStation.

2.4. Data handling

2.4.1. Approach 1: XCMS and principal component analysis (PCA)

GC–MS total ion chromatogram (TIC) data were exported to AIA format data files by ChemStation Software. NetCDF formats data in the AIA format data files were then processed by XCMS software [19,20]. XCMS software, which is freely available under an open-source license at http://metlin.scripps.edu, incorporates nonlinear retention time alignment, matched filtration, peak detection and peak matching. For grouping, bandwidth ("bw") was set to 5 (standard deviation or half width at half maximum). A list with 263 features (ion, retention time, intensity) was obtained, after excluding the variables (features) from blank samples. Normalization was achieved by relative peak areas to the internal standard myristic-d_{27} acid (feature M312T1003, with "*M*" = mass, "*T*" = retention time in s).

A matrix of data with rows (different samples) and columns (features) was used for principal component analysis (PCA) by SIMCA-P V11.0 (Umetrics, Sweden). Pattern recognition based on principal component analysis (PCA) was accomplished after Pareto scaling (Par). In Pareto scaling, the intensity of each variable was scaled by the square root of that variable's standard deviation [21,22].

2.4.2. Approach 2: AMDIS deconvolution, Mass Profiler Professional (peak alignment) and PCA

GC–MS data were processed into ELU and FIN files (two different formats of files) using AMDIS (Automated Mass Spectral Deconvolution and Identification System, NIST). ELU files contain unidentified masses and FIN files have the information of identified masses. Hereby, deconvolution is performed and lists with unidentified features (compounds) and identified features (targets, identified in selected library) are obtained. The ELU and FIN files were applied for further peak alignment (Mass Profiler Professional software) and statistical analysis (PCA).

2.4.3. Identification

The identification of the metabolites in *Arabidopsis* samples was performed by combining mass spectra and database consultation (NIST05, Fiehn Library [18]). Further validation was done using the literature data.

3. Results and discussion

3.1. Critical factors in sample preparation

Success in metabolomics greatly depends on the quality of sample preparation [23]. The variability inevitably introduced by the preprocessing step (extraction, fractionation and/or derivatisation) should be kept to an absolute minimum. Then, it would be able to detect features (metabolites) that are statistically up- or down-regulated between different species or sample groups [24].

During the initial part of this work, we experienced several difficulties when applying described sample preparation methods. Several problems of them were related with the derivatisation

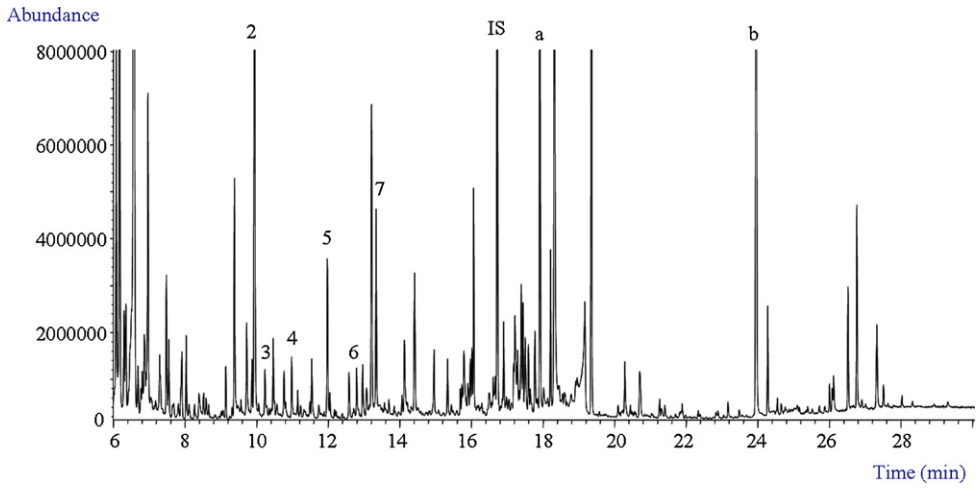

Fig. 1. Total ion chromatogram (TIC) of one *Arabidopsis* sample. Peak label see Table 1. Peak "a" is L-ascorbic acid; peak "b" is sucrose and peak "IS" is internal standard myristic acid-d_{27}.

steps. Silylation, which is the most widely used derivatisation method to block active hydrogen atoms (as in alcohols, carbohydrates, acids, phenols and amines), is prone to artifacts. Several problems were encountered related with incomplete derivatisation or due to the influence of residual silylation agent on the chromatography which could reduce column life.

The quality of reagents and the type of reaction vials used appear to be critical parameters. It was necessary to use fresh reagents for each series of analyses. Different suppliers were tested and the above mentioned sources were found to be reliable and result in the least amount of background signals.

Since residual silylating agents typically degrade column and inlet performance, removal of residual MSTFA was considered. Several approaches were evaluated. The addition of a small amount of silica was tested in order to create a two phase system to remove the silylating agent in a selective and controlled way. Approximately 50 mg silica and 200 μL iso-octane were added to the derivatised sample (10 μL methoxyamine solution + 90 μL MSTFA) and the sample was vortexed. The silica efficiently bound residual MSTFA, but unfortunately, most of the peaks from standard compounds (except derivatised sucrose which survived the procedure), especially test solutes with two active functional groups (e.g. 1,4-butanediamine, 1,4-butanedioic acid, aminocyclopropanecarboxylic acid, etc.) disappeared as well. Polyvinyl alcohol (PVA), butylamine and N,N-diisopropylethylamine (EDIPA) were evaluated as alternatives, but in all cases target metabolites were also removed from the derivatised sample solution and/or extra peaks were generated. Finally, it was decided not to remove the silylating agent. Minimization of the silylating agent was considered and split injection was used.

Initially, the derivatisation reaction was carried out in screw cap reagent tubes. Tests were carried out on several standard mixtures with concentrations varying between 10 μg/mL and 100 μg/mL consisting of amino acids, sugars, fatty acids, etc. In several cases, the measure RSDs (relative standard deviations) of the peak areas for the analysis of these standards were above 15%, which is too high for metabolomic studies considering the values will further increase when analyzing biological material instead of the standards [24]. Additionally sample stability testing revealed that the measure RSDs for the each derivatised standard further increased when the samples were stored (at +5 °C) for more than 24 h. In order to maximize the accuracy of the methodology, automated sample

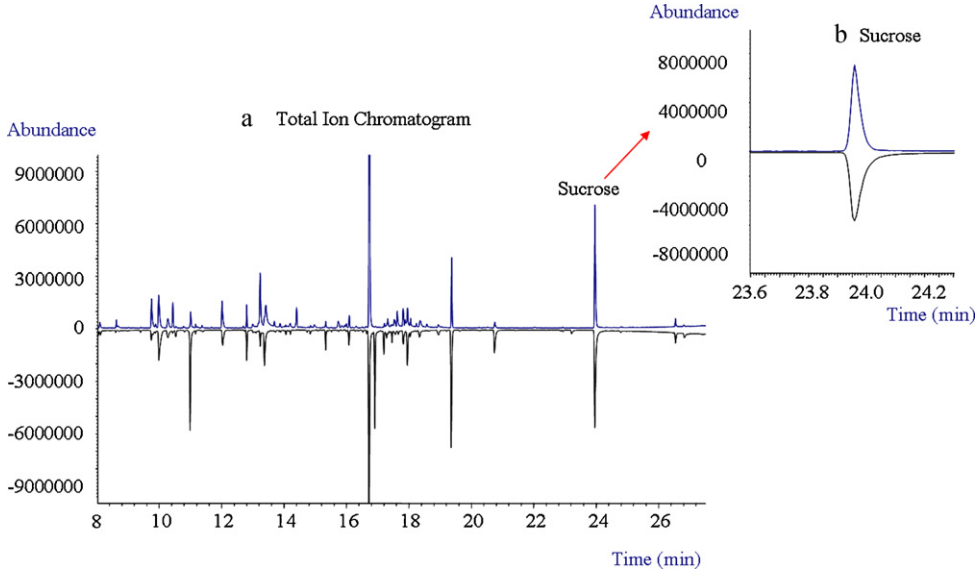

Fig. 2. A mirror image of two chromatograms acquired at different days and after instrument maintenance (liner changed and column changed).

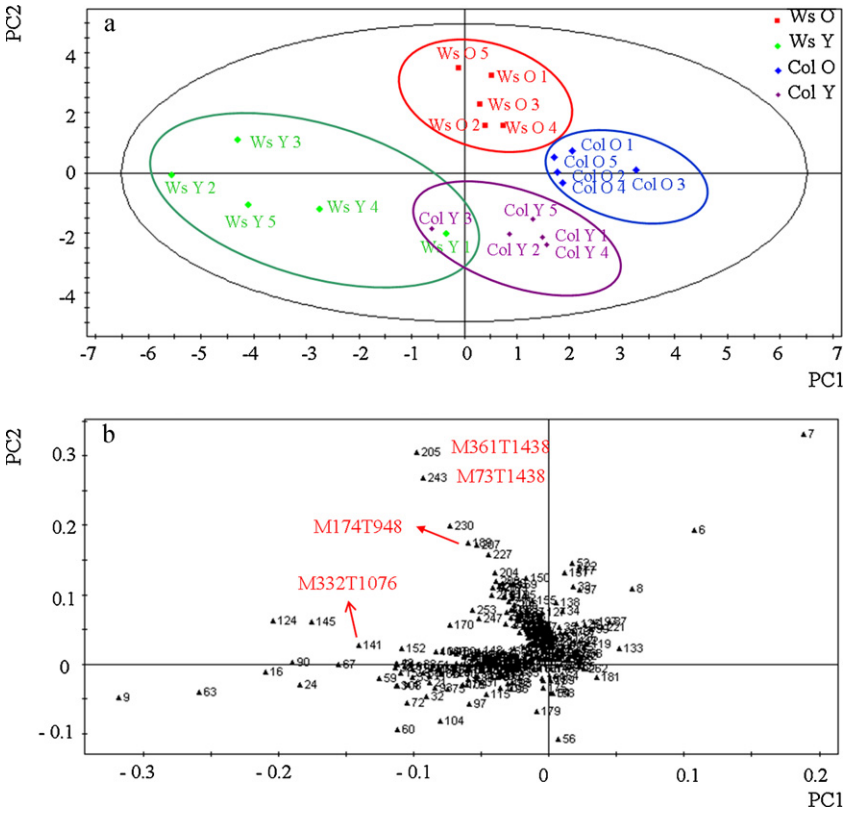

Fig. 3. PCA analysis for different types of *Arabidopsis* samples (approach 1): (a) PCA score plot; (b) PCA loading plot. $A = 3$; $R^2X = 0.772$; $R^2X\,[1] = 0.434$; $R^2X\,[2] = 0.251$.

preparation with a commercially available system was therefore applied. This allowed to minimize the sample preparation time and, more importantly, to keep the time between each derivatisation step and analysis constant.

3.2. Automated sample preparation

A detailed overview of the sample preparation method is described in Supplemental Table 1. The best results in terms of reproducibility were obtained when adding 70 μL of both reagent solution to conically shaped silanized high recovery vials capped with teflon lined septa. The increased reagent volume, together with the conical shape of the vial, allowed for more efficient vortex mixing compared to the original methodology [12]. By using silanized vial walls and the teflon lined septa, side reactions were effectively suppressed leading to a minimal amount of artifact signals in the blank analyses. In order to test the analytical variability of the entire methodology, the reproducibility was measured for the analysis of 6 aliquots of the same extract of an *Arabidopsis* species. The corresponding data are shown in Table 1. The average of the relative standard deviations (RSDs) for the raw peak areas was 6.87%. Surprisingly little or no improvement was observed when calculating the RSD's of the peak areas relative to the area of the internal standard. This is related to the fact that there is a constant time gap between the analysis and the derivatisation procedure. Small changes in the profile, due to, for example, slow loss or inter-exchange of the trimethylsilyl groups between the many species in the solutions, are thereby occurring. This demon-

Table 1
Repeatability of the sample preparation and GC–MS analysis for some selected analytes.

Peak no.	Compound (name)	Retention time (min)	RSD (%, $n = 6$) on retention time	RSD (%, $n = 6$) on area	RSD % on relative area
IS	Myristic acid-d_{27}	16.727	0.01	5.78	0.00
1	2-Hydroxypyridine	6.519	0.04	8.16	3.40
2	Phosphoric acid	9.966	0.05	6.65	8.15
3	L-Threonine 1	10.224	0.04	6.42	8.14
4	Fumaric acid	10.940	0.01	5.33	1.70
5	Aspartic acid 1	12.002	0.03	2.03	4.94
6	D-Malic acid	12.794	0.01	3.10	3.18
7	L-Glutamic acid 1	13.338	0.01	10.64	15.02
8	Purine riboside	21.776	0.03	8.75	11.06
9	Dehydroascorbic acid 1	16.863	0.03	8.96	5.90
10	L-Sorbose 2	17.235	0.01	9.93	5.79
11	D-Mannose 2	17.435	0.01	11.10	13.56
12	Lactulose 1	23.867	0.02	3.76	6.64
13	Allo-inositol	17.245	0.01	5.06	2.92
14	D-(+)Trehalose	24.752	0.02	8.94	6.72
15	Cellobiose 1	24.444	0.02	5.26	7.98
Average				6.87	6.57

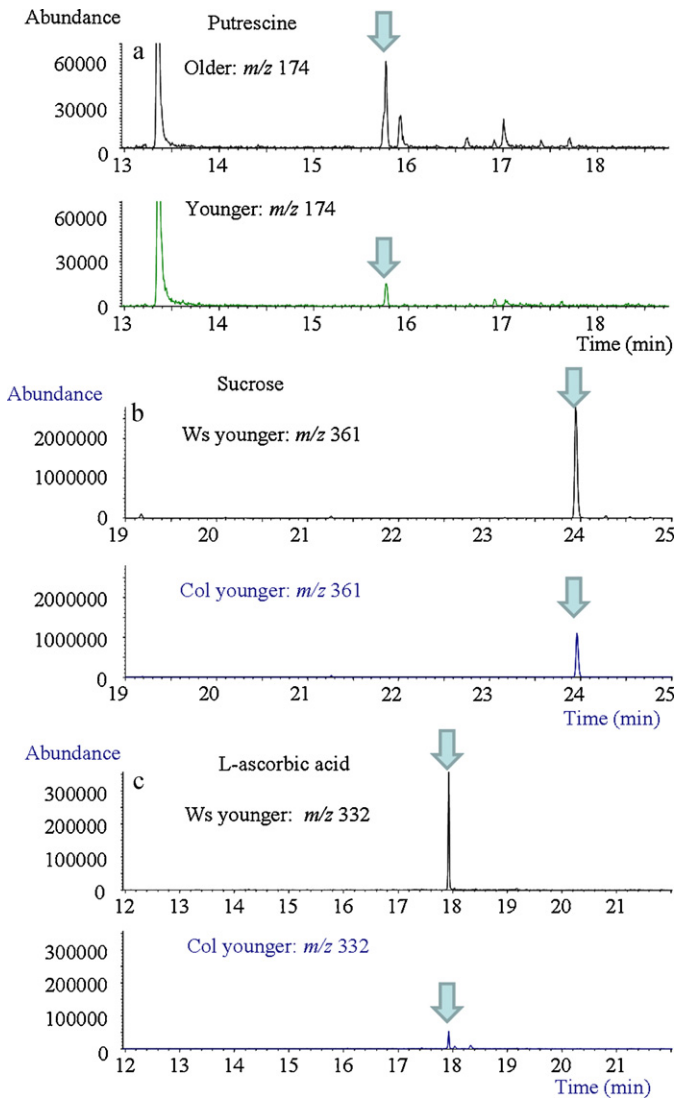

Fig. 4. Extract ion chromatograms: (a) *m/z* 174 of older and younger samples (putrescine); (b) *m/z* 361 of Ws and Col younger samples (sucrose); (c) *m/z* 332 of Ws and Col younger samples (L-ascorbic acid).

strates again that time control of the entire analytical workflow is of critical importance. Note that the analysis of the different *Arabidopsis* samples as discussed below is using the retention time locked method.

3.3. GC–MS

A typical total ion chromatogram obtained for an *Arabidopsis* sample is shown in Fig. 1. The identity of a number of selected compounds and figures of merit of the measurements are given in Table 1. This is relevant information as the quality of feature recognition is highly connected to the minimization of the analytical variability. Fluctuations in retention time and molecular weight measurement could very well blur the distinction between positional isomers often characterized by similar retention times and identical mass spectra. While the MS accuracy on "low" resolution quadrupole instrumentation is nowadays excellent and stable for several months, drifts in retention time can occur if this is not carefully controlled, especially after, e.g. injector maintenance. Retention time locking (RTL), which is a simple procedure involving 5 analyses of a retained compound of choice at various head pressures, allows to re-lock a profile at a set retention time for each peak, in this way greatly improving the reproducibility of retention times [25,26]. The RSDs in Table 1 show the low variability of the latter for 6 consecutive analyses. In Fig. 2, a mirror image of two chromatograms (obtained from different samples) acquired at different days and after instrument maintenance (liner changed, column changed) is shown. Relocking was performed and new samples were analyzed. It is clear that retention time stability is excellent when RTL is performed allowing comparison of data sets acquired on different days.

3.4. Multivariate analysis, metabolite identification and interpretation

3.4.1. Approach 1: XCMS and principal component analysis (PCA)

In total, 20 samples were collected, prepared and analyzed. Five samples were investigated for each type of *A. thaliana* consisting of Columbia "young" and "old" and of Wassilewksija "young" and "old", abbreviated as Col Y, Col O, Ws Y and Ws O, respectively. XCMS was used for peak alignment and 263 features were obtained after excluding the features from the blank samples. Note that several features were obtained for each molecule because the XCMS software allocates each recognized mass signal as a different feature (variable). Each feature is characterized by a mass (*M*) and a time (*T*) in Dalton and seconds, respectively. The software therefore tags the features by a $M_{xxx}T_{yyy}$ code which is for example used in Table 2. To facilitate, e.g. the graphical representations discussed further, each feature is also identified by an ID code. Differences in feature abundance then allow chemometrics treatment of the data. The obtained features were treated with the SIMCA-P Software to construct the PCA plots. These allow a tangible representation of multivariate data in a multidimensional space by making a projection in a two-dimensional plane whereby the differences between the samples are maximized. PCA plots not only reveal groups of samples, trends and outliers with a score plot, but also show the relationships between samples and features in a loading plot [20]. Because it is an unsupervised method, the approach is unbiased. The score plot shown in Fig. 3a was obtained when plotting the scores of the two first principal components (PC1 and PC2) for the 20 samples. It can be easily seen that the classification of these four groups is quite distinct in this case. Not so surprisingly an ecotype differentiation (between Ws and Col) is obtained on the main PC1 axis which explains 43.4% of the difference between the samples. More interestingly the age differentiation between the plants is visible on PC2 (*Y*-axis).

The loading plot (Fig. 3b) allows to interpret the score plot and to present which features are the most influential. These loadings of the principal components unravel the "large or small correlation" and the "positive or negative correlation" in which the measured features contribute to the scores [20]. Features which are grouped together contribute similar information and are correlated in the loading plot. For instance, M361T1438 (ID 205) and M73T1438 (ID 243, both corresponding to sucrose) are positively correlated. When the numerical value of M361T1438 increases, the numerical value of M73T1438 is also increasing. It is an expected behavior for different MS fragments of the same molecule. The distance to the plot's origin is also relevant. Features which are located far away from origin have a stronger impact on the model. For instance, the feature M332T1076 (ID 141) separates Ws type from Col type. The features which mostly mark the differences between the four groups (located far from the origin 0,0 in Fig. 3b) are listed in Table 2.

Taking a closer look at features, M174T948 (ID 189) shows that older leaves have a higher peak of M174T948 than younger ones, as evidenced by the extracted ion chromatograms in the raw data

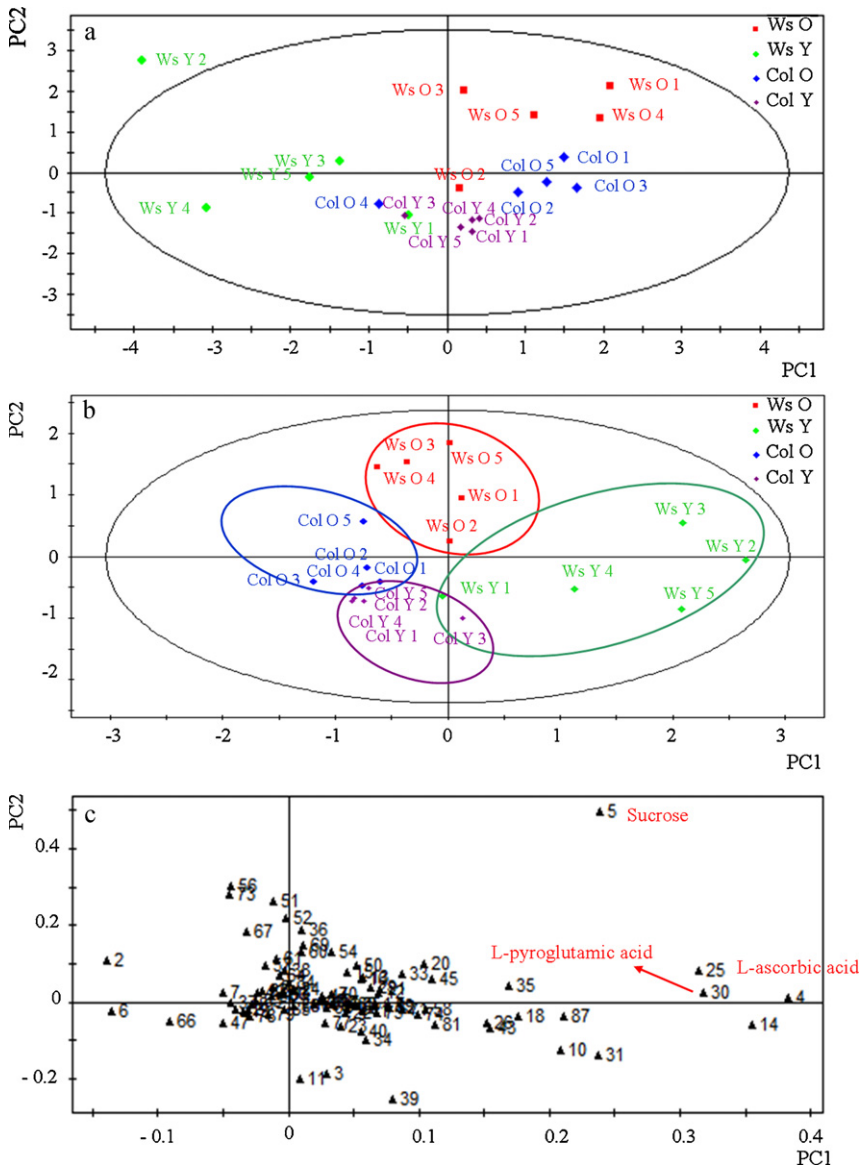

Fig. 5. PCA analysis for different types of *Arabidopsis* samples (approach 2): (a) PCA score plot (ELU files); $A = 2$; $R^2X = 0.368$; $R^2X\,[1] = 0.224$; $R^2X\,[2] = 0.144$. (b) PCA score plot (FIN files); (c) PCA loading plot (FIN files). $A = 3$; $R^2X = 0.609$; $R^2X\,[1] = 0.322$; $R^2X\,[2] = 0.197$.

file (Fig. 4a). This feature was identified as putrescine by the Nist 05 and Fiehn libraries which combine mass spectra with retention index and cover different kinds of metabolites (fatty acids, amino acids, sugars, amines, etc.) [18]. Putrescine is one of the polyamines which are implicated in many biological processes, such as growth, development and abiotic stress responses [27,28]. It can be synthesized in the root system and exported to the leaves. The concentration can be affected by the growth conditions and parameters, such as plant age [29]. As a result, there was more putrescine in older leaves. A similar result was obtained by Friedman et al. [29]. More putrescine was found in exudates of older as compared to younger sunflower plants which indicated a physiological role of polyamines in plant development linked to ageing in plants [29,30].

Features M361T1438, M362T1438, M217T1438, M73T1438 and M147T1438 all were identified as sucrose. Leaves from Ws had indeed a higher level of M361T1438 than Col ones (Fig. 4b). Sucrose concentrations also varied between the younger and older *Arabidopsis* samples. This observation was a good proof of within-plant variation in concentrations of sucrose [31].

The concentration of the feature M332T1076 (ID 141) in the Ws ecotype was higher than in Col (Fig. 4c). The compound was identified as L-ascorbic acid. Ascorbic acid (AA) is not only an important anti-oxidant, but also influences induction of flowering [32]. AA was suggested as an inhibitory effect on the time of floral introduction [33]. For further support, it was approved that elevation of AA content in *Arabidopsis* would lead to a delay in flowering [32]. Here, Ws ecotype has higher concentration of AA. This could be the reason why Ws ecotype induces flowering a bit later than Col.

3.4.2. *Approach 2: AMDIS deconvolution, Mass Profiler Professional (peak alignment) and PCA*

In the second approach, peak alignment of ELU and FIN files was processed by Mass Profiler Professional after AMDIS deconvolution.

For ELU files, totally 673 features were obtained. Entities were filtered based on their frequency values. Those which appeared in more than 50% of samples in at least one condition were chosen. Finally, 231 features were left for further PCA study. Fig. 5a

Table 2
Feature list by two approaches.

Retention time (min)	ID (PCA)	Approach 1	Identification (FIN files)	Approach 2
13.21	145	M156T793	L-Pyroglutamic acid	L-Pyroglutamic acid
13.28	124	M73T797		
13.35	60	M84T801	L-Glutamic acid	
13.35	104	M174T801	L-Glutamic acid	
14.45	90	M140T867		
15.76	189	M174T948	Putrescine	
17.93	141	M332T1076	L-Ascorbic acid	L-Ascorbic acid
19.33	9	M73T1160		
19.35	16	M217T1161		
19.36	67	M305T1162		
19.36	24	M147T1161		
23.96	205	M361T1438	Sucrose	Sucrose
23.96	207	M362T1438	Sucrose	
23.96	230	M217T1438	Sucrose	
23.96	243	M73T1438	Sucrose	
23.96	227	M147T1438	Sucrose	
6.60	6	M173T396		
6.58	7	M115T395		
6.60	8	M116T396		
9.97	63	M299T598		
13.13	56	M140T788		

"*M*" = mass and "T" = retention time (s).

shows the score plot, in which the two principal components only explain 36.8% of the difference between the samples. Compared to approach 1, it is not impressive.

FIN files containing identified information were later tested. Totally 119 compounds were obtained and 88 compounds were left after filtering by frequency. PCA was then used for statistical analysis. PCA score plot with similar classified information as approach 1 was achieved (Fig. 5b). Three principal components explain 60.9% of the difference between the samples. The loading plot (Fig. 5c) presents some mostly influential features for the differences between the four groups. Several features such as L-pyroglutamic acid, L-ascorbic acid and sucrose were exactly the same as approach 1. It was quite meaningful to use the identified features for PCA analysis. This could be useful for target study and also with good confirmation.

4. Conclusions

An automated work flow including automated sample preparation, retention time locked GC–MS analysis and automated data analysis has been described. Good reproducibility was achieved by automated sample preparation. Retention time locked GC–MS data made the peak alignment much easier. Two approaches XCMS followed by principal component analysis (approach 1) and AMDIS deconvolution with Mass Profiler Professional and PCA (approach 2) were compared. XCMS was very useful as a preprocessing method of mass spectrometry data for metabolite profiling. Features obtained by these two approaches were quite comparable. It could be feasible to use both identified and unidentified features for statistical analysis as a better confirmation. The detection of several features that were up- or down regulated in the different species succeeded. In short, this automated whole procedure can be considered as an interesting approach for metabolomic study.

Acknowledgements

The study has been partly supported by the foundation (Nos. 20835006 and 20805045) from National Natural Science Foundation of China. The Unit Plant Hormone Signaling and Bio-imaging was supported by the Fund for Scientific Research-Flanders (grant G.0313.05 to DVDS) and a PhD fellowship to JD from the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT Vlaanderen).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.chroma.2011.01.024.

References

[1] M.M. Koek, F. Bakels, W. Engel, A. van den Maagdenberg, M.D. Ferrari, L. Coulier, T. Hankemeier, Anal. Chem. 82 (2010) 156.
[2] K.K. Pasikanti, P.C. Ho, E.C.Y. Chan, J. Chromatogr. B 871 (2008) 202.
[3] K. Yuan, H. Kong, Y. Guan, J. Yang, G. Xu, J. Chromatogr. B 850 (2007) 236.
[4] S.A. Fancy, O. Beckonert, G. Darbon, W. Yabsley, R. Walley, D. Baker, G.L. Perkins, F.S. Pullen, K. Rumpel, Rapid Commun. Mass Spectrom. 20 (2006) 2271.
[5] P. Jonsson, H. Stenlund, T. Moritz, J. Trygg, M. Sjostrom, E.R. Verheij, J. Lindberg, I. Schuppe-Koistinen, H. Antti, Metabolomics 2 (2006) 135.
[6] M. Vallejo, A. García, J. Tuñón, D. García-Martínez, S. Angulo, J.L. Martin-Ventura, L.M. Blanco-Colio, P. Almeida, J. Egido, C. Barbas, Anal. Bioanal. Chem. 394 (2009) 1517.
[7] L. Torras-Claveria, S. Berkov, O. Jauregui, J. Caujape, F. Viladomat, C. Codina, J. Bastida, Phytochem. Anal. 21 (2010) 80.
[8] O. Fiehn, G. Wohlgemuth, M. Scholz, T. Kind, D.Y. Lee, Y. Lu, S. Moon, B. Nikolau, Plant J. 53 (2008) 691.
[9] K. Schneider, J.O. Krömer, C. Wittmann, I. Alves-Rodrigues, A. Meyerhans, J. Diez, E. Heinzle, Microb. Cell. Fact. 8 (2009) 12.
[10] S. Mas, S.G. Villas-Boas, M.E. Hansen, M. Akesson, J. Nielsen, Biotechnol. Bioeng. 96 (2007) 1014.
[11] O. Fiehn, J. Kopka, R.N. Trethewey, L. Willmitzer, Anal. Chem. 72 (2000) 3573.
[12] O. Fiehn, in: J. Salinas, J.J. Sanchez-Serrano (Eds.), Arabidopsis Protocols, Humana Press, 2006, p. 439.
[13] B.C. Grande, M.S.G. Falcon, M.R. Comesana, J.S. Gándara, J. Agric. Food Chem. 49 (2001) 3145.
[14] R.M. González-Rodríguez, R. Rial-Otero, B. Cancho-Grande, J. Simal-Gándara, J. Chromatogr. A 1196–1197 (2008) 100.
[15] M. Pateiro-Mourea, E. Martinez-Carballo, M. Arias-Estevez, J. Simal-Gándara, J. Chromatogr. A 1196–1197 (2008) 110.
[16] L. Rey-Salgueiro, X. Pontevedra-Pombal, M. Álvarez-Casas, E. Martínez-Carballo, M.S. García-Falcón, J. Simal-Gándara, J. Chromatogr. A 1216 (2009) 5235.
[17] R.M. González-Rodríguez, B. Cancho-Grande, J. Simal-Gándara, J. Chromatogr. A 1216 (2009) 6033.
[18] T. Kind, G. Wohlgemuth, D.Y. Lee, Y. Lu, M. Palazoglu, S. Shahbaz, O. Fiehn, Anal. Chem. 81 (2009) 10038.
[19] P. Gao, C. Lu, F. Zhang, P. Sang, D. Yang, X. Li, H. Kong, P. Yin, J. Tian, X. Lu, A. Lu, G. Xu, Analyst 133 (2008) 1214.
[20] C.A. Smith, E.J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, Anal. Chem. 78 (2006) 779.
[21] L. Eriksson, E. Johansson, N. Kettaneh-Wold, S. Wold, Multi-and Megavariate Data Analysis: Principles and Applications, Umetrics, 2001.
[22] H.M. Parsons, C. Ludwig, U.L. Günther, M.R. Viant, BMC Bioinform. 8 (2007) 234.

[23] M.H. Shin, D.Y. Lee, K.-H. Liu, O. Fiehn, K.H. Kim, Anal. Chem. 82 (2010) 6660.
[24] S.-A. Fancy, K. Rumpel, in: F. Wang (Ed.), Methods in Pharmacology and Toxicology, Humana Press INC, 2008, p. 317.
[25] N. Etxebarria, O. Zuloaga, M. Olivares, L.J. Bartolomé, P. Navarro, J. Chromatogr. A 1216 (2009) 1624.
[26] L.M. Blumberg, S.K. Matthew, Anal. Chem. 70 (1998) 3828.
[27] P.T. Evans, R.L. Malmberg, Ann. Rev. Plant Physiol. Plant Mol. Biol. 40 (1989) 235.
[28] J.M. Tanguy, Plant Growth Regul. 34 (2001) 135.
[29] R. Friedman, N. Levin, A. Altman, Plant Physiol. 82 (1986) 1154.
[30] N. Bagni, R. Pistocchi, in: R.D. Slocum, H.E. Flores (Eds.), Biochemistry and Physiology of Polyamines in Plants, Crc Press, 1991, p. 105.
[31] S.Z. Merritt, J. Chem. Ecol. 22 (1996) 1133.
[32] C. Barth, M.D. Tullio, P.L. Conklin, J. Exp. Bot. 57 (2006) 1657.
[33] P.L. Conklin, C. Barth, Plant Cell Environ. 27 (2004) 959.