Anal Bioanal Chem
DOI 10.1007/s00216-015-9042-2

RESEARCH PAPER

Evaluation of changes induced in rice metabolome by Cd and Cu exposure using LC-MS with XCMS and MCR-ALS data analysis strategies

Meritxell Navarro-Reig[1] · Joaquim Jaumot[1] · Alejandro García-Reiriz[1,2] · Romà Tauler[1]

Received: 30 July 2015 / Revised: 7 September 2015 / Accepted: 10 September 2015

Abstract The comprehensive analysis of untargeted metabolomics data acquired using LC-MS is still a major challenge. Different data analysis tools have been developed in recent years such as XCMS (various forms (X) of chromatography mass spectrometry) and multivariate curve resolution alternating least squares (MCR-ALS)-based strategies. In this work, metabolites extracted from rice tissues cultivated in an environmental test chamber were subjected to untargeted full-scan LC-MS analysis, and the obtained data sets were analyzed using XCMS and MCR-ALS. These approaches were compared in the investigation of the effects of copper and cadmium exposure on rice tissue (roots and aerial parts) samples. Both methods give, as a result of their application, the whole set of resolved elution and spectra profiles of the extracted metabolites in control and metal-treated samples, as well as the values of their corresponding chromatographic peak areas. The effects caused by the two considered metals on rice samples were assessed by further chemometric analysis and statistical evaluation of these peak area values. Results showed that there was a statistically significant interaction between the considered factors (type of metal of treatment and tissue). Also, the discrimination of the samples according to both factors was possible. A tentative identification of the most discriminant metabolites (biomarkers) was assessed. It is finally concluded that both XCMS- and MCR-ALS-based strategies provided similar results in all the considered cases despite the completely different approaches used by these two methods in the chromatographic peak resolution and detection strategies. Finally, advantages and disadvantages of using these two methods are discussed.

Keywords Metabolomics · LC-MS · XCMS · MCR-ALS · Cu · Cd · Rice · Sample discrimination

Introduction

Metabolomics can be defined as the exhaustive profiling study of all metabolites contained in an organism. It is known that external perturbations imposed on organisms can produce changes in their metabolome. These perturbations can be environmental changes; physical, abiotic, or nutritional stresses; mutation; and transgenic events [1–3]. Therefore, metabolomics is a powerful approach to study molecular mechanisms and metabolic pathways implicated in the response to different perturbations and in the organism defense strategies against them. Over the last decade, data processing has been a challenge in untargeted metabolomics due to the extreme complexity of the experimental data sets, especially in the case of combining a MS detector with chromatographic techniques such as LC or GC. As a consequence, software programs for automated processing of data have been introduced, such as MetAlign [4], MZmine [5], or XCMS [6], among others.

In the last years, XCMS has become a favorite method among the metabolomic community for feature detection,

Electronic supplementary material The online version of this article (doi:10.1007/s00216-015-9042-2) contains supplementary material, which is available to authorized users.

✉ Joaquim Jaumot
joaquim.jaumot@idaea.csic.es

[1] Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, 08034 Barcelona, Spain

[2] Departamento de Química Analítica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Instituto de Química Rosario (IQUIR-CONICET), Suipacha 531, Rosario S2002LRK, Argentina

Published online: 24 September 2015

and it has been used for a broad range of applications. In brief, XCMS is a tool dedicated to chromatographic feature detection which includes automatic processing of huge size full-scan LC-MS data and estimates candidate metabolites by using peak detection and retention time correction algorithms and methods. For each proposed candidate, XCMS gives *p* value (statistical test comparing the integrated peak areas of this candidate in control versus treated samples) and fold change (defined as the ratio of the integrated peak areas of the treated samples versus the control samples) [7, 8].

MCR-ALS is also a popular chemometric method used for the resolution of pure contributions in unresolved mixtures [9]. MCR-ALS is used in a wide variety of applications as, for instance, the resolution of overlapped chromatographic peaks in environmental samples. MCR-ALS has been recently proposed as an alternative approach to detect potential biomarkers in untargeted metabolomics studies [10]. MCR-ALS decomposes the experimental LC-MS data matrix into their factor contributions which can be assigned to the chromatographic elution profiles and to the mass spectra of each resolved component. The main difference between these two approaches lies in peak detection and resolution. While XCMS identifies each feature characterized by its retention time and a unique *m*/*z* value, MCR-ALS resolves mathematical components characterized by their elution profiles and mass spectra (with more than one possible MS feature assigned to the same elution profile) [6, 10]. With the aim of comparing these two approaches, in the present work, the same metabolomic data set was processed by means of XCMS and by MCR-ALS, and further evaluated by using other chemometric methods for exploration and discrimination purposes. The proposed untargeted metabolomic approach has been used to assess the effects of cadmium and copper treatment on Japanese rice.

Plants are complex organisms exposed to a set of abiotic and biotic stresses [11]. One of these abiotic stresses is the pollution by toxic metals present in the environment. These metals can be found as constituents of the Earth's crust and geological processes, but human activities, such as mining, agriculture, and a wide range of industrial activities, can drastically alter their geochemical cycles and distribution on earth surface [12, 13]. These anthropogenic activities caused that the level of some of these toxic metals in the environment increased notably in recent years. Although the discovery of adverse health effects resulting from toxic metals has caused the decrease of emissions in most of the developed countries during the last century, there are still some metals like cadmium, whose emissions increased during the twentieth century, due to its large industrial use and reduced recycling [14]. Among toxic metals, cadmium and copper have been listed on the priority list of hazardous materials by the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) in 2013 [13]. These two pollutants are readily absorbed by roots and rapidly translocate to the aerial parts of plants [15]. Since diet is the primary source of exposure to these metals for the general population, intensive research has been performed on the accumulation of these pollutants in edible plants [13]. In this work, Japanese rice (*Oryza sativa japonica* Nipponbare) has been used as a target organism because it is one of the model organisms frequently used in plant metabolomics and is also an edible plant [1, 3, 13].

The metabolomic study presented in this work considers two categorical factors related to the metal exposure: rice tissue sample analyzed (root or aerial part) and metal of treatment (Cd or Cu). Metabolomic datasets commonly use statistical experimental designs, where different dose groups, multiple time points, diverse sample groups, or various subjects are simultaneously investigated [16, 17]. For this reason, comprehensive data analysis methods able to deal with this type of complex designs are required. In this work, statistical evaluation of the different investigated effects has been performed using different multivariate data analysis methods such as ANOVA-simultaneous component analysis (ASCA), principal component analysis (PCA), and partial least-squares discriminant analysis (PLS-DA).

Experimental

Reagents

Cadmium chloride hydrate (≥98.0 %), copper(II) sulfate pentahydrate (≥98.0 %), and ammonium acetate (≥98.0 %) were from Sigma-Aldrich (Steinheim, Germany). HPLC-grade water, acetonitrile (≥99.8 %), and methanol (≥99.8 %) were supplied by Merck (Darmstadt, Germany). Chloroform was obtained from Carlo Erba (Peypin, France). Piperazine-*N*, *N*′-bis(2-ethanesulfonic acid) (PIPES) (≥99.0 %) was used as internal standard (Sigma-Aldrich, Steinheim, Germany).

Solutions containing 10, 50, and 100 μM of cadmium (Cd) and copper (Cu) were prepared weekly by diluting a 1000 μM stock solution of these metals. Stock solutions were prepared weekly by dissolution of the appropriate amounts of cadmium chloride hydrate and copper(II) sulfate salts. All the solutions were stored at 6 °C until their use.

Water used for plant watering, for preparing cadmium and copper solutions, and during the extraction procedure was purified using an Elix 3 coupled to a Milli-Q system (Millipore, Belford, MA, USA), and filtered through a 0.22-μm nylon filter integrated into the Milli-Q system.

Plant growth, stress treatment, and metabolite extraction

Oryza sativa japonica Nipponbare seeds, obtained from the Center for Research in Agricultural Genomics (CRAG) at Autonomous University of Barcelona, were incubated for

2 days at 30 °C in a wet environment. After this period, seeds were planted in 3.0×3.0 cm individual pots and grown on an Environmental Test Chamber MLR-352H (Panasonic®) for 22 days under white fluorescent light. Temperature, relative humidity, and light long-day conditions at the chamber were set as described in the Electronic supplementary material (ESM) Table S1. During the first 10 days of growth, rice plants were watered with Milli-Q water three times a week. Since then, plant treated samples were subjected to irrigation water containing different concentrations of Cd and Cu, whereas the plant control samples were watered with milli-Q water until harvest. Metal concentrations used for stressing rice plants were 10, 50, and 1000 μM, and for every concentration, two trays containing 18 pots were used. The lower concentration was set to 10 μM, in agreement with the lowest reported metal concentration producing noticeable changes in plants [13, 18, 19], and the higher concentration was set to 1000 μM because it is the highest metal concentration inducing changes in plants without causing their death [13, 18, 19]. In order to avoid differences in the growth of individual plants, the position of the trays inside the chamber was changed daily following a random design, and the volume of irrigation water was controlled and set at 200 mL per tray. After harvest, roots and aerial part were separated and, immediately, metabolism was quenched by freezing at liquid nitrogen temperature. Samples were stored at −80 °C until extraction.

Before extraction, aerial parts and roots were ground under liquid nitrogen to a fine powder and lyophilized overnight until dryness. Metabolite extraction was carried out by dispersing 40 mg of the dried tissue in 1 mL of MeOH in a 2.0-mL Eppendorf tube. Then, the mixture was vortexed for 1 min and sonicated for 10 min; this step was repeated twice. After centrifuging for 20 min at 14,100×*g*, a 750-μL aliquot of the supernatant was transferred to a 1.5-mL Eppendorf tube. Then, 500 μL of chloroform and 400 μL of water were added. After that, the mixture was vortexed for 1 min, incubated for 15 min at −4 °C, and centrifuged for 20 min at 14,100×*g*. Finally, a 750-μL aliquot of aqueous fraction was transferred to a 1.5-mL Eppendorf tube, evaporated to dryness under nitrogen gas, and reconstituted with 450 μL of acetonitrile/water (1:1 *v*/*v*). For internal standard quantification, 50 μL of 50 mg/L solution of the internal standard (PIPES) was added to the extract. For each tray, two replicates were done. All of the extracts were stored at −80 °C until analyzed and were filtered through 0.2-μm nylon filters before injection (Pall Life Sciences, Port Washington, NY, USA).

HPLC-MS analysis

Chromatographic separation was performed on an Acquity UHPLC system (Waters, Milford USA), equipped with a quaternary pump, an autosampler, and a column oven. An HILIC TSK gel Amide-80 column (250×2.0 mm^2 i.d., 5 μm) with a 2.0 mm×1 cm i.d. guard column of the same material provided by Tosoh Bioscience (Tokyo, Japan) was used for analytical separation of metabolites. Elution gradient was performed using solvent A (acetonitrile) and solvent B (ammonium acetate 3 mM at pH 5.5, adjusted with acetic acid) as follows: 0–3 min, isocratic gradient at 5 % B; 3–27 min, linear gradient from 5 to 70 % B; 27–30 min, isocratic gradient at 70 % B; 30–32 min back to the initial conditions at 5 % B; and from 32 to 40 min, at 5 % B. The mobile phase flow rate was 0.15 mL/min and the injection volume was 5 μL.

The mass spectrometer was an LCT Premier XE-time-of-flight (TOF) analyzer (Waters, Milford USA) equipped with an electrospray (ESI) as ionization source in negative and positive modes. Nitrogen (purity >99.98 %) was used as cone and desolvation gas at flow rates of 50 and 600 L/h, respectively. Desolvation temperature was set to 350 °C, and electrospray voltages were set to 3.0 kV (positive mode) and to 2.2 kV (negative mode). The mass acquisition range was 90–1000 *m*/*z*.

Data analysis

Waters raw chromatographic data files (.raw format) were converted to the standard CDF format by the Databridge function of MassLynx™ v 4.1 software (Waters, USA).

These data files were then imported into the MATLAB environment (release 2014b; The Mathworks Inc., Natick, MA, USA) by using the MATLAB Bioinformatics Toolbox (version 4.3.1) and in-house built routines. Finally, every LC-MS-analyzed rice sample gave a data matrix containing the acquired retention times on the rows and the detected *m*/*z* values on the columns. In order to facilitate calculations, the total number of columns (i.e., *m*/*z* values) was reduced by using a binning approach (grouping mass values into a number of bins within a particular *m*/*z* range, in this case 0.05 amu). Every analyzed sample gave a data matrix with 1020 rows (retention time from 0 to 40 min) and 18,200 columns (from 90 to 1000 amu at 0.05 resolution). In the case of XCMS, raw chromatographic data files in CDF format were directly imported into MetaboNexus bioinformatics platform [20] without applying the binning approach.

Peak areas analysis

Two different methodologies were used and compared for the calculation of chromatographic peak areas: XCMS and MCR-ALS. In order to ascertain the effect of the treatment with the two metals, chromatographic peak areas obtained using any of these two methods were analyzed using PCA, PLS-DA, and ASCA. Before applying these chemometric methods, peak areas were autoscaled (mean-centered and scaled by their

standard deviation) to give equal weight (scale) to each one of the detected features.

XCMS XCMS approach allows an automatic processing of data for feature detection and calculation of chromatographic peak areas [6]. A typical XCMS analysis starts with the application of the *centWave* data processing algorithm which basically consists of two main steps. First, dominant mass spectra features are identified in this domain by using the so-called regions of interest (ROIs). In these identified ROIs, the presence of a chromatographic peak is denoted by a signal which at a particular *m*/*z* value has intensity over a particular preselected threshold value. The second step is the identification and modeling of chromatographic peaks by means of a wavelet transformation and a Gaussian shape curve fitting approach. Then, non-relevant features are dismissed by considering only those that are present in more than a certain percentage of all the samples (commonly 50 %). Finally, chromatographic peaks of the same component in different samples are aligned by means, for instance, of the *obiwarp* algorithm [21]. For more detailed information about the XCMS algorithm see the work of Smith [6] and Tautenhahn [22].

In this work, MetaboNexus bioinformatics platform [20] has been used to import and pre-process raw chromatographic data files in CDF format. MetaboNexus pre-processing platform relies on the XCMS package in R language environment and it provides a dashboard of controls to handle pre-processing in an intuitive manner with available pre-sets for different instruments. In this work, the settings were manually adjusted starting with the pre-sets corresponding to an HPLC/Q-TOF analyzer. The optimization of these parameters in our particular case was not straightforward. For this reason, the full analysis was repeated using different combinations of the parameters with variations from the default settings, and the results were compared to decide which the best parameters were. Finally, the *centWave* algorithm was employed as a feature detection method using 30 ppm as the maximal tolerated *m*/*z* deviation in consecutive scans, and allowing chromatographic peak widths ranging from 10 to 60 s. The number of peaks across samples of intensity higher than 1000 was fixed to 5, and the signal-to-noise threshold was set to 10. Peak integration was carried out using a Mexican hat approach considering a minimum difference in *m*/*z* for peaks with overlapping retention time of −0.0025. Regarding chromatographic peak alignment, the *obiwarp* algorithm [21] was selected for retention time correction. Grouping parameters were set to 5 s for the bandwidth of Gaussian smoothing kernel to apply to the peak density chromatogram whereas the width of overlapping *m*/*z* slices to use for creating peak density chromatograms and grouping peaks across samples was set to 0.025 amu. Finally, the minimum percentage of samples at where the same peaks need to be present in at least one sample class was set to 70 %.

The final output is a data table that contains the selected features (identified by their exact *m*/*z* values) in the rows, and the area of these features for each sample in the columns. Finally, sample areas were normalized by using the area of the PIPES internal standard.

Multivariate curve resolution by alternating least squares MCR-ALS is a chemometric method used for the resolution of pure contributions in unresolved mixtures [9]. MCR-ALS can be used to resolve a wide variety of datasets from different research fields, like hyphenated and multidimensional chromatographic systems, -omics data, process analysis, spectroscopic images, environmental data tables, etc., as it has been already described in the literature [23–25].

In this work, MCR-ALS has been used to resolve the elution and mass spectra profiles of the metabolites obtained in the full-scan untargeted LC-MS analysis of the rice sample extracts before and after metal treatment. MCR-ALS decomposes every individual experimental dataset arranged in a data matrix according to the following bilinear model:

$$\mathbf{D} = \mathbf{C}\mathbf{S}^{\mathbf{T}} + \mathbf{E} \quad (1)$$

Where **D** (size *IxJ*) represents the experimental LC-MS data matrix (from a single rice sample) in which the rows are the MS spectra at all retention times (i=1,…I), and the columns are the chromatograms at all *m*/*z* channels (j=1,…J). According to Eq. 1, **D** matrix is decomposed into the product of two factor matrices, **C** and $\mathbf{S}^{\mathbf{T}}$, that corresponds respectively to the matrix of the resolved elution profiles, **C** (size *IxN*), and to the matrix of their corresponding mass spectra, $\mathbf{S}^{\mathbf{T}}$ (size *NxJ*). *N* represents the total number of resolved components considering during MCR-ALS analysis. **E** matrix (size *IxJ*) contains the residuals not explained by the model using the *N* considered components.

This data analysis strategy can be easily extended to the simultaneous analysis of several samples. For instance, in the case of this work, a total number of 128 samples have been simultaneously considered: 16 aerial part rice control samples, 24 aerial part rice samples treated with Cd at different concentrations (10, 50, and 1000 μM), 24 aerial part rice samples treated with Cu at different concentrations (10, 50, and 1000 μM), 16 root control rice samples, 24 root rice samples treated with Cd at different concentrations (10, 50, and 1000 μM), and 24 root rice samples treated with Cu at different concentrations (10, 50, and 1000 μM). All these samples were arranged in a single column-wise augmented data matrix ($\mathbf{D_{aug}}$), containing the 128 individual data matrices ($\mathbf{D_x}$ where x=1,…128), one for each rice sample, settled one on the top of

the other. This long column-wise augmented matrix ($\mathbf{D_{aug}}$) is also decomposed using a bilinear model such as:

$$\mathbf{D_{aug}} = \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \\ \mathbf{D}_3 \\ \vdots \\ \mathbf{D}_{128} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \\ \mathbf{C}_3 \\ \vdots \\ \mathbf{C}_{128} \end{bmatrix} \mathbf{S}^{\mathbf{T}} + \begin{bmatrix} \mathbf{E}_1 \\ \mathbf{E}_2 \\ \mathbf{E}_3 \\ \vdots \\ \mathbf{E}_{128} \end{bmatrix} = \mathbf{C_{aug}}\mathbf{S}^{\mathbf{T}} + \mathbf{E_{aug}} \quad (2)$$

Before applying MCR-ALS, every individual data matrix from every sample ($\mathbf{D_x}$) was normalized dividing by the chromatographic peak area of the internal standard (PIPES) in the considered sample. In order to accelerate and reduce memory requirements of MCR-ALS calculations, every chromatogram of each sample was divided into six separate chromatographic time windows. In this way, a final number of six column-wise augmented data matrices, each one corresponding to one of the six chromatographic time windows, were independently analyzed by MCR-ALS.

MCR-ALS solves Eq. 1 ($\mathbf{D}$, single data matrix case) or Eq. 2 ($\mathbf{D_{aug}}$, augmented data matrix case) starting with an initial guess of the number of components needed to explain sufficiently well the considered matrix by its singular value decomposition (SVD) [26]. This number should be large enough to include most of the metabolites extracted from the rice samples giving a significant MS chromatographic signal and, also, to consider possible background and solvent signal contributions. Next, an initial estimate of either $\mathbf{C}$ or $\mathbf{S^T}$ factor matrices should be provided. For instance, this estimate can be obtained from the purest MS signals in the data set gathered by a variable detection method such as SIMPLISMA [27]. Then, the estimation of $\mathbf{C}$ or $\mathbf{S^T}$ factor matrices is performed by means of an alternating least squares optimization under preselected constraints. In the case of this work, the applied constraints were only non-negativity (elution and spectra profiles) and spectral normalization (equal height) [28]. Resolved mass spectra in $\mathbf{S^T}$ are enforced to be the same for the common constituents in the different analyzed samples, whereas elution profiles resolved in $\mathbf{C_{aug}}$ matrix are allowed to be different for each one of the samples ($\mathbf{C_x}$ (x=1,…128)), as shown in Eq. 2. It is important to point out here that in the MCR-ALS approach, there is no need to correct for the unavoidable changes in the elution profiles of the same metabolite in different samples and chromatographic runs, for instance in their retention times (peak shifting) and in their profile shapes, differently to what occurs with XCMS (see above). This is a clear advantage of MCR-LS compared to XCMS, as it will be discussed later because it fits better with the natural behavior of LC-MS data. It is also important to emphasize that in this case of LC-MS data, MCR-ALS results will be not affected by uncertainties due to rotational ambiguities since MS spectral resolution is very high and also due to the large number of simultaneously analyzed individual rice sample data matrices (up to 128), which gives very robust results. Finally, from MCR-ALS results, a data table that contains the resolved components in the rows and the peak area of these components for each sample in the columns is obtained.

Discrimination and evaluation of metal effects on rice samples by chemometric analysis of metabolite concentration (peak areas) changes

Once metabolite relative concentration changes were estimated by XCMS and MCR-ALS, various methods were used to evaluate them and to discriminate samples according to metal treatment. PCA [29], PLS-DA [30], and ASCA [31] were used for this purpose and are briefly described in the ESM.

In this work, PCA has been used to explore the behavior of tissue (aerial part or root) samples when they were treated either with Cd or Cu metals. PLS-DA has been used to discriminate between root and aerial parts of rice samples, and between samples treated either by Cd or Cu (Venetian blind cross-validation was used to assess the results of the PLS-DA model). Finally, ASCA analysis (performed on a well-balanced experimental design) allowed the interpretation of the sources of the experimental variance: the tissue sample (root or aerial part) and the metal of treatment (Cd or Cu). Possible interactions between these two factors have been also evaluated. In order to assess the statistical significance of the considered factors, a permutation test can be done. In this work, the number of permutations for each model was set to 10,000.

Chemometric software

PCA, PLS-DA, and ASCA were performed by using PLS Toolbox 7.8 (Eigenvector Research Inc., Wenatche, WA, USA) working under MATLAB (The Mathworks, Natick, MA, USA). MCR-ALS was carried out using MCR-ALS toolbox freely available at www.mcrals.info. XCMS was performed using MetaboNexus interactive data analysis platform [20].

Results and discussion

Two categorical factors (tissue sample and metal of treatment) were considered. First, the statistical significance of these two factors was evaluated by means of ASCA analysis of LC-MS peak areas from MCR-ALS resolved components and from XCMS detected features. In addition, the possibility of differentiating samples according to the experimental factors was studied using an exploratory analysis of LC-MS peak areas

using PCA and PLS-DA. Table 1 shows a summary of the data matrices considered in this work.

Assessment of the effects of experimental factors

As is described below, in all the cases studied here, XCMS and MCR-ALS data analysis gave comparable results. Furthermore, results confirmed that the treatment with cadmium and copper had a significant effect on all treated samples. XCMS was applied to the entire full-scan chromatograms, and the output was a table containing 1627 features. Not all of these detected features could be considered an independent metabolite since some of them could be adducts or isotopic masses coming from the same metabolite, and others could be assigned to electric signals or other noise contributions. This table was used to build $\mathbf{D}_{\mathbf{xcms\_areas}}$ matrix (see Table 1). This number of XCMS features detected could be reduced by using additional applications such as, for instance, CAMERA [32]. This software facilitates compound annotation and identification by considering adducts and isotopic peaks which reduce the final number of features detected by XCMS. However, in this work, this was not used since this selection was performed by chemometric analysis (see below). MetaboNexus allowed a fast analysis of the data set. The entire full-scan chromatograms were resolved in approximately 10 min. However, the bottleneck of XCMS approach was the selection of the optimal pre-processing parameters that required testing several combinations of the feature detection and peak alignment parameters, which increase the total time of analysis considerably.

MCR-ALS was used to analyze the six column-wise augmented data matrices described in the methods section. Figure 1 shows a TIC chromatogram of an aerial part sample with the selected chromatographic regions highlighted. Between 25 and 30 MCR-ALS components were resolved for each column-wise augmented data matrix with a minimum explained variance of 98 % (an example of an SVD used to select the number of components considered in the resolution of each window is shown in ESM Fig. S1). This selection considered a sufficiently large number of components to include all significant metabolite contributions and other interfering MS signal contributions (instrumental background, solvent, etc.) that could appear among the resolved components. A total number of 165 MCR-ALS components (resolved peaks) were needed to explain sufficiently the data variance. In this case, 29 from the 165 resolved components were associated with noisy signals, like background and solvent contributions, and hence, they were ignored in subsequent analyses. Finally, 136 components were used to build the $\mathbf{D}_{\mathbf{mcr\_areas}}$ matrix (see Table 1).

Figure 2 shows the final results for one of these 136 components resolved in the simultaneous analysis of 32 aerial part rice samples, 24 treated with copper at three concentrations (10, 50, and 1000 μM), and 8 control samples (not treated with copper). This example shows a component resolved in the second chromatographic region, from 5.1 to 7.1 min. A total of 30 components were resolved in this particular interval. Elution profile of this resolved component in the different aerial part samples is depicted in Fig. 2a, where a clear decrease is observed in the height and area of the resolved chromatographic peaks in this profile between control and Cu-treated samples. However, no clear distinction is observed between the heights or areas of these peaks for the different levels of metal, probably reflecting that at the lowest concentration of the metal tested in this study, the profile of this particular metabolite is already significantly affected, without any further change of it at higher Cu concentration. Figure 2b shows the corresponding resolved mass spectrum for this component. A high intense mass signal was found at *m*/*z* 515.10 with two lower intensity signals (see inset) at *m*/*z* 516.10 and 517.10, which are isotopic contributions. Some other small signals at different *m*/*z* values are present also in this spectrum (see other insets). They may be possible adducts of the main component, or also they can be very minor

Table 1 Design of experimental data matrices

Matrix	Size[a]	Description	Uses
$\mathbf{D}_{\mathbf{mcr\_areas}}$	128×136	Full data set	Evaluation of the effect of tissue sample, of metal of treatment and of their interaction by ASCA
$\mathbf{D}_{\mathbf{xcms\_area}}$	128×1627		
$\mathbf{Cd}_{\mathbf{mcr\_areas}}$	64×136	Roots and aerial part samples treated with Cd	Evaluation of the effect of tissue sample by PCA and PLS-DA
$\mathbf{Cd}_{\mathbf{xcms\_areas}}$	64×1627		
$\mathbf{Cu}_{\mathbf{mcr\_areas}}$	64×136	Roots and aerial part samples treated with Cu	
$\mathbf{Cu}_{\mathbf{xcms\_areas}}$	64×1627		
$\mathbf{L}_{\mathbf{mcr\_areas}}$	48×136	Aerial part samples treated with Cd or Cu	Evaluation of the effect of metal of treatment by PCA and PLS-DA
$\mathbf{L}_{\mathbf{xcms\_areas}}$	48×1627		
$\mathbf{R}_{\mathbf{mcr\_areas}}$	48×136	Roots samples treated with Cd or Cu	
$\mathbf{R}_{\mathbf{xcms\_areas}}$	48×1627		

Subscripts indicate the kind of data: *-mcr_areas* is the data matrix containing areas of the MCR-ALS resolved components; *-xcms_areas* is the data matrix containing peak areas obtained by XCMS

[a] Size: number of samples×136 MCR-ALS resolved components or 1627 XCMS detected features

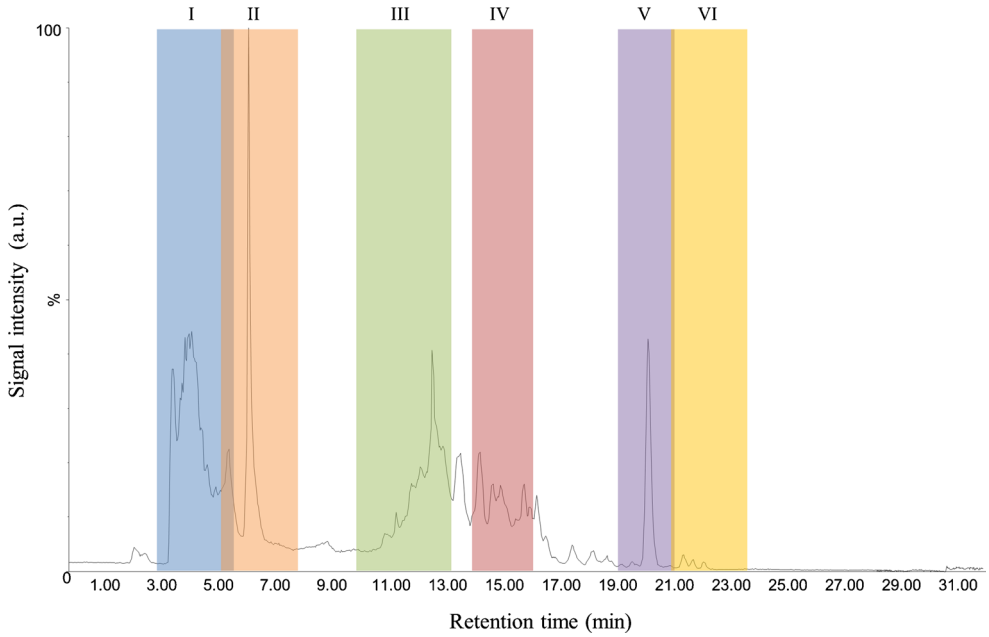

Fig. 1 Example of a TIC chromatogram of an aerial part sample with the selected important chromatographic regions highlighted

strongly coeluted metabolites, not resolved by MCR-ALS in an independent component due to their very low variance contribution during the ALS optimization. The resolution time for MCR-ALS was about 15 min for each chromatographic window. Although the full MCR-ALS approach workflow (mass spectral binning, selection of chromatographic time windows, resolution and selection of relevant components) required a larger amount of time. However, most of these steps did not require user participation, and so, automation and parallel computation could reduce the total time of analysis significantly.

Chromatographic information provided by XCMS and MCR-ALS strategies was rather similar. As mentioned before, XCMS algorithm detects features in all the cases that at a particular *m*/*z* value the signal is higher than a given threshold. Some of these features can be assigned to artifacts (instrumental noise, background and solvent contributions) and in other cases a single metabolite can give multiple features due to the

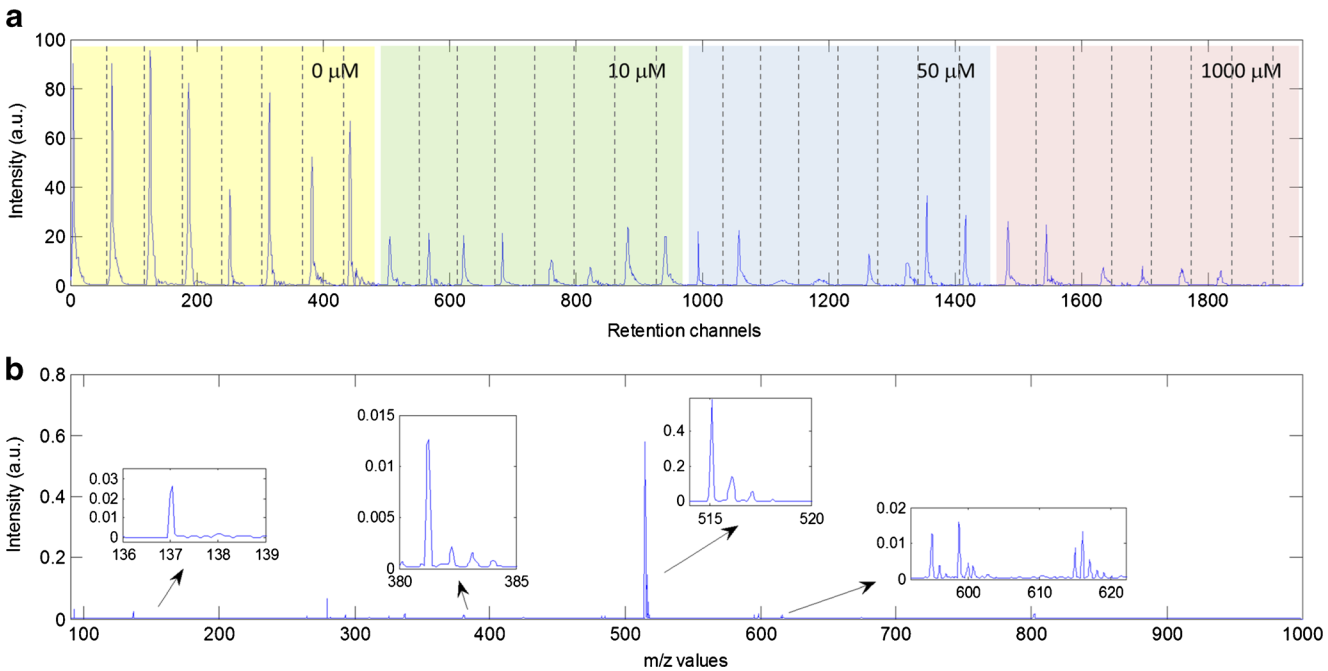

Fig. 2 Example of a resolved MCR component: aerial part rice samples treated with copper. **a** Resolved elution profiles, **b** resolved mass spectra. *Insets*: zoom in different *m*/*z* regions. *Dashed lines* separate each individual sample

detection of isotopic peaks or adducts. These two facts reduced the total number of independent features related to metabolites. The opposite case was found when considering the MCR-ALS resolved components. In this case, the final number of resolved components and, subsequently, the size of the $\mathbf{D}_{\mathbf{mcr\_areas}}$ matrix were significantly lower than the size of the $\mathbf{D}_{\mathbf{xcms\_areas}}$ matrix. This fact is caused because MCR-ALS resolves the components according to their elution profiles and their corresponding mass spectra. Every MCR-ALS resolved component will be associated with an elution profile and a mass spectrum which, however, may include various features at different *m*/*z* values. In this case, these features could be easily assigned to isotopic peaks of the nominal *m*/*z* values or different adducts of the same metabolite. Therefore, a single MCR-ALS resolved component provides information on several features related to the same metabolite grouped together. A more difficult situation could be found when two metabolites present a very strongly overlapped elution profile. If the number of components resolved by MCR-ALS resolution is too low, these two metabolites could appear as being resolved in the same component and, therefore, the resolved MS spectrum will have the features of the two compounds. Then, the interpretation of the resolved mass spectrum can be more difficult. Nevertheless, this information is still provided by MCR-ALS despite the smaller dimension of the $\mathbf{D}_{\mathbf{mcr\_areas}}$ matrix compared to $\mathbf{D}_{\mathbf{xcms\_areas}}$. It is possible that some information retrieved by the XCMS approach can be left in the residuals of the MCR-ALS model and lost for further analysis. This could be the case of rather small signals that contributes to a small quantity of the total variance of the data. Nevertheless, it is important to perform a deep study on those samples to differentiate if these little signals were due to the samples (and their treatment) or the previously described artifacts (background, solvent contributions).

Finally, when the use of the two approaches is compared, it may be argued that the XCMS workflow, as implemented in MetaboNexus, is more straightforward and that the final results can be reached faster. However, as stated above, the selection of the optimal pre-processing parameters in XCMS is a very critical aspect of the obtained results, and they should be properly optimized in order to have meaningful results. On the contrary, the MCR-ALS approach is more robust, although not so easy to use for the non-experienced user and, in principle, more time consuming. This robustness is demonstrated by the fact that satisfactory results can also be obtained when dealing with lower resolution data. For instance, MCR-ALS has successfully overcome inherent difficulties of GC-MS and CE-MS data analysis, such as multiple MS signals for the same metabolite due to derivatization in GC-MS and large migration time shifts between samples in CE-MS [33, 34].

As a result, $\mathbf{D}_{\mathbf{mcr\_areas}}$ and $\mathbf{D}_{\mathbf{xcms\_areas}}$ peak area data matrices were studied using an ASCA two-factor model with interaction considering the two factors, the sample tissue (root or aerial part) and the metal of treatment (Cd or Cu). Balanced experimental design considered 32 samples for each combination of the factors: root-Cd, root-Cu, aerial-Cd, and aerial-Cu. Results showed that both factors had a statistically significant effect (p=0.0001). Interaction between the kind of metal and the tissue analyzed was significant (p<0.050) in both cases (XCMS and MCR-ALS data). These results would indicate that the effects of each one of these two factors (type of metal and rice tissue) were dependent on the levels of the other.

Discrimination of rice samples due to metal treatment

Once peak areas were obtained with both XCMS and MCR-ALS, discrimination of samples according to the effects produced by the two metals (Cu and Cd) was investigated. With this aim, PCA and PLS-DA were applied to LC-MS peak areas of the different features or components obtained with both methodologies, XCMS and MCR-ALS, respectively. In both cases, results were analogous.

Tissue sample effect

Samples treated with Cd and samples treated with Cu were first studied separately (matrices $\mathbf{Cd}_{\mathbf{mcr\_areas}}$, $\mathbf{Cu}_{\mathbf{mcr\_areas}}$, $\mathbf{Cd}_{\mathbf{xcms\_areas}}$, $\mathbf{Cu}_{\mathbf{xcms\_areas}}$ described in Table 1). For both treatments, root samples were discriminated from aerial part samples by both PCA and PLS-DA. On the other hand, samples were also distinguished according to the metal treatment (Cd or Cu) when ASCA scores plot was considered (32 samples for each factor). Table 2 summarizes the figures of merit of the PLS-DA models.

As example of these results, Fig. 3a shows the scores plot for Cu treatment when PLS-DA was applied to MCR-ALS resolved peak areas, where root samples were discriminated from aerial part of rice samples by LV1, explaining 36 % of X-data variance. Figure 3b depicts the same results for Cd treatment, where root samples were discriminated from aerial part samples by LV1, explaining 45 % of X-data variance. As a conclusion, for both metal treatments, roots were clearly distinguished from aerial parts of rice samples with a Mathew's correlation coefficient (MCC) equal to 1.0, either for XCMS and for MCR-ALS results (analogous XCMS scores plots were obtained and shown in the ESM).

Type of metal effect

In order to evaluate more specifically the effect of metal treatment, root and aerial parts of rice samples were evaluated separately (matrices $\mathbf{L}_{\mathbf{mcr\_areas}}$, $\mathbf{R}_{\mathbf{mcr\_areas}}$, $\mathbf{L}_{\mathbf{xcms\_areas}}$, $\mathbf{R}_{\mathbf{xcms\_areas}}$ described in Table 1). For both tissues, samples treated with Cd were separated from samples treated with Cu by both PCA and PLS-DA. Samples treated with different metals were also discriminated when ASCA scores plots were considered.

Table 2 Figures of merit of the PLS-DA models

Matrix	Number of latent variables	X-variance explained	Y-variance explained	R^2 CV	MCC
$\mathbf{Cd}_{\mathbf{mcr\_areas}}$	2	50.23	81.19	0.54	1.000
$\mathbf{Cu}_{\mathbf{mcr\_areas}}$	2	51.70	94.61	0.86	1.000
$\mathbf{Cd}_{\mathbf{xcms\_areas}}$	2	37.23	94.76	0.85	1.000
$\mathbf{Cu}_{\mathbf{xcms\_areas}}$	2	50.22	96.82	0.95	1.000
$\mathbf{L}_{\mathbf{mcr\_areas}}$	2	41.57	91.59	0.66	1.000
$\mathbf{R}_{\mathbf{mcr\_areas}}$	2	49.67	77.61	0.59	1.000
$\mathbf{L}_{\mathbf{xcms\_areas}}$	2	40.47	96.77	0.94	1.000
$\mathbf{R}_{\mathbf{xcms\_areas}}$	2	49.19	91.60	0.82	1.000

As an example, Fig. 4a gives the scores plot for aerial parts of rice samples when PLS-DA was applied to resolved MCR-ALS peak areas, where samples treated with Cd were discriminated from samples treated with Cu by LV1, explaining 32 % of X-data variance. The scores plot for root samples is shown in Fig. 4b, where samples treated with Cd were distinguished from samples treated with Cu by LV1, explaining 44 % of X-data variance. For both tissues, samples exposed to Cd were clearly separated from samples exposed to Cu with a Mathew's correlation coefficient (MCC) equal to 1.0 (analogous XCMS results and scores plots were obtained, see ESM). Quality parameters of the obtained PLS-DA results are reported in Table 2. It can be seen that MCC values are close to one in all cases, although slightly better R^2 were obtained for XCMS data.

Identification of possible metabolite biomarkers of the effects of metals (Cu and Cd) on rice tissues

This study is concluded by the identification of some of the most relevant *m*/*z* values associated with the peak areas obtained using both strategies (XCMS and MCR-ALS). Variable importance on projection (VIP) scores (see ESM for the mathematical definition) were used to detect the most important features in each case.

Metabolomic interpretation of the changes caused by the different factors requires the identification of those variables (metabolites) having a VIP score higher than the average (greater than 1 rule), using for instance high-resolution MS combined with tandem MS. In this study and for brevity, only the six metabolites with the highest VIP scores for each PLS-DA model were considered. In order to select these six metabolites, the *m*/*z* values associated with the peak areas with the highest VIP scores were used. For instance, Fig. 5 gives the PLS-DA VIP scores plot for Cd-treated samples from peak areas obtained by MCR-ALS and XCMS ($\mathbf{Cd}_{\mathbf{mcr\_areas}}$ and $\mathbf{Cd}_{\mathbf{xcms\_areas}}$ matrices described in Table 1), showing the most relevant features (MCR-ALS resolved peaks or XCMS features, respectively) for discriminating between roots and aerial part rice samples. Results confirmed again that most of the *m*/*z* values obtained from XCMS peak areas were the same than those retrieved from the MCR-ALS resolved components.

The same procedure was also applied to the samples treated with Cu ($\mathbf{Cu}_{\mathbf{mcr\_areas}}$ and $\mathbf{Cu}_{\mathbf{xcms\_areas}}$ matrices) and it was also applied to the matrices containing only root samples or

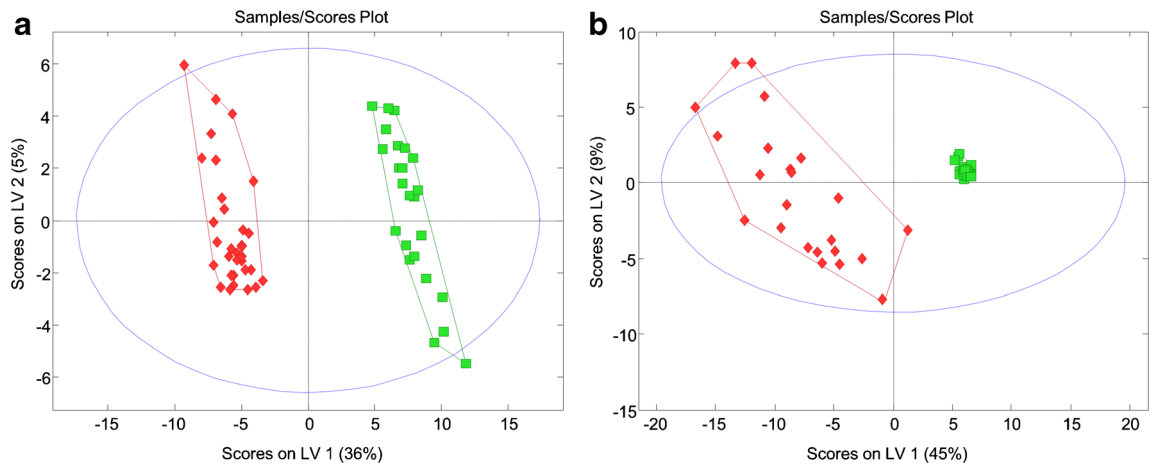

Fig. 3 PLS-DA results for tissue sample factor and MCR-ALS data. *Red diamonds* (♦) are aerial part samples and *green squares* (■) are root samples. **a** Scores plot for Cu treatment. **b** Scores plot for Cd treatment

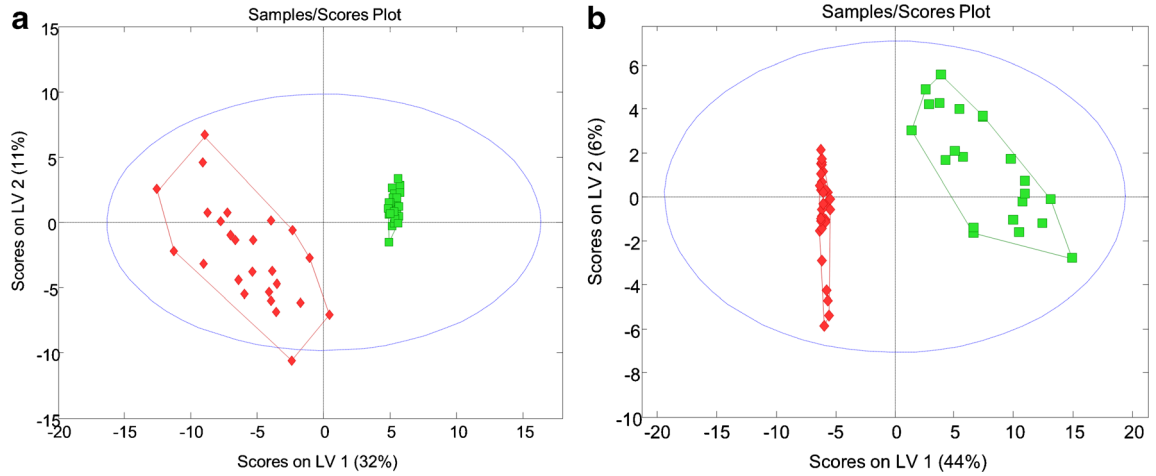

Fig. 4 PLS-DA results for metal of treatment factor and MCR-ALS data. *Red diamonds* (♦) are Cd samples and *green squares* (■) are Cu samples. **a** Scores plot for aerial part samples. **b** Scores plot for root samples

only aerial part samples ($\mathbf{L}_{\mathbf{mcr\_areas}}$, $\mathbf{R}_{\mathbf{mcr\_areas}}$, $\mathbf{L}_{\mathbf{xcms\_areas}}$, and $\mathbf{R}_{\mathbf{xcms\_areas}}$ matrices). Most of the relevant *m*/*z* values obtained using XCMS approach were almost the same than those obtained by MCR-ALS. In some cases, MCR-ALS resolved components gave more than one mass signal, for instance, peak number 100 in Fig. 5a with 337.09 and 199.80 *m*/*z* values. As mentioned before, these additional mass signals usually correspond to isotopic peaks and/or adducts of the same metabolite, but it could also correspond to two metabolites with extremely highly overlapped elution profiles (embedded peaks) which had been resolved in the same MCR-ALS component.

At this point, some characteristics of each method regarding metabolite identification can be discussed. MCR-ALS has the advantage of resolving the mass spectrum corresponding to every elution profile simultaneously. In this mass spectra, several features can be detected such as isotopic peaks, adducts eluted at the same retention time, and some solvent/background contributions. It is also possible that these different features can come from various metabolites eluting at the same retention times. This can be considered a drawback of the MCR-ALS since it makes their identification more laborious and time consuming. However, we think that these steps can be handled reasonably well, and new tools to do this in a more general and accurate way need to be developed. In the case of XCMS, the excessive number of detected features can be reduced by using appropriate complementary software (such as the previously mentioned CAMERA). However, in our experience, the detection of metabolites with similar behavior is not so straightforward such as in MCR-ALS where different metabolites in the same resolved component can be easily identified by their different MS signals.

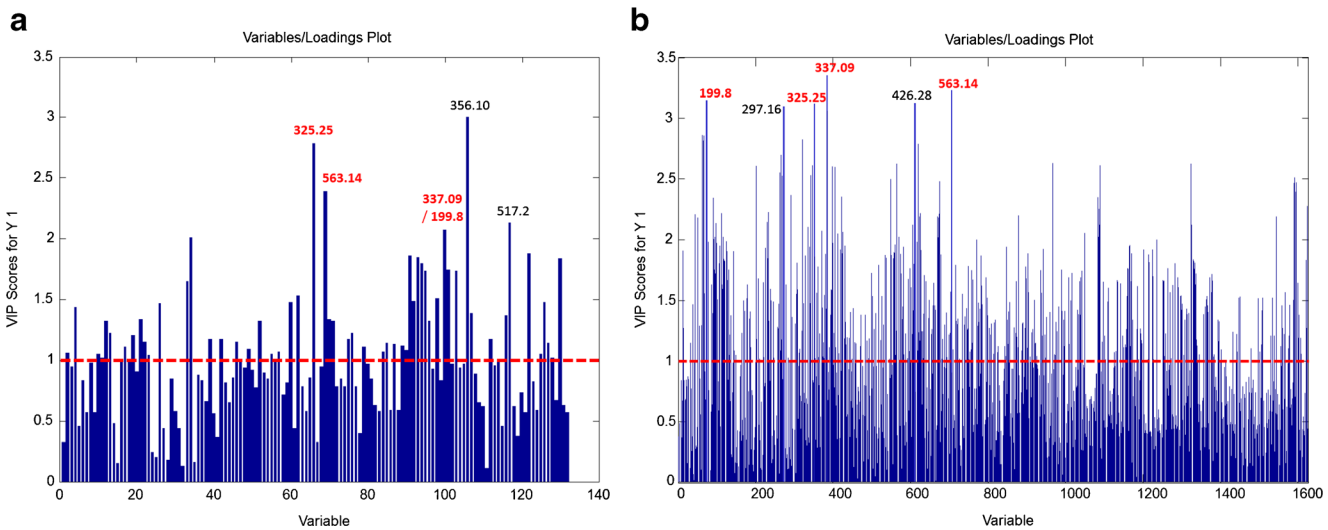

Fig. 5 PLS-DA VIP scores for samples treated with Cd. **a** Results for peak areas obtained with MCR-ALS. **b** Results for peak areas obtained with XCMS. *Numbers* indicate the *m*/*z* values associated with the peak areas with a higher VIP score. *Red* numbers indicate those that are equal for MCR-ALS and XCMS

Despite the main goal of this work was not the detailed biological interpretation of the observed metabolite changes, the tentative identification of some metabolites was attempted to confirm the reliability of the compared methodologies. In order to do this, *m*/*z* values of resolved peaks were compared with *m*/*z* values obtained from public databases such as MassBank [35], Metlin [36], and HMDB [37]. Since MS spectra of the components resolved by MCR-ALS were obtained from binned data, their accuracy is low (0.05 amu) and they cannot be used for a good identification of the metabolites. Therefore, once one particular component has been also identified by their precise elution profile, the retention time of its peak maximum in a given sample can be easily estimated and its accurate mass spectrum recovered from the original raw data file using MassLynx™ v 4.1 software. On the contrary, when XCMS was used, no binning approach was necessary. Therefore, the *m*/*z* accurate value of the feature detected was directly used for biomarker identification. Results of these tentative identifications are shown in Table 3.

From the results shown in Table 3, a preliminary biochemical interpretation can be attempted. For example, most of the metabolites shown in Table 3 are glycosides, such as 4-methylumbelliferyl glucoside, or other metabolites related to their biosynthesis, such as 6‴-*O*-sinapoylsaponarin or stearyl citrate [38]. Glycosides are very abundant in plants, and they have been reported to be altered by metal exposure in *Arabidopsis thaliana* [1, 39]. Glycosides have been also reported to form strong metal complexes, which would explain why their concentration changes so much by metal treatment [40, 41].

Conclusions

Both XCMS and MCR-ALS appeared to be powerful methodologies for metabolomic studies. These two approaches gave very similar results for all the experimental metabolomics systems investigated in the present work. XCMS resolved the entire full-scan chromatogram very fast, allowing a direct calculation of chromatographic peak areas for every *m*/*z* value giving a statistically significant MS signal, resulting in a huge number of features. In contrast, MCR-ALS led to a lower number of features since isotopic peaks and metabolites with very similar chromatographic profiles may be described in the same resolved elution profile with a MS spectrum including more than one *m*/*z* value. A comparison between the two approaches showed that XCMS was more straightforward for non-experienced users and required a smaller amount of processing time. However, MCR-ALS was more robust since it allowed working with more complex data (due to large time shifts, background contributions, and lower mass resolution) and did not require the optimization of parameters for each LC-MS instrument.

Chemometric evaluation of the peak areas for the candidate metabolites obtained by XCMS and MCR-ALS approaches allowed the statistical assessment of the effects caused by the metal exposure. ASCA statistical evaluation showed that there was a significant interaction between analyzed tissue sample and the type of metal of treatment. Further exploration by PCA and PLS-DA showed that metal exposure allowed the discrimination of rice samples considering both studied factors (type of metal and type of tissue). The identification of principal molecular biomarkers related to metal treatment

Table 3 Tentative identification of molecular markers using relevant *m*/*z* values considering the VIP scores

Highest mass ion	Retention time (min)	Proposed metabolite	Adduct	Adduct mass	Error *m*/*z* (ppm)	Factor
199.0392	26.87	Camalexin	M−H	199.0408	8.2	Part of the plant
325.2400	6.33	Avocadyne 4-acetate	M−H	325.2384	4.8	
337.0895[a]	6.33	4-Methylumbelliferyl glucoside	M−H	337.0929	10.1	
563.1365[a]	20.21	Apigenin 7-*O*-[beta-D-apiosyl-(1→2)-beta-D-glucoside]	M−H	563.1406	7.2	
799.1999[a]	20.42	6‴-*O*-Sinapoylsaponarin	M−H	799.2091	11.5	
593.2798	16.75	7,8-Dihydrovomifoliol 9-[rhamnosyl-(1→6)-glucoside]	M+Hac-H	593.2815	2.7	
279.0528[a]	27.52	7-Hydroxy-2-methyl-4-oxo-4H-1-benzopyran-5-carboxylic acid	M+Hac-H	279.0510	6.4	Metal of Treatment
329.0898[a]	6.1	1-*O*-Vanilloyl-beta-D-glucose	M−H	329.0878	6.1	
330.09[a]	19.25	Koenimbine	M+K-2H	330.0902	5.6	
331.0554	19.51	*S*-7-Methylthioheptylhydroximoyl-L-cysteine	M+K-2H	331.0558	1.2	
481.2549	6.17	Stearyl citrate	M+K-2H	481.2573	4.9	
446.1357[a]	28.6	(*R*)-2-Hydroxy-7,8-dimethoxy-2H-1,4-benzoxazin-3(4H)-one 2-glucoside	M+Hac-H	446.1304	11.8	

[a] Metabolites relevant also in the ASCA analysis

appears to be an important issue in plant biology studies to assess their metabolic changes. A preliminary evaluation of what metabolites were affected by metal exposure highlighted that glycoside family of compounds were much affected. Further work is pursued to perform a more exhaustive identification of metabolites by using other MS technologies, such as higher resolution MS instruments and MS/MS approaches. The combination of these advanced MS technologies with chemometric procedures will facilitate a more complete identification of molecular markers related to rice exposure to metals and the improvement of knowledge about metabolic changes.

Acknowledgments The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement n. 320737. Also, recognition from the Catalan government (grant 2014 SGR 1106) is acknowledged. JJ acknowledges a CSIC JAE-Doc contract cofounded by the FSE, and AGR thanks CONICET for a fellowship.

Conflict of interest The authors declare that they have no competing interests.

References

1. Fukusaki E, Kobayashi A (2005) Plant metabolomics: potential for practical operation. J Biosci Bioeng 100(4):347–354
2. Xiao JF, Zhou B, Ressom HW (2012) Metabolite identification and quantitation in LC-MS/MS-based metabolomics. TrAC Trends Anal Chem 32:1–14
3. Roessner U, Luedemann A, Brust D, Fiehn O, Linke T, Willmitzer L, Fernie AR (2001) Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. Plant Cell 13(1):11–29
4. Lommen A (2009) Metalign: Interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. Anal Chem 81(8):3079–3086
5. Pluskal T, Castillo S, Villar-Briones A, Orešič M (2010) MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics 11:395
6. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem 78(3):779–787
7. Johnson CH, Ivanisevic J, Benton HP, Siuzdak G (2015) Bioinformatics: the next frontier of metabolomics. Anal Chem 87(1):147–156
8. Tautenhahn R, Patti GJ, Rinehart D, Siuzdak G (2012) XCMS online: A web-based platform to process untargeted metabolomic data. Anal Chem 84(11):5035–5039
9. Jaumot J, de Juan A, Tauler R (2015) MCR-ALS GUI 2.0: new features and applications. Chemometr Intell Lab 140:1–12
10. Farrés M, Piña B, Tauler R (2014) Chemometric evaluation of Saccharomyces cerevisiae metabolic profiles using LC-MS. Metabolomics 11:210–224
11. Cramer GR, Urano K, Delrot S, Pezzotti M, Shinozaki K (2011) Effects of abiotic stress on plants: a systems biology perspective. BMC Plant Biol 11:163
12. D'Alessandro A, Taamalli M, Gevi F, Timperio AM, Zolla L, Ghnaya T (2013) Cadmium stress responses in Brassica juncea: hints from proteomics and metabolomics. J Proteome Res 12(11): 4979–4997
13. Villiers F, Ducruix C, Hugouvieux V, Jarno N, Ezan E, Garin J, Junot C, Bourguignon J (2011) Investigating the plant response to cadmium exposure by proteomic and metabolomic approaches. Proteomics 11(9):1650–1663
14. Järup L (2003) Hazards of heavy metal contamination. Brit Med Bull 68(1):167–182
15. Ahsan N, Nakamura T, Komatsu S (2012) Differential responses of microsomal proteins and metabolites in two contrasting cadmium (Cd)-accumulating soybean cultivars under Cd stress. Amino Acids 42(1):317–327
16. Antti H, Ebbels TMD, Keun HC, Bollard ME, Beckonert O, Lindon JC, Nicholson JK, Holmes E (2004) Statistical experimental design and partial least squares regression analysis of biofluid metabonomic NMR and clinical chemistry data for screening of adverse drug effects. Chemometr Intell Lab 73(1 SPEC. ISS): 139–149
17. Johnson HE, Lloyd AJ, Mur LAJ, Smith AR, Causton DR (2007) The application of MANOVA to analyse Arabidopsis thaliana metabolomic data from factorially designed experiments. Metabolomics 3(4):517–530
18. Aina R, Labra M, Fumagalli P, Vannini C, Marsoni M, Cucchi U, Bracale M, Sgorbati S, Citterio S (2007) Thiol-peptide level and proteomic changes in response to cadmium toxicity in Oryza sativa L. roots. Environ Exp Bot 59(3):381–392
19. Roth U, Von Roepenack-Lahaye E, Clemens S (2006) Proteome changes in Arabidopsis thaliana roots upon exposure to Cd 2+. J Exp Bot 57(15):4003–4013
20. Huang SM, Toh W, Benke PI, Tan CS, Ong CN (2014). MetaboNexus: an interactive platform for integrated metabolomics analysis. Metabolomics 10:1084–1093
21. Prince JT, Marcotte EM (2006) Chromatographic alignment of ESI-LC-MS proteomics data sets by ordered bijective interpolated warping. Anal Chem 78(17):6140–6152
22. Tautenhahn R, Bottcher C, Neumann S (2008) Highly sensitive feature detection for high resolution LC/MS. BMC Bioinformatics 9:504
23. De Juan A, Jaumot J, Tauler R (2014) Multivariate Curve Resolution (MCR). Solving the mixture analysis problem. Anal Methods 6(14):4964–4976
24. de Juan A, Tauler R (2007) Factor analysis of hyphenated chromatographic data. Exploration, resolution and quantification of multicomponent systems. J Chromatogr A 1158(1–2):184–195
25. Ruckebusch C, Blanchet L (2013) Multivariate curve resolution: a review of advanced and tailored applications and challenges. Anal Chim Acta 765:28–36
26. Golub GH, Loan CFV (1996) Matrix computations, 3rd edn. Johns Hopkins University Press, Baltimore
27. Windig W, Guilment J (1991) Interactive self-modeling mixture analysis. Anal Chem 63(14):1425–1432
28. Tauler R, Maeder M, de Juan A (2010) Multiset data analysis: extended multivariate curve resolution. In: Brown SD, Tauler R, Walczak, B (ed) Comprehensive Chemometrics, vol 2. Elsevier B.V., Amsterdam, pp 473–505
29. Wold S, Esbensen K, Geladi P (1987) Principal component analysis. Chemometr Intell Lab 2(1–3):37–52
30. Barker M, Rayens W (2003) Partial least squares for discrimination. J Chemom 17(3):166–173
31. Jansen JJ, Hoefsloot HCJ, Van Der Greef J, Timmerman ME, Westerhuis JA, Smilde AK (2005) ASCA: analysis of multivariate data obtained from an experimental design. J Chemom 19(9):469–481

32. Kuhl C, Tautenhahn R, Böttcher C, Larson TR, Neumann S (2012) CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. Anal Chem 84(1):283–289
33. Ortiz-Villanueva E, Jaumot J, Benavente F, Piña B, Sanz-Nebot V, Tauler R (2015) Combination of CE-MS and advanced chemometric methods for high-throughput metabolic profiling. Electrophoresis 36:2324–2335
34. Schmidtke LM, Blackman JW, Clark AC, Grant-Preece P (2013) Wine metabolomics: objective measures of sensory properties of semillon from GC-MS profiles. J Agric Food Chem 61(49):11957–11967
35. Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K, Ojima Y, Tanaka K, Tanaka S, Aoshima K, Oda Y, Kakazu Y, Kusano M, Tohge T, Matsuda F, Sawada Y, Hirai MY, Nakanishi H, Ikeda K, Akimoto N, Maoka T, Takahashi H, Ara T, Sakurai N, Suzuki H, Shibata D, Neumann S, Iida T, Tanaka K, Funatsu K, Matsuura F, Soga T, Taguchi R, Saito K, Nishioka T (2010) MassBank: a public repository for sharing mass spectral data for life sciences. J Mass Spectrom 45(7):703–714
36. Tautenhahn R, Cho K, Uritboonthai W, Zhu Z, Patti GJ, Siuzdak G (2012) An accelerated workflow for untargeted metabolomics using the METLIN database. Nat Biotechnol 30(9):826–828
37. Wishart DS, Knox C, Guo AC, Eisner R, Young N, Gautam B, Hau DD, Psychogios N, Dong E, Bouatra S, Mandal R, Sinelnikov I, Xia J, Jia L, Cruz JA, Lim E, Sobsey CA, Shrivastava S, Huang P, Liu P, Fang L, Peng J, Fradette R, Cheng D, Tzur D, Clements M, Lewis A, de Souza A, Zuniga A, Dawe M, Xiong Y, Clive D, Greiner R, Nazyrova A, Shaykhutdinov R, Li L, Vogel HJ, Forsythei I (2009) HMDB: a knowledgebase for the human metabolome. Nucleic Acids Res 37(Suppl 1):D603–D610
38. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 40(D1):D109–D114
39. Sun X, Zhang J, Zhang H, Ni Y, Zhang Q, Chen J, Guan Y (2010) The responses of Arabidopsis thaliana to cadmium exposure explored via metabolite profiling. Chemosphere 78(7):840–845
40. Satterfield M, Brodbelt JS (2001) Structural characterization of flavonoid glycosides by collisionally activated dissociation of metal complexes. J Am Soc Mass Spectrom 12(5):537–549
41. Gyurcsik B, Nagy L (2000) Carbohydrates as ligands: coordination equilibria and structure of the metal complexes. Coordin Chem Rev 203(1):81–149