Accepted Manuscript

Title: Told through the wine: an liquid chromatography–mass spectrometry interplatform comparison reveals the influence of the global approach on the final annotated metabolites in non-targeted metabolomics

Author: Ramon Díaz Hector Gallart-Ayala Juan V. Sancho Oscar Nuñez Tatiana Zamora Claudia P.B. Martins Félix Hernández Santiago Hernández-Cassou Javier Saurina Antonio Checa

PII: S0021-9673(16)00030-3
DOI: http://dx.doi.org/doi:10.1016/j.chroma.2016.01.010
Reference: CHROMA 357203

To appear in: *Journal of Chromatography A*

Received date: 17-10-2015
Revised date: 19-12-2015
Accepted date: 5-1-2016

Please cite this article as: Ramon Díaz, Hector Gallart-Ayala, Juan V.Sancho, Oscar Nuñez, Tatiana Zamora, Claudia P.B.Martins, Félix Hernández, Santiago Hernández-Cassou, Javier Saurina, Antonio Checa, Told through the wine: an liquid chromatographyndashmass spectrometry interplatform comparison reveals the influence of the global approach on the final annotated metabolites in non-targeted metabolomics, Journal of Chromatography A http://dx.doi.org/10.1016/j.chroma.2016.01.010

Told through the wine: an liquid chromatography–mass spectrometry interplatform comparison reveals the influence of the global approach on the final annotated metabolites in non-targeted metabolomics

Ramon Díaz[a], Hector Gallart-Ayala[b,c], Juan V. Sancho[a], Oscar Nuñez[b,d], Tatiana Zamora[a], Claudia P.B. Martins[e], Félix Hernández[a], Santiago Hernández-Cassou[b], Javier Saurina[b] and Antonio Checa[b,c,*]

[a] *Research Institute for Pesticides and Water, University Jaume I, Av. Sos Baynat S/N, 12071 Castellón, Spain*

[b] *Department of Analytical Chemistry, Universitat de Barcelona, Martí i Franquès 1-11, E-08028 Barcelona, Spain*

[c] *Department of Medical Biochemistry and Biophysics, Division of Physiological Chemistry II, Karolinska Institutet, Scheeles väg 2, SE - 17177, Stockholm, Sweden*

[d] *Serra Húnter Fellow, Generalitat de Catalunya, Barcelona, Spain*

[e] *Thermo Fisher Scientific, 355 River Oaks Parkway, 95134, San Jose, CA, USA*

* Corresponding author: Antonio Checa

Department of Medical Biochemistry and Biophysics, Division of Physiological Chemistry II,

Karolinska Institutet, Scheeles väg 2, SE - 17177, Stockholm, Sweden Phone: +46 8 524 87 86;

Fax: +46 8 736 04 39

e-mail: antonio.checa@ki.se

Highlights

- Two independent metabolomics platforms were employed for wine discrimination
- The ability of each platform on identifying discriminating markers was compared
- High divergence was observed in the original identified metabolites
- Identified metabolites were cross-validated between platforms

Abstract

This work focuses on the influence of the selected LC-HRMS platform on the final annotated compounds in non-targeted metabolomics. Two platforms that differed in columns, mobile phases, gradients, chromatographs, mass spectrometers (Orbitrap [Platform#1] and Q-TOF [Platform#2]), data processing and marker selection protocols were compared. A total of 42 wines samples from three different protected denomination of origin (PDO) were analyzed. At the feature level, good (O)PLS-DA models were obtained for both platforms (Q^2[Platform#1]=0.89, 0.83 and 0.72; Q^2[Platform#2]=0.86, 0.86 and 0.77 for Penedes, Ribera del Duero and Rioja wines respectively) with 100% correctly classified samples in all cases. At the annotated metabolite level, platforms proposed 9 and 8 annotated metabolites respectively which were identified by matching standards or the MS/MS spectra of the compounds. At this stage, there was no coincidence among platforms regarding the suggested metabolites. When screened on the raw data, 6 and 5 of these compounds were detected on the other platform with a similar trend. Some of the detected metabolites showed complimentary information when integrated on biological pathways. Through the use of some examples at the annotated metabolite level, possible explanations of this initial divergence on the results are presented. This work

shows the complications that may arise on the comparison of non-targeted metabolomics platforms even when metabolite focused approaches are used in the identification.

Keywords: Metabolomics, Metabolite identification, LC-HRMS, Wine

1. Introduction

Food characterization approaches are useful to increase our knowledge of products that regularly reach our tables. In recent years and thanks to the advances in technology and instrumentation, new powerful tools for the characterization and differentiation of foodstuffs and beverages have emerged. Among them metabolomics, defined as the study of small molecules (<1500 Da) occurring in a given biological system, has proven its utility in many research areas, including biology and physiology[1], plant science [2] and food characterization and discrimination [3, 4]. Specifically, several works have used metabolomics approaches in wine origin authentication [5-7].

Metabolomic methods can be classified into targeted and non-targeted modes. Targeted methods are mainly focused on selected pre-defined group or groups of metabolites and can usually provide quantitative measurements of the selected compounds, while non-targeted metabolomics aims at the pseudo-quantitative detection of many groups of compounds. Though absolute quantification is not possible, non-targeted methods are useful in the characterization (fingerprinting) and classification of samples based on the differential abundance of the compound responses among the classes. These approaches have been applied to the identification and authentication of complex alcoholic beverages, such as wine. Specifically, targeted

approaches in wine characterization and discrimination have mostly been based on the determination of naturally occurring compounds such as organic acids [8], volatile [9] and phenolic compounds [10, 11]. In non-targeted approaches, on the other hand, the different families of metabolites can be simultaneously characterized and analyzed to provide a specific sample fingerprint with discriminating features that may [12-16] or may not [6] be later identified.

Though non-targeted metabolomics approaches are potentially much more informative than the targeted ones, in practice the annotation of the features obtained either by using databases or by matching with pure standards is required. As the metabolomic fingerprint is usually enough to discriminate the origin of the wines, further steps of identification are not usually carried out. This complicates the comparison with other laboratories and the translation of the method into another with a higher throughput or more cost-effective.

Some studies have been carried out in liquid chromatography-high resolution mass spectrometry (LC-HRMS) based metabolomics to evaluate the influence of the instruments [17, 18] on the results, concluding that despite the expected differences, a high degree of overlapping exists when the only difference corresponds to the mass spectrometer. However, differences between metabolomic platforms from different laboratories include several additional variance factors, such as sample treatment, column, mobile phases and data processing. The influence of the analytical platform on the metabolic profiling level of urine samples has been recently evaluated [19]. High convergence between the spectral information from the different platforms was obtained, and an approach to evaluate how this translates into fully identified or annotated features was suggested as future work [19].

In the present work we compared the performance of two different non-targeted LC-HRMS platforms from two different laboratories to tackle the discrimination of Spanish red wines from

three different protected denominations of origin (PDO). The ability of both platforms indiscovering discriminating markers at the annotated metabolite level for each PDO was inspected. Finally, discriminating metabolites individually suggested for each platform were compared and obtained results were put in a biological context.

2. Materials and methods

2.1. Reagents

All reagents used were of analytical grade unless otherwise specified. Formic acid (99%) was obtained from Merck (Darmstadt, Germany) and Sharlau (Barcelona, Spain). LC-MS grade acetonitrile and water were purchased from Riedel-de-Haën (Seelze, Germany). HPLC grade methanol was purchased from Sharlau. Polyphenol standards were purchased from Sigma-Aldrich (St. Louis, MO). Mobile phases were filtered using 0.45 μm nylon membrane filters (Whatman, Clifton, NJ).

2.2. Samples and sample pretreatment

Forty-two commercial wines from three different Spanish PDO (14 Penedes, 14 Ribera del Duero and 14 Rioja) were purchased from retail stores. More detailed information about samples is supplied in Supplementary Table 1. Samples were opened and two different 5 mL aliquots were withdrawn, frozen at -20ºC and distributed between both platforms (Figure 1). Frozen samples were analyzed within a two weeks period after aliquotation. On the day of analysis, wines were thawed at room temperature and diluted 1:1 with double deionized water. Prior to injection, wine samples were filtered through 0.22 μm teflon membranes (Scharlab, Barcelona, Spain) for

Platform#1 and centrifuged at 10000 *g* for Platform#2. Experimental workflow is summarized in Figure 1.

2.3. UHPLC-HRMS methods

2.3.1 Platform#1. Orbitrap based metabolomics

An ultra-high performance liquid chromatography (UHPLC) system (Accela; Thermo Fisher Scientific, San José, CA, US) was coupled to a single-stage Orbitrap instrument (Exactive HCD; Thermo Fisher Scientific, Bremen, Germany) equipped with a heated electrospray (H-ESI II) ionization source operating in the negative ion mode. The chromatographic separation was performed in a Hypersil Gold aQ (100 mm x 2.1 mm i.d., 1.9 µm particle size) column, from Thermo Fisher Scientific, using a gradient elution of 0.1% formic acid in water (solvent A) and acetonitrile (solvent B) at 600 µL min^{-1}. The column temperature and sample tray were held at 25°C and 4°C, respectively. The UHPLC-HRMS details of Platform#1 are summarized in Supplementary Table 2.

2.3.2. Platform#2. Q-TOF based metabolomics.

A Waters Acquity UPLC system (Waters, Milford, MA, USA) was interfaced to a hybrid quadrupole-orthogonal acceleration-TOF mass spectrometer (Xevo G2 Q-TOF, Waters Micromass, Manchester, UK), using an orthogonal Z-spray-ESI interface operating in negative ion mode. The UHPLC separation was performed using an Acquity UPLC BEH C18 column (100 × 2.1mm, 1.7 µm particle size) from Waters at a 300 µL min^{-1} flow rate. Mobile phase A and B consisted of water and methanol, both containing 0.01% formic acid. The column

temperature and sample tray were held at 40°C and 4°C, respectively. The UHPLC-HRMS details of Platform#2 are summarized in Supplementary Table 2.

2.4. Data processing and statistical analysis

Platform#1. For global analysis, LC-HRMS(/MS) raw files were first converted to mzXML format using MSConvert. Automatic peak detection and integration was performed using XCMS package for R. Centwave peak picking algorithm was used based on previous instrumental data analysis experience (peak width = 6-15 s, S/N threshold = 5, ppm = 15). Peak grouping was carried out using a density grouping method with an initial band width of 15 s. The resulting data matrix was then exported as comma separated values (CSV) files. Standard t-tests were conducted for univariate statistics using R (Version 3.1.0). SIMCA-P+® (Version 13, Umetrics AB, Sweden) software was used to perform multivariate statistical analysis. Principal Component Analysis (PCA) and Partial Least Square Discriminant Analysis (PLS-DA) were applied to build descriptive and predictive models. The validity and robustness of the models were evaluated by $R^2(Y)$ and $Q^2(Y)$ parameters. Quality of cross-validation $Q^2(Y)$ value was assessed by using 2000 permutation tests as previously described [20]. Features in multivariate analysis were ranked according to their Variable Importance in Projection (VIP) values. Cross-validation models were used to calculate jackknife uncertainty of VIP.

Platform#2. UPLC-(Q)TOF MS data was converted from proprietary to netCDF format using Databridge application (within MassLynx v4.1; Waters Corporation) and pre-processed using XCMS package for R. Centwave peak picking algorithm was used based on previous instrumental data analysis experience (peak width = 5-20 s, S/N threshold = 10, ppm = 25). Peak

grouping was carried out using a density grouping method with an initial band width of 30 s. Then, peak retention time was aligned using a non-linear locally weighted scatter plot smoothing (LOESS) correction allowing a maximum of five missing samples to consider a peak group as "well behaved" and eligible to be included as an element of the retention time correction spine. Data was normalized to overcome instrumental drift due to cumulative interface contamination. Iterative LOESS normalization was performed, making the median intensities of all features in every sample equal as previously described [21].

Multivariate analysis was carried out using SIMCA-P+® (Version 12, Umetrics, Sweden). Partial Least Square Discriminant Analysis (PLS-DA) was employed to distinguish samples according to their origin. Model was evaluated based on classification results after cross-validation. Then, orthogonal PLS-DA (OPLS-DA) analysis was used to highlight the markers to differentiate between classes.

2.5. Variable annotation

For simplification purposes, only compounds of plant origin were considered in both platforms. When possible, compounds were confirmed by matching retention time and spectra of analytical standards. Otherwise, characteristic fragments in MS/MS were screened for each compound.

Platform#1. Features showing significant VIP values combined with low jackknife errors (VIP-jackniffe > 1) were submitted to annotation process.

Platform#2. For candidate markers selection OPLS-DA S-Plot was used confronting every group with the rest. Those features with p(corr) > 0.7 were submitted for annotation.

For both platforms, the combination of accurate mass measurement of the precursor ions obtained in full scan mode in combination with the HRMS(/MS) fragmentation was used for identification

purposes. METLIN (http://metlin.scripps.edu) and the human metabolome database (http://www.hmdb.ca) were used as databases for metabolite annotation.

3. Results and discussion

The present study aimed at comparing the performance of two different metabolomics platforms on the identification of discriminating compounds in wine authentication. Samples belonging to three different PDO that had already been discriminated according to their (poly)phenolic fingerprint were selected in order to address two questions. The main objective was to compare the impact of the metabolomic workflow on the discriminating compounds reported. Then we aimed to inspect how an LC-HRMS based metabolomics approach could improve and complement the generated knowledge of a less specific LC-UV platform [5] at the metabolite level. Samples were analysed using two different HRMS instrument configurations widely employed in metabolomics analyses, Orbitrap (Platform#1) and Q-TOF (Platform#2). Reversed phase chromatography was selected due to the nature of compounds involved, mainly polyphenols, though specific columns were used by each platform. For the same reason, negative ionization mode was chosen.

The results showed that despite the ability of the platforms to distinguish the wine classes both at the spectral and at the annotated metabolite level, a strong divergence among the annotated metabolites involved in the discrimination occurred. Some of these discriminating compounds could be detected on the other platform when they were screened in a targeted manner, showing the usefulness of employing different platforms in non-targeted metabolomics.

3.1. PDO discrimination of samples at the feature level

At the extracted features level, PDO classes were separated using both experimental setups (Figure 2a-b). Features obtained from Platform#1 presented proportionally lower *m/z* relative to Platform#2. This observation is in accordance with previous experiments where Orbitrap and Q-TOF platforms were used [22].

Both platforms used a pooled QC from all samples to control for system stability. On Platform#1, no algorithm was employed for data correction and only compounds presenting a %CV≤30 on the QCs were submitted to later statistical analysis. On Platform#2, data was normalized using LOESS normalization which adjusts the local median of log fold changes of peak intensities between samples in the data set to be approximately zero across the whole peak intensity [21]. Normalization used in Platform#2 resulted in a higher control of the experimental variation as shown by the clustering of the QCs (Figure 2b). In both platforms, separation of the three classes using unsupervised PCA was observed across PC1 for Penedes wines with a higher degree of overlapping between Ribera del Duero and Rioja. This is in accordance with our previous results based on the analysis of the LC-UV profiles of the samples [5]. This was attributed to the higher similarities in climatological conditions resulting from a closer geographical location of the PDO and also from the homogeneity on the grape variety, where Tempranillo grapes are predominantly used (Supplementary Table 1). However, Penedes samples obtained from Tempranillo grape variety did not cluster with the other two PDO, thus suggesting that growing conditions do affect the fingerprint more than grape type in terms of PCA variance. The same predominance of PDO over grape variety is supported by previous findings [5].

Classification of samples according to their PDO was accomplished by means of PLS-DA and OPLS-DA for Platforms #1 and #2 respectively. In both cases, cross-validated models using all extracted features allowed the correct classification of the wines according to PDO (Table 1).

Presence of overfitting due to the high feature to sample ratio was tested by comparing the obtained Q^2 to the values of 2000 permutated models [20]. The value of Q^2 obtained was lower than the value of the selected model in all cases (Table 1), thus suggesting that no overfitting occurred ($p < 1/2000$). This is also supported by the plot of the distribution of the permutated models in all cases (Supplementary Figure 1).

3.2. PDO discrimination of samples at the annotated metabolite level

The top discriminating features between PDO classes were annotated using the selected protocol for each platform (See Section 2.5). Annotation was performed both after operator visual spectral inspection and using automated procedures based on accurate mass, referencing to public or in-house databases and, when possible, using commercially available compounds. Annotated compounds that were not from plant origin or presented an improbable retention time (*e.g.* highly retained polar compounds) were not considered. In addition, when more than three possible candidates were returned by the database, the feature was also discarded. This approach reduced the number of potentially discriminating features (See Table 2), but enabled proper identification of the compounds by means of MS/MS fragmentation. It also facilitated the later inter-platform comparison and interpretation of the results in biological pathways which is the desirable endpoint of most metabolomic studies. A total of 9 and 8 features were identified for Platforms #1 and #2 respectively (Table 2), although none of them was common.

PCA models built using only annotated features on each platform resulted in a clear class separation for Platform#2 (Figure 2d). In the case of Platform#1, separation between Ribera del Duero and Rioja was not complete using the unsupervised model (Figure 2c). OPLS-DA models built and validated as described in Section 3.1 using annotated metabolites for each platform

provided similar results to the ones obtained when all features were used (Table 1). This shows that though the restrictions imposed in the annotation drastically reduced the number of variables to be used, annotated metabolites were still important for the performance of the separation in both platforms. None of the platforms allowed to identify characteristic compounds of Rioja wines based on the selected criteria. Separation for this class at the metabolite level was thus based on the lack of the characteristic annotated compounds for each of the other two PDO.

3.3. Interplatform comparison of annotated metabolites

The final goal of the workflow consisted of the integration of the results obtained from both platforms. Estimated retention times were obtained by injecting a mixture of polyphenols in both platforms (Figure 3a, Supplementary Table 3). Annotated compounds for each of the platforms presenting mass errors below 5 ppm at the expected retention time ±15% were screened in the raw files of the other platform and areas were extracted (Figure 3a, Supplementary Table 4). Compounds annotated in one platform that were validated in the other are presented on Table 2. A total of 6 and 5 features from Platforms #1 and #2 were identified on Platforms #2 and #1, respectively. This shows the influence of the platform, including the whole pipeline of analysis on the reported results of the metabolomics workflow. The aforementioned capability of each instrument to detect lower or higher masses was also reflected on the distribution of the *m/z* for identified compounds (Table 2). The resulting PCA model performed including only the annotated features common for both platforms showed a high degree of similarity between them (Figure 2e-f). This consistency was also maintained for the commonly identified metabolites (Figures 3 b-c and 4). (O)PLS-DA models performed for each platform only using features

identified in both platforms also allowed the correct classification of all samples according to PDO (Table 1).

3.4. Pathway related annotated metabolite comparisons

Each platform was able to identify one resveratrol derived compound as discriminating feature for Penedes wines (Table 2, Figure 3b-c). This is in accordance with our previous publication, where trans- and cis-piceid were also identified. Using Platform#2, only trans-piceid could be validated. In Platform#1 both cis- and trans-piceid (glucosides of resveratrol) were identified. On the other hand, Platform#2 identified a methylated form of the glucoside that presented a much higher fold change. The compound was annotated using the accurate mass of the deprotonated molecule and the presence of a methyl-resveratrol fragment (Supplementary Figure 2a). As two different compounds presented the same accurate mass and fragment, trans- form was selected as the annotated compound based on the retention time pattern of both resveratrol and piceid, as they elute before the cis- form (Figure 3a, Supplementary Table 3). An inspection of the compound on the raw data of Platform#1 showed that elevated levels of methyl resveratrol glucoside could also be found using this platform. Further inspection of the table obtained during data processing showed that the compound was discarded before submitting to the multivariate analysis because it presented a high %CV on the QCs and thus it was not originally identified. This example illustrates one of the risks of excluding features by a fixed %CV value on the QCs as compounds presenting high fold changes can still be identified despite presenting high %CV. As it was available on the original polyphenol mixture, trans-resveratrol was also screened on the samples (Figure 3d). Though none of the platforms was able to identify it as a discriminating compound, pattern shown in Penedes wines strengthens the results found for its derivate

compounds. For the previous study [5], low absorbance of resveratrol was the most likely reason why the compound was not identified using the LC-UV profile. In the case of LC-HRMS platforms, a possible explanation is the correlation with the other resveratrol derived compounds that showed stronger differences.

Flavanol related compounds were also identified by both platforms (Figure 4). Platform#1 found increased levels of the flavanols catechin and epicatechin flavanols in Penedes wines (Figure 4 a-b) and of their derived compounds gallocatechin and epigallocatechin in Ribera del Duero wines (Figure 4 c-d). Both findings could be reproduced in Platform#2. A digalactoside of myricetin was found increased in Ribera del Duero wines (Figure 4e) by Platform#2. Integration of the compound in Platform#1 resulted in an equal significant change, but fold change relative to the other two PDO was 12 times lower (Figure 4e). This could be attributed to the presence of interfering coeluting compounds causing ion suppression and affecting the robustness of the quantification of this compound in Platform#1 (Supplementary Figure 2b). In a previous study, myricetin and one of its glucosides were found increased in Ribera del Duero wines [5]. As in the case of the resveratrol derived compounds, the use of the non-targeted platform reinforces the idea of an upregulation of this part of the pathway (Figure 4) on this PDO. Though myricetin was not originally identified by none of the platforms, manual integration of the compound showed its increased levels in Ribera del Duero wines.

These examples show an important issue that may arise when comparing non-targeted, but also targeted metabolomics platforms where different chromatographic methods are used. In the present case, this strengthens the confidence of the found markers as they were obtained using two different chromatographic methods. On the other hand, when results are not validated, it may be difficult to make a decision on which result is the correct one if a labelled internal standard for

the compound is not included on the analysis. Sample pre-treatment is another potential source that may introduce ion suppression related differences. In the present case, the different treatments carried out (filtration vs centrifugation, See Section 2.2) were not expected to cause significant differences. In addition the relative abundance between the depronotaned molecule $[M-H]^-$ and its product ions indicates a high fragmentation for the myricetin glucoside in Platform#1. This effect was observed for the other glucosides identified on Table 2 and may explain the original lack of detection of these compound on Platform#1. This shows the challenge that must be faced on the optimization of source parameters in a non-targeted metabolomics approaches. In the present case, selected source voltages in Platform#1 were optimized using the standard mix of polyphenols used to compare between platforms (Supplementary Table 3). This resulted in an increase of in-source fragmentation for polyphenols presenting sugar moieties on their structure that complicated their detection.

Finally, the importance of the role of the chromatographic retention in non-targeted metabolomics was pointed out in the case of gallic acid. This compound was detected and identified as a marker of Penedes wines by Platform#1 in agreement with our previous publication [5]. However, gallic acid was not identified in the wine samples analysed by Platform#2. In this case the closeness of the compound to the void volume, in combination with the low *m/z* value could be a possible lack of identification of this compound using Platform#2.

4. Conclusions

Final identification of the metabolites remains one of the bottlenecks of non-targeted metabolomics platforms. Through the use of a restricted system where only a family of compounds as a fraction of the metabolome was used to facilitate the comparison (polyphenols

and plant related molecules), this work shows the difficulties of obtaining robust results in terms of annotated metabolites at the end of the workflow. Direct comparisons of the results among platforms may result in a difficult issue due to differences in chromatography, mass spectrometer system and optimization and statistical biomarker selection tools. These complications are expected to increase when carrying out a comparison dealing with a higher coverage of the metabolome. On the other hand, we show that the use of different platforms allows the results to be both complemented and validated as compared to single platform based approaches, thus increasing the confidence in the obtained outcomes.

Acknowledgements

The authors are grateful to Serveis Centrals d´Instrumentació Científica (SCIC) of University Jaume I for the use of UPLC-QTOF MS (QTOF Premier) and to Nuria León and the Public Health Laboratory of Valencia for technical support on the use of Q-Exactive. This work was supported by the Spanish Ministerio de Ciencia y Tecnología (project CTQ2008-04776/BQU), Fundació Bancaixa (project P1-1B2010-50) and the Generalitat Valenciana (Research Group of Excellence, Prometeo/2009/054; PrometeoII/2014/023).

References

[1] M. Anas Kamleh, K. Spagou, E. J. Want, Metabolic Profiling in Disease Diagnosis, Toxicology and Personalized Healthcare, Curr. Pharm. Biotechnol. 12 (2011) 976-995.

[2] J.L. Wolfender, S. Rudaz, Y.H. Choi, H.K. Kim, Plant Metabolomics: From Holistic Data to Relevant Biomarkers, Curr. Med. Chem. 20 (2013) 1056-1090.

[3] J.M. Cevallos-Cevallos, J.I. Reyes-De-Corcuera, E. Etxeberria, M.D. Danyluk, G.E. Rodrick, Metabolomic analysis in food science: a review, Trends Food Sci. Tech. 20 (2009) 557-566.

[4] A. Checa, J. Saurina, Chapter 6 - Metabolomics and PDO, Comprehensive Analytical Chemistry, Elsevier, 2013, pp. 123-143.

[5] D. Serrano-Lourido, J. Saurina, S. Hernández-Cassou, A. Checa, Classification and characterisation of Spanish red wines according to their appellation of origin based on chromatographic profiles and chemometric data analysis, Food Chem. 135 (2012) 1425-1431.

[6] A.E. Springer, J. Riedl, S. Esslinger, T. Roth, M.A. Glomb, C. Fauhl-Hassek, Validated Modeling for German White Wine Varietal Authentication Based on Headspace Solid-Phase Microextraction Online Coupled with Gas Chromatography Mass Spectrometry Fingerprinting, J. Agr. Food Chem. 62 (2014) 6844-6851.

[7] A. Versari, V.F. Laurie, A. Ricci, L. Laghi, G.P. Parpinello, Progress in authentication, typification and traceability of grapes and wines by chemometric approaches, Food Res. Int. 60 (2014) 2-18.

[8] L.T. Gallagher, J.S. Heo, M.A. Lopez, B.M. Ray, J. Xiao, A.P. Umali, A. Zhang, S. Dharmarajan, H. Heymann, E.V. Anslyn, Pattern-based discrimination of organic acids and red wine varietals by arrays of synthetic receptors, Supramol. Chem. 24 (2012) 143-148.

[9] M.P. Fabani, M.J.A. Ravera, D.A. Wunderlin, Markers of typical red wine varieties from the Valley of Tulum (San Juan-Argentina) based on VOCs profile and chemometrics, Food Chem. 141 (2013) 1055-1062.

[10] L. Jaitz, K. Siegl, R. Eder, G. Rak, L. Abranko, G. Koellensperger, S. Hann, LC–MS/MS analysis of phenols for classification of red wine according to geographic origin, grape variety and vintage, Food Chem. 122 (2010) 366-372.

[11] E. Salvatore, M. Cocchi, A. Marchetti, F. Marini, A. de Juan, Determination of phenolic compounds and authentication of PDO Lambrusco wines by HPLC-DAD and chemometric techniques, Anal. Chim. Acta 761 (2013) 34-45.

[12] M. Arbulu, M.C. Sampedro, A. Gómez-Caballero, M.A. Goicolea, R.J. Barrio, Untargeted metabolomic analysis using liquid chromatography quadrupole time-of-flight mass spectrometry for non-volatile profiling of wines, Anal. Chim. Acta 858 (2015) 32-41.

[13] R. Ragone, P. Crupi, S. Piccinonna, C. Bergamini, F. Mazzone, F. Fanizzi, F. Schena, D. Antonacci, Classification and chemometric study of Southern Italy monovarietal wines based on NMR and HPLC-DAD-MS, Food Sci. Biotechnol. 24 (2015) 817-826.

[14] R. Diaz, M. Ibanez, J.V. Sancho, F. Hernandez, Building an empirical mass spectra library for screening of organic pollutants by ultra-high-pressure liquid chromatography/hybrid quadrupole time-of-flight mass spectrometry, Rapid Commun. Mass Spectrom. 25 (2011) 355-369.

[15] R. Díaz, O.J. Pozo, J.V. Sancho, F. Hernández, Metabolomic approaches for orange origin discrimination by ultra-high performance liquid chromatography coupled to quadrupole time-of-flight mass spectrometry, Food Chem. 157 (2014) 84-93.

[16] J. Rubert, O. Lacina, C. Fauhl-Hassek, J. Hajslova, Metabolic fingerprinting based on high-resolution tandem mass spectrometry: a reliable tool for wine authentication?, Anal. Bioanal. Chem. 406 (2014) 6791-6803.

[17] G. Glauser, N. Veyrat, B. Rochat, J.-L. Wolfender, T.C.J. Turlings, Ultra-high pressure liquid chromatography–mass spectrometry for plant metabolomics: A systematic comparison of high-resolution quadrupole-time-of-flight and single stage Orbitrap mass spectrometers, J. Chromatogr. A, 1292 (2013) 151-159.

[18] H.G. Gika, G.A. Theodoridis, M. Earll, R.W. Snyder, S.J. Sumner, I.D. Wilson, Does the Mass Spectrometer Define the Marker? A Comparison of Global Metabolite Profiling Data Generated Simultaneously via UPLC-MS on Two Different Mass Spectrometers, Anal. Chem. 82 (2010) 8226-8234.

[19] J.-C. Martin, M. Maillot, G. Mazerolles, A. Verdu, B. Lyan, C. Migné, C. Defoort, C. Canlet, C. Junot, C. Guillou, C. Manach, D. Jabob, D.-R. Bouveresse, E. Paris, E. Pujos-Guillot, F. Jourdan, F. Giacomoni, F. Courant, G. Favé, G. Le Gall, H. Chassaigne, J.-C. Tabet, J.-F. Martin, J.-P. Antignac, L. Shintu, M. Defernez, M. Philo, M.-C. Alexandre-Gouaubau, M.-J. Amiot-Carlin, M. Bossis, M. Triba, N. Stojilkovic, N. Banzet, R. Molinié, R. Bott, S. Goulitquer, S. Caldarelli, D. Rutledge, Can we trust untargeted metabolomics? Results of the metabo-ring initiative, a large-scale, multi-instrument inter-laboratory study, Metabolomics 11 (2015) 807-821.

[20] J. Westerhuis, H.J. Hoefsloot, S. Smit, D. Vis, A. Smilde, E.J. van Velzen, J.M. van Duijnhoven, F. van Dorsten, Assessment of PLSDA cross validation, Metabolomics 4 (2008) 81-89.

[21] K.A. Veselkov, L.K. Vingara, P. Masson, S.L. Robinette, E. Want, J.V. Li, R.H. Barton, C. Boursier-Neyret, B. Walther, T.M. Ebbels, I. Pelczer, E. Holmes, J.C. Lindon, J.K. Nicholson, Optimized Preprocessing of Ultra-Performance Liquid Chromatography/Mass Spectrometry Urinary Metabolic Profiles for Improved Information Recovery, Anal. Chem. 83 (2011) 5864-5872.

[22] G. Dervilly-Pinel, S. Weigel, A. Lommen, S. Chereau, L. Rambaud, M. Essers, J.-P. Antignac, M.W.F. Nielen, B. Le Bizec, Assessment of two complementary liquid chromatography coupled to high resolution mass spectrometry metabolomics strategies for the screening of anabolic steroid treatment in calves, Anal. Chim. Acta 700 (2011) 144-154.

Figure captions

Figure 1. Summary of the experimental workflow.

Figure 2. Plots of scores obtained for the PCA analysis for Platforms #1 and #2 at the all features level (a-b), platform specific annotated metabolite level (c-d) and interplatform validated metabolites (e-f). In all cases, samples are coloured by PDO.

Figure 3. (a) Retention time correlation function between both platforms. Represented standards (+) and annotated compounds (•) are shown in Supplementary Tables 3 and 4. (b-d) Extracted areas of the different resveratrol derived markers as a function of their PDO. All areas were normalized to the mean area value of the lower abundant class for each platform to facilitate the comparison. P1 = Platform#1; P2 = Platform#2.

Figure 4. Extracted areas of the catechin related compounds found increased in Penedes (a-b) and Ribera del Duero (c-e) wines. All areas were normalized to the mean area value of the lower abundant class for each platform to facilitate the comparison. P1 = Platform#1; P2 = Platform#2

Figure 1

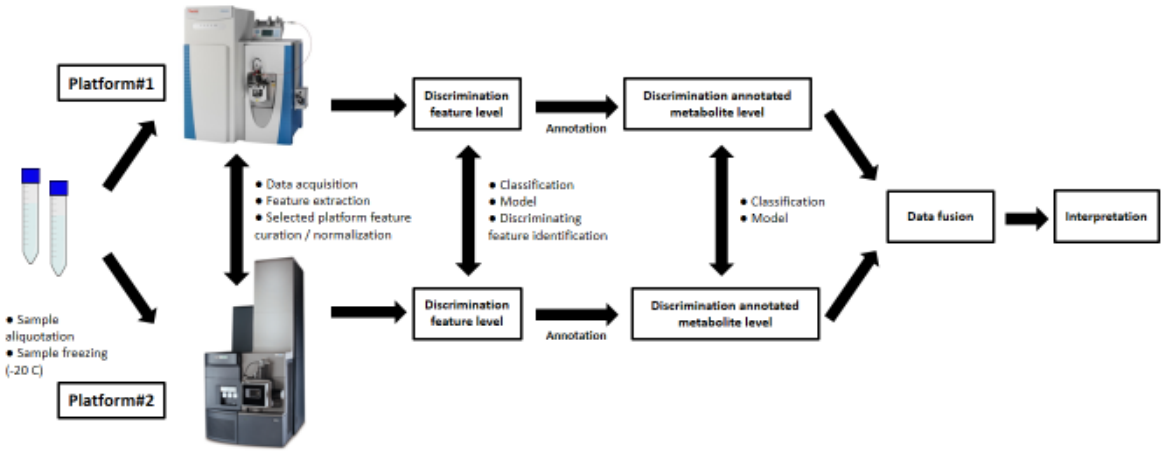


Figure 2

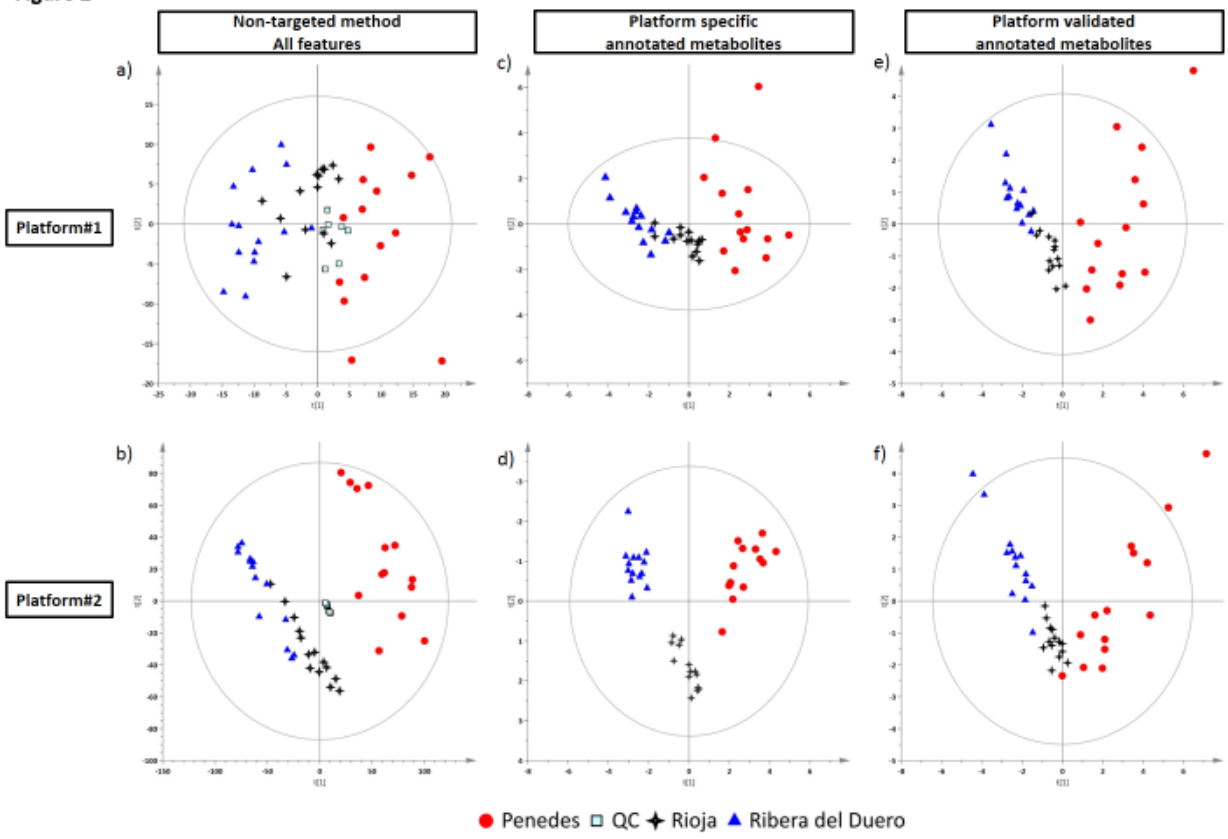


Figure 3

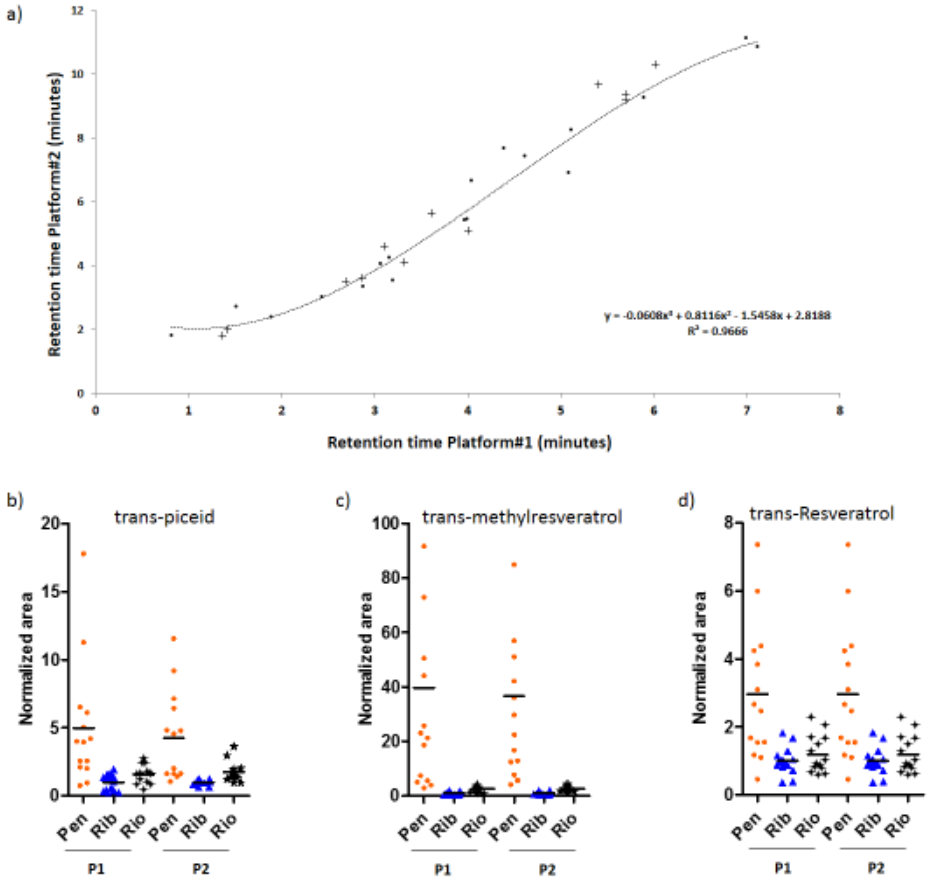
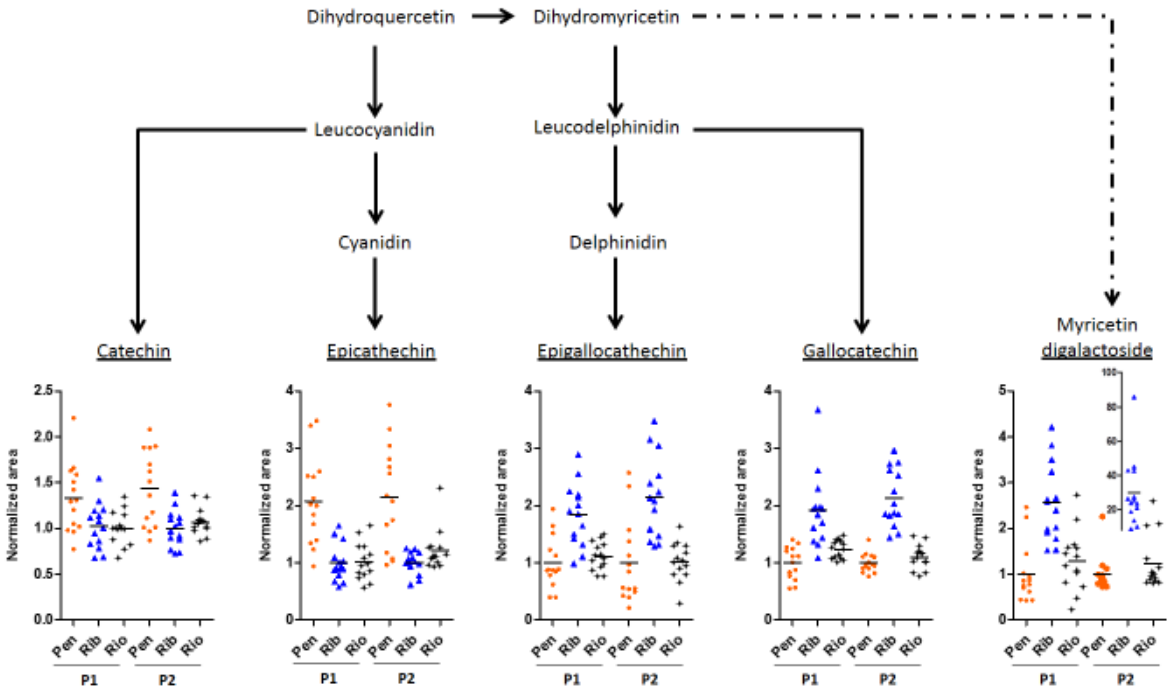


Figure 4



P	FC	p-val	FC	p-val	FC	p-val	FC	p-val	FC	p-val
1	1.31	0.0012	2.06	<0.0001	1.75	<0.0001	1.72	<0.0001	2.24	<0.0001
2	1.39	<0.0001	1.94	<0.0001	2.13	<0.0001	2.03	<0.0001	26.91	<0.0001

Table 1. Summary of OPLS-DA models obtained for Platform 1 (Q-Exactive) and 2 (Q-TOF) working in the negative mode at the extracted feature level.

		Penedes				Ribera				Rioja		
Platform	**LV**	**Q^2 [a]**	**p[b]**	**%CC[c]**	**LV**	**Q^2 [a]**	**p[b]**	**%CC[c]**	**LV**	**Q^2 [a]**	**p[b]**	**%CC[c]**
1	1+1	0.89	$<5x10^{-4}$	100	1+1	0.83	$<5x10^{-4}$	100	1+1	0.72	$<5x10^{-4}$	100
2	1+0	0.86	$<5x10^{-4}$	100	1+1	0.86	$<5x10^{-4}$	100	1+1	0.77	$<5x10^{-4}$	100

a) Q2 obtained by the sevenfold cross-validation procedure obtained in SIMCA

b) Probability to obtained a Q^2 value equal or above to the one obtained for a specific model model when 2000 permutations are performed

c) Percentage of correctly classified samples

Table 2. Metabolites separately identified in platform 1 and 2 and results of the valiadatio (identification) on the other platform. FC = Fold-change. P=Penedes, RD= Ribera del Duero, RIO= Rioja.

Platform 1						
RT (s)	**m/z (Theoretical)**	**Mass Error (ppm)**	**Compound (HMDB number)**	**P.D.O**	**FC (p-val)**	**Validated[c] (FC / p-val)**
199	289.0720 (289.07176)	1.4	Epicathechin[a] (HMDB33973)	P	2.06 (<0.0001)	YES (1.94/<0.0001)
82	173.0449 (173.0455)	-3.5	Shikimic acid (HMDB03070)	P	3.02 (<0.0001)	YES (6.45/9.40E-03)
49	159.0292 (159.0299)	-4.4	2-Methyl-4-oxopentanedioic acid (HMDB39447)	P	1.70 (<0.0001)	NO
48	169.0135 (169.0142)	-4.1	Gallic acid[a] (HMDB05807)	P	1.54 (<0.0001)	NO
85	305.0668 (305.0667)	0.3	Gallocatechin[b] (HMDB38365)	RD	1.72 (<0.0001)	YES (2.03/<0.0001)
161	305.0668 (305.0667)	0.3	Epigallocatechin[b] (HMDB38361)	RD	1.75 (<0.0001)	YES (2.13/<0.0001)
172	329.0880 (329.0878)	0.6	3'-Glucosyl-2',4',6'-trihydroxyacetophenone (HMDB40621)	RD	1.71 (<0.0001)	YES (1.30/<0.0001)
240	389.1238 (389.1242)	-1.0	trans-piceid[a] (HMDB30564)	P	3.87 (<0.0001)	YES (3.10/<0.0001)
273	389.1245 (389.1242)	0	cis-piceid[a] (HMDB31422)	P	4.18 (0.0008)	NO
Platform 2						
562	287.0927 (287.0925)	0.6	5'-Methoxy-O-desmethylangolensin[b] (HMDB41686)	P	4.78 (<0.0001)	YES (62.38/<0.0001)
276	481.0966 (481.0988)	-4.5	(-)-Epigallocatechin 3' (or 7)-glucuronide[b] (HMDB41638 /HMDB41640)	RD	7.99 (<0.0001)	NO
219	165.0186 (165.0193)	-4.3	3,4-Methylenedioxybenzoic acid (HMDB32613)	RIO	3.04 (<0.0001)	NO
618	403.1386 (403.1398)	-3.2	4'-Methylresveratrol 3-glucoside[b] (HMDB34117)	P	32.56 (<0.0001)	YES (22.95/<0.0001)
551	431.1335 (431.1348)	-3.0	Trichocarposide[b] (HMDB31723)	P	10.30 (<0.0001)	YES (19.40/<0.0001)
581	535.1816 (535.1821)	-1.0	Hydroxypinoresinol glucoside (HMDB33281/HMDB33282)	P	17.28 (<0.0001)	YES (3.60/0.0005)
339	641.1349 (641.1359)	-1.7	Myricetin 3,3'-digalactoside[b] (HMDB37850)	RD	30.12 (<0.0001)	YES (2.24/<0.0001)
325	655.1147 (655.1152)	-0.7	Gossypetin 8-glucuronide 3-glucoside (HMDB39103)	RD	26.91 (<0.0001)	NO

[a] Identification based on retention time and MS/MS of analytical standards

[b] Identified using MS/MS fragmentation pattern from databases

[c] Compounds identified on the other platform based on the predicted retention time (See Figure 3a) and MS/MS fragments