H-1 NMR metabolomics applied to Bordeaux red wines

Louis Gougeon, Grégory da Costa, François Guyon, Tristan Richard

^{1}H NMR metabolomics applied to Bordeaux red wines

Louis GOUGEON[1], Gregory DA COSTA[1], François GUYON[2], Tristan RICHARD[1,*]

1. Univ. Bordeaux, ISVV, EA 4577, USC 1366 INRA, Unité de Recherche Œnologie, 210 chemin de Leysotte, F-33882 Villenave d'Ornon, France.

2. Service Commun des Laboratoires, 3 avenue du Dr. Albert Schweitzer, 33600 Pessac, France.

* Corresponding author:

Phone: (+33)5557575957

Email address: tristan.richard@u-bordeaux.fr

Abstract

The q-NMR metabolomics has already demonstrated its potential for classifying wines of different geographical origins, grape varieties, or vintages. This study focuses on the characterisation of Bordeaux red wines, seeking to discriminate them from others produced in the major French wine regions. A sampling of 224 commercial French wines was analysed by ^{1}H NMR and forty compounds were quantified. Non-supervised and supervised statistical analyses revealed a singular imprint of Bordeaux wines in comparison with other French wines, with classification rates ranging from 71% to 100%. Within the Bordeaux vineyards, red wines from the different Bordeaux subdivisions were analysed from different vintages. Our results indicate that q-NMR metabolomics enables the differentiation of Médoc and Libournais vineyard highlighting the most discriminant constituents. In addition, the effects of wine evolution during bottle aging and vintage on Bordeaux red wines were pointed out and discussed.

Keywords

Wine classification; red wines; Bordeaux wines; q-NMR; multivariate statistics

1. Introduction

According to the Conseil Interprofessionnel du Vin de Bordeaux (CIVB), Bordeaux wines are unchallenged leaders of overall exports of French Protected Designation of Origin (PDO) in 2016, representing 39% of the total volume and 46% of the total value. In total, there are nearly sixty PDO in Bordeaux, distributed in five main subdivisions: Médoc, Graves, Libournais, Entre-deux-Mers, and Blaye-Bourg. Three-quarters of the vineyard are destined for the cultivation of red grapes. Merlot is the main red cultivar followed by Cabernet Sauvignon and Cabernet Franc, to which minor red grape varieties such as Malbec, Petit Verdot or Carmenere are added. The goal of this work is to identify specific chemical markers allowing to characterize Bordeaux red wines. Specific fingerprintings of Bordeaux red wines were investigated by comparison with wines from others French regions, and classification was refined in order to establish distinctions also among Bordeaux wines.

With the rise of metabolomics, many studies have been focussed on wine classification combining quantitative analytical data with advanced mathematical and statistical techniques (Geana, Popescu, Costinel, Dinca, Ionete, Stefanescu, et al., 2016). To achieve this goal different instrumental approaches could be used including liquid and gas chromatographies coupled with different type of detectors such as infrared spectroscopy, mass spectroscopy, Raman spectroscopy, ultra-low radioactivity or NMR (Médina, Salagoïty, Guyon, Gaye, Hubert, & Guillaume, 2013). For over ten years now, quantitative ^{1}H NMR spectrometry (q-NMR) has been the object of numerous developments concerning foods and beverages analysis (Bharti & Roy, 2012; Pauli, Gödecke, Jaki, & Lankin, 2012), and enological practices in particular (Amargianitaki & Spyros, 2017; Hong, 2011). Depending on the objectives, the classifications focused on grape varieties, geographic origin, vintage (Anastasiadi, Zira, Magiatis, Haroutounian, Skaltsounis, & Mikros, 2009; Godelmann, Fang, Humpfer, Schütz, Bansbach, Schäfer, et al., 2013) or winemaking techniques (López-Rituerto,

Cabredo, López, Avenoza, Busto, & Peregrina, 2009; Mazzei, Spaccini, Francesca, Moschetti, & Piccolo, 2013). This versatile and high resolution technic provides quickly access to the simultaneous determination of several wine constituents without pretreatment. Together with multivariate statistical techniques, q-NMR has become a recognized technique for wine traceability whose effectiveness has been confirmed by several scientific publications (Godelmann, et al., 2013; Gougeon, Da Costa, Le Mao, Ma, Teissedre, Guyon, et al., 2018; Gougeon, Da Costa, Richard, & Guyon, 2019; López-Rituerto, Savorani, Avenoza, Busto, Peregrina, & Engelsen, 2012; Papotti, Bertelli, Graziosi, Silvestri, Bertacchini, Durante, et al., 2013; Son, Hwang, Kim, Ahn, Park, Van Den Berg, et al., 2009).

In the present study, red wines from main French DPO (Bordeaux, Beaujolais, Burgundy, Côtes du Rhône, Languedoc-Roussillon and Loire Valley) and different vintages (2004 to 2017) were investigated by q-NMR spectroscopy coupled to multivariate statistical analysis. In the context of blended wines such Bordeaux wines, we evaluated the ability of q-NMR to discriminate: (i) Bordeaux wines from the other French PDO, (ii) Bordeaux appellations. In addition, the effects of wine evolution during bottle aging and vintage on Bordeaux red wines were investigated.

2. Material and Methods

2.1 Sample collection

A total of 224 commercial French red wines from 2004 to 2017 vintages were collected. Their classification is shown on Table S1. Six french wine-producing regions are represented: Bordeaux (n=127), Beaujolais (n=15), Burgundy (n=20), Côtes du Rhône (n=20), Languedoc-Roussillon (n=24) and Loire Valley (n=18). Inside Bordeaux sampling, six subdivisions are distinguished: Generic (n=15), Blaye-Bourg (n=5), Entre-deux-Mers (n=8), Graves (n=16),

Libournais (n=36) and Médoc (n=47). Wines were stored a 4°C and analysed under a week after bottle opening.

2.2 Sample preparation

Wine samples were prepared according to our previous work (Gougeon, et al., 2018). Briefly, after centrifugation, 420 µL of wine were directly mixed with 120 µL of phosphate solution (1M, pH 2.6), and 60 µL of a D_2O solution containing trimethylsilylpropanoic acid sodium salt (TMSP) were used as the frequency reference and calcium formate was used as internal standard for quantitation. The pH was fine-adjusted to 3.1 using a semi-automatic small scale system (BTpH, Bruker BioSpin, Germany) using a 1M HCl solution.

2.3 ^{1}H NMR spectra acquisition

^{1}H NMR spectra acquisitions were performed on a 600 MHz spectrometer (Bruker BioSpin). NMR analysis was carried out a constant temperature of 293°K. Three pulse sequences were used to collected the spectra, as described in a previous study (Gougeon, et al., 2018). The number of sampling points was 64 K using a 20.0229 ppm spectral width. The relaxation delay was set to 5 s and the sampling time was set to 2.726 s.

2.4 ^{1}H NMR spectra post-acquisition treatments

All ^{1}H NMR spectra were aligned thanks to TMPS signal ($\delta = 0$ ppm). An exponential weighting function corresponding to 0.3 Hz line broadening was applied before applying Fournier transformation using Topspin® software (version 3.2, Bruker Biospin, Germany). Phase correction was performed manually to have the greatest reproducibility (Bharti & Roy, 2012). Baseline correction was performed automatically with MestReNova® NMR software (version 11.0.3, Mestrelab Research, Spain)

2.5 Compound quantification

Forty compounds were identified on the ^{1}H NMR spectra (Table S2). Quantitation was performed semi-automatically on the MestReNova using the plugin Simple Mixture Analysis (SMA). Validation method, LOD, LOQ, CV was described in our previous study (Gougeon, et al., 2018).

2.6 Chemometrics

For statistical analysis, the data were normalized and then introduced into RStudio 1.1.447 and BioStatFlow 2.8.3 for pattern recognition. Principal components analysis (PCA) was performed to visualize the acquired data and observe the discrete trend between samples. To create a more reasonable model, partial least squares discriminant analysis combined to orthogonal signal correction filters (OSC-PLS-DA) was performed to sharpen the separation between observations groups (Wehrens, 2011). OSC-PLS-DA is also helpful to understand which component carries the class separating information, in addition with one-way ANOVA. The quality of the OSC-PLS-DA model was indicated by R^2Y and Q^2 metrics. The R^2Y describes the percentage of variation explained by the model and Q^2 indicates the predictive ability of the model (Worley & Powers, 2013). The performance of models was evaluated by internal leave one-out cross validation (LOOCV) (Riedl, Esslinger, & Fauhl-Hassek, 2015).

3. Results and discussion

French wines are the products of different terroirs, grape varieties and viti-vinicultural practices. In this study, 224 French commercial wine samples mainly from Bordeaux area (n=127) and from others French wine-producing regions (n=97) were analysed by ^{1}H NMR

metabolomics. ^{1}H NMR spectra were acquired similarly to our previous article with only a slight adjustment concerning pH control (Gougeon, et al., 2018). Compounds were quantified by targeted analysis using the Global Spectral Deconvolution method (GSD) (Cobas, Seoane, Domínguez, Sykora, & Davies, 2011), allowing the semi-automatized quantification of forty molecules listed in Table S2. Multivariate analyses were used to discriminate the wine key features concerning their geographical origins, grape varieties or vintages. For wine separation, PCA was performed and followed by OSC-PLS-DA to create a more reasonable regression model. Orthogonal signal correction is a PLS-based data filtering technique that removes the uncorrelated information, and consequently it is more exclusively focussed on the variables of interest (Gavaghan, Wilson, & Nicholson, 2002). In addition, the model predictivity was evaluated by internal leave one-out cross validation (LOOCV). In the LOOCV procedure, each training set is created by randomly taking all the samples except one and the test set is the sample left out. Thus, for N samples, N different training sets and N different test sets are created. Each sample is predicted once, which provides a conservative estimate of the prediction ability of the PCA model (Forina, Lanteri, Casolino, & Oliveri, 2004).

3.1. Classification of Bordeaux versus French wines

The 127 Bordeaux wine samples from six different appellations (Bordeaux generic, Blaye and Bourg, Entre-deux-Mers, Graves, Libournais, and Médoc) and from thirteen different vintages (ranged from 2004 to 2016) were compared to the 97 wines from other French geographic areas (Beaujolais, 15; Burgundy, 20; Côtes du Rhône, 20; Languedoc-Roussillon, 24; Loire Valley, 18) from different vintages (ranged from 2004 to 2017) by OCS-PLS-DA (Figure 1). A tendency to discriminate Bordeaux wines from other French wines is clearly observed (R^2Y and Q^2 value of 0.75 and 0.82, respectively) despite the wide range of vintages analysed. Most

Bordeaux red wines are located in negative PC1 values, whereas others French wines are placed in positive PC1 values (Figure 1A). To quantify the predictive ability of the model, the leave one-out cross validation (LOOCV) was used. The overall correct classification of the wine origin for Bordeaux wines versus French wines is near 95% (Table S3). This highlights the specificity of Bordeaux red wines in comparison to all other French wines. As shown in Figure 1C, Bordeaux red wines contain more proline (mean, 1.4; range, 0.5–4.1 g/L), phenethyl alcohol (mean, 98; range, 58–143 mg/L), succinic acid (mean, 1.8; range, 1.3–2.7 g/L), gallic acid (mean, 125; range, 69–182 mg/L), arabinose (mean, 0.3; range, 0.0–0.6 g/L), galacturonic acid (mean, 1.8; range, 0.4–3.1 g/L), methanol (mean, 181; range, 54–365 mg/L) and isopentanol (mean, 242; range, 132–366 mg/L). On the contrary, they contain less lactic acid (mean, 0.7; range, 0.4–1.0 g/L), caffeic acid (mean, 3; range, 0.0–22 mg/L), ethyl lactate (mean, 85; range, 16–183 mg/L) and 2,3-butanediol (mean, 363; range, 217–655 mg/L) than others French wines.

Many of these compounds, such as proline, phenethyl alcohol, gallic acid or succinic acid could be directly or indirectly associated to grape variety. Proline is one of the most abundant amino acid in grape and wine. Its content in wine was correlated to grape varieties (Son, Kim, van den Berg, Hwang, Park, Lee, et al., 2008). Cabernet Sauvignon and Merlot, which are the main varieties of Bordeaux region, are known to be rich in proline (Huang & Ough, 1991). Phenethyl alcohol is a highly aromatic alcohol synthetized by bacteria, fungi, and yeasts from L-phenylalanine (Etschmann, Sell, & Schrader, 2003). Its content in wines depends of alcoholic fermentation but also on L-phenylalanine level in grapes. Son et al. already highlighted that phenethyl alcohol could be a chemical marker of grape variety (Son, et al., 2008). They found higher contents in Cabernet Sauvignon wines than in Shiraz wines. But not only the grape variety influences the content of phenetyl alcohol, the same authors also observed significantly higher phenethyl alcohol levels in Cabernet Sauvignon wines produced

191 in Australia than in those produced in France or in South Korea. Gallic acid has been
192 observed to discriminate against wines grape variety (Hu, Yue, Zhu, Wen, Zhang, & Hardie,
193 2015). In particular, in agreement with our results, it has been shown that Cabernet Sauvignon
194 wines contain high levels of gallic acid (Zhu, Hu, Lu, & Xu, 2018). Succinic acid is produced
195 by yeasts during fermentation from sugar or amino acids (Coulter, Godden, & Pretorius,
196 2004). It was pointed out as a discriminant factor of grape variety (Hu, Yue, Zhu, Wen,
197 Zhang, & Hardie, 2015). Moreover its content was affected by vintage (Ali, Maltese, Toepfer,
198 Choi, & Verpoorte, 2011) and winemaking practices (Mazzei, Spaccini, Francesca, Moschetti,
199 & Piccolo, 2013). As for succinic acid, the contents of some compounds could be associated
200 to the influence of different factors. Low lactic acid levels observed on Bordeaux wine
201 suggest an indirect climate influence. Lactic acid is produced by malolactic fermentation from
202 malic acid which is directly linked to the ripeness of the grape berry, so to climate (Lonvaud-
203 Funel, 1999). Caffeic acid has already been identified as a discriminative factor for wines
204 made from vines treated by different cultural processes (De Pascali, Coletta, Del Coco, Basile,
205 Gambacorta, & Fanizzi, 2014). Wines from cover crop vines had higher levels of caffeic acid
206 than vines that had undergone soil tillage. It would seem that cultural practices are potentially
207 responsible for the differences observed in our case. Finally, methanol, galacturonic acid, and
208 arabinose are present in high levels specifically in Bordeaux wines (Figure 1C). These three
209 compounds are connected with skin degradation on the must. Indeed, galacturonic acid and
210 arabinose are two of the building blocks of pectins, which form the skin berry (Müller-
211 Maatsch, Bencivenni, Caligiani, Tedeschi, Bruggeman, Bosch, et al., 2016). Methanol is
212 produced through pectin hydrolysis by enzymes (Revilla & González-SanJosé, 1998). This
213 observation suggests a pre-fermentation process effect on Bordeaux red wines. A possible
214 explanation can be the will of produce long-keeping wines, which means an extended
215 maceration to extract the maximum content of polyphenols. So there is a set of factors that

configure the specificity of Bordeaux red wines including grape varieties, climate and winemaking practices.

To confirm that Bordeaux wines have a singular fingerprinting, pairwise comparisons between Bordeaux wines and those of the others French wine-producing regions were performed by unsupervised statistical analysis (PCA) as shown on Figure 2. Biplots indicate separation and the overall correct prediction rates for Bordeaux wines were ranged between 71 to 100%, for Languedoc and Burgundy wines, respectively (Table S3). These estimates are comparable with those observed by Godelmann et al. (2013), where geographical origin of wines from five German production areas could be predicted 89% correctly in average, with rates ranging from 59 to 100% (Godelmann, et al., 2013).

The grape variety, independently from vintage, seems to be one of the main discriminative factors between Bordeaux wines and those of other French wine-producing regions. Bordeaux wines were produced with major parts of Cabernet Sauvignon and Merlot, whereas wines from other French regions were produced with different varieties: Pinot noir for Burgundy, Gamay for Beaujolais, and Cabernet Franc for Loire Valley wines. Role of grape varieties in wine discrimination was already proved in several studies (Anastasiadi, Zira, Magiatis, Haroutounian, Skaltsounis, & Mikros, 2009; Fan, Zhong, Fauhl-Hassek, Pfister, Horn, & Huang, 2018; Son, et al., 2008). Nevertheless, wines from regions using the same varieties than Bordeaux, such as Languedoc, were discriminated. Cultural and fermentation practices seem to have also an important role concerning wine classification. As previously mentioned, some discriminant compounds could be directly or indirectly linked to these practices such as caffeic acid, methanol, galacturonic acid, and arabinose.

3.2. Comparison of Bordeaux appellations

In order to investigate Bordeaux wines, NMR metabolomics was used to discriminate wines from different subdivisions of Bordeaux (Graves, Libournais and Médoc). In addition, the effect of wine aging in bottle and vintage were investigated on these 127 Bordeaux wine samples.

3.2.1. Geographic origin

Bordeaux vineyard can be expanded in five main subdivisions: Blaye and Bourg, Entre-deux-Mers, Graves, Libournais and Médoc. These areas differ with respect to their soil, mesoclimate and grape varieties. To examine the capacity of NMR combined to multivariate statistical analysis to distinguish subdivisions, the Bordeaux wine samples from Graves (n=16), Libournais (n=36) and Médoc (n=47) areas and different vintages (ranged from 2004 to 2014) were compared.

Initially, PCA was performed over all of the wines from the three appellations (Figure S1). On one hand, the PCA score plot between the three appellations showed separation by the first principal component, with overlapping of the Graves samples, resulting in low predictability. The statistical analysis revealed that Graves wines are not discriminated neither from Libournais nor Médoc wines. These data seem to preclude the discrimination of Graves wines from the two others Bordeaux appellations. On the other hand, unsupervised classification suggests a trend to distinguish between Libournais and Médoc. To confirm this observation, an OSC-PLS-DA was performed only with data created with Libournais and Médoc wines. Results are shown in Figure 3. The OSC-PLS-DA score plots revealed clear separation between Libournais and Médoc wines with high values of R^2Y and Q^2, of 0.77 and 0.68, respectively. The model was further validated using LOOCV. The mean overall correct classification of Médoc versus Libournais wines was 85% (83 and 87% for Médoc and

Libournais, respectively) (Table S4). Among the discriminant parameters Médoc wines contained more shikimic and lactic acids but less proline and ethanol (Figure 3).

Average concentration of shikimic acid was 69 mg/L (range 23-95 mg/L) in Médoc wines whereas, in Libournais wines, average concentration was 20 mg/L (range 0-55 mg/L). Shikimic acid is a phenolic compound identified as a chemical marker of grape varieties (Godelmann, et al., 2013). Even if the Bordeaux wines are blended from different grape varieties, Cabernet Sauvignon is the dominant variety in Médoc PDO's and Merlot Libournais appellation. It has been reported that shikimic acid levels in Cabernet Sauvignon wines are higher than in Merlot wines (Mardones, Hitschfeld, Contreras, Lepe, Gutiérrez, & von Baer, 2005). Our results are in agreement with this observation. Concerning the proline contents, they reached 2.0 g/L on average (range 1.1-4.1 g/L) in the Libournais wines against 1.3 g/L (range 0.5-1.9 g/L) in the Médoc ones. Huang et al. have shown that Merlot contained more proline than Cabernet Sauvignon, with 1.4 and 0.8 g/L in a Merlot and Cabernet Sauvignon grape juice, respectively (Huang & Ough, 1991). This result is also in agreement with the distribution of grape varieties in the two areas. Both shikimic acid and proline levels suggest a grape variety effect on the observed wines discrimination. Concerning lactic acid, our results indicate that Médoc wine contain more lactic acid (mean, 0.73; range, 0.55–1.02 g/L) than Libournais ones (mean, 0.57; range, 0.39–0.77 g/L). As mentioned previously, lactic acid contents can be linked to malic acid levels in grapes, so indirectly linked with grape ripeness. The ethanol contents observed are consistent with this hypothesis. Ethanol levels are significantly lower in Médoc wines (mean 12.6, range 11.2-13.6%) than in Libournais wines (mean 13.4, range 12.3-14.4%). Therefore, it would seem that the combination of grape varieties, climate and cultural practices contributed to discriminate these close geographical designations as previously observed by Pereira et al. on Bordeaux's monovarietal wines (Pereira, Gaudillère, Van Leeuwen, Hilbert, Maucourt, Deborde, et al., 2007).

291

292 *3.2.2. Wine aging and vintages*

293 To observe the effects of vintage and wine evolution during bottle aging, the NMR-data of the
294 Bordeaux red wines (n=127) obtained from thirteen different vintages (2004 to 2016) were
295 analysed. First, to observe the influence of wine aging independently of vintage, the youngest
296 wines (n=28, vintages 2013 to 2016) were compared to the oldest wines (n=37, vintages 2004
297 to 2007) by OSC-PLS-DA (Figure 4). The analysis revealed a clear separation among the two
298 groups with R^2Y and Q^2 values of 0.75 and 0.78, respectively. The main compounds
299 responsible for these differences are xylose, epicatechin, catechin, tyrosine and citric acid
300 which are more present in young wines, and ethyl acetate, ethyl lactate, fructose, caffeic acid,
301 syringic acid, succinic acid and shikimic acid which are more present in older wines (Figure
302 4C). Among the compounds responsible for the discrimination of the wines some could be
303 directly connected to wine evolution during aging in bottle. Catechin and epicatechin are
304 involved in a series of polymerization reactions with different compounds inducing a decrease
305 of the free-compounds during aging due to precipitation. Similarly to our results, Cassino et
306 al. observed the same pattern for xylose with a decrease during wine aging in bottles and an
307 increase of esters (ethyl acetate and ethyl lactate) (Cassino, Tsolakis, Bonello, Gianotti, &
308 Osella, 2018). These results indicate a clear evolution during bottle aging that will influence
309 the discrimination between wines at a given point in time in addition to the differences
310 between vintages.

311 In order to observe the specific effects of wine evolution and vintage, the six different
312 vintages with more than ten wine samples were submitted to multiple statistical approaches
313 (2005, n=10; 2007, n=12; 2010, n=10; 2011, n=15; 2012; n=23 and 2013, n=10). Vintages
314 were pairwise compared by OSC-PLS-DA to sharpen the observed separations and followed
315 by cross-validations using LOOCV. Figure 5 shows the comparison between 2010 and 2011

316 vintages. OSC-PLS-DA was able to separate these vintages (R^2Y = 0.81 and Q^2 = 0.83) giving
317 a mean overall correct classification of 75% (70 and 80% for 2010 and 2011, respectively
318 (Table S5). The same procedure was applied to compare one vintage to another (Table S5)
319 giving overall correct classifications ranging from 59% (2012 versus 2013) to 100% (2005
320 versus 2010). Overall, it could be noted that the more vintages are distinct the higher are the
321 classification scores. The average correct classification percentages of 2005 and 2007
322 vintages versus the other ones (2010 to 2013) are 87 and 90%, respectively. In contrast, these
323 percentages are reduced to 61 and 59% for comparisons between 2010 versus 2011 and 2012
324 versus 2013, respectively. This could be due to the wine evolution during bottle aging at a
325 given point in time, as previously mentioned. A wide span of time between vintages facilitates
326 their distinction.
327 To reduce the effect of wine aging in bottle and investigate the specific effect of vintage, the
328 four successive vintages 2010 to 2013 were analysed. Compared to 2011, 2010 vintages
329 presented more sugars and 2,3-butanediol, but less tyrosine, threonine, alanine, catechin,
330 xylose and lactic acid (Figure 5). In Bordeaux, 2010 was a good year for viticulture, with a
331 weather allowing good grape ripeness, which could be an explanation for upper level of
332 sugars and lower level of lactic acid compared to wines from 2011. High levels of amino
333 acids on 2011 could be explained by more rainfall inducing greatest roots absorption. In the
334 same manners levels of amino acids and xylose in 2011 wines were increased in comparison
335 to those of 2012 (Figure S2). Moreover, glycerol, isobutanol and ethyl acetate are negatively
336 correlated to wines from 2012 vintage, but lactic acid, catechin, epicatechin and 2,3-
337 butanediol are positively correlated. Finally, 2013 was a bad vintage in Bordeaux inducing
338 grape ripeness difficulties highlighted by low values of technologic ripeness parameters such
339 as sugars to acids ratio. Discriminative variables of 2013 wines compared to wines from 2012
340 are low levels of ethanol, which confirm low sugars levels, and myo-inositol (Figure S3). In

addition, high malic acid levels confirm the bad technological ripeness. All these observations confirm the climatic influence on wine discrimination between successive vintages.

4. Conclusion

In this study, 224 commercial wines produced in the six major French wine regions were analysed using ^{1}H NMR experiments and multivariate analysis. Forty compounds were quantified by conventional targeted analysis. Advanced data analysis and chemometrics allowed the discrimination of wines on different levels. Bordeaux wines present a singular brand in comparison to the five other major French producing areas. Despite the differences within each geographical designation, NMR spectrometric analyses coupled to multivariate approaches allows the discrimination of a specific area. In addition, the analysis shows that it is possible to discriminate the two major Bordeaux sub-regions Libournais and Médoc. The grape variety composition alone does not explain all the differences observed. Soil and viticultural practices also have a significant influence, as evidenced by the chemical diversity of the molecular markers that discriminate these wines. Moreover, the differentiation of wines from different vintages is made possible in a relatively limited geographical area such as Bordeaux wines by combination of different factors including wine aging and climatic influences. Two successive vintages could be distinguished using specific markers of vitivinicultural conditions. In addition, the wine evolution during bottle aging induces the variation of some specific markers such as polyphenolic compounds. Nevertheless, the analysis during aging merit further study to monitor the wine evolution over time.

Conflict of interest

The authors have no conflicts of interest to declare.

366 **Acknowledgments**

367 The authors thank Annick Moing for their scientific advices concerning statistical analysis.
368 Elodie Guittard and Caroline Baz are warmly thanked for their collaboration in the sample
369 collection. This work was supported by the Bordeaux Metabolome Facility and MetaboHUB
370 (ANR-11-INBS-0010 project), the Conseil Régional d'Aquitaine, Conseil Interprofessionnel
371 du Vin de Bordeaux (CIVB) and FranceAgriMer program (grant number 2014–0785).

372

373 **Appendix A. Supplementary data**

374

375 **Figure legends**

376 **Figure 1.** OSC-PLS-DA (A) score plot, (B) loading plot, and (C) one way ANOVA
377 performed with normalised data on 12 most discriminant parameters based on q-NMR
378 normalized data from Bordeaux (red, n=127) versus other French regions (blue, n=97) red
379 wines (validation parameters of the model: $Q^2 = 0.82$ and $R^2Y = 0.75$).

380

381 **Figure 2.** Principal component analysis (PCA) score-plot based on q-NMR data of Bordeaux
382 red wines (blue, n=127) and versus those from (A) Beaujolais (n=15), (B) Burgundy (n=20),
383 (C) Côtes du Rhône (n=20), (D) Languedoc-Roussillon (n=24), and (E) Loire Valley (n=18).

384

385 **Figure 3.** OSC-PLS-DA (A) score plot, (B) loading plot, and (C) one way ANOVA
386 performed with normalised data on 16 most discriminant parameters based on normalized q-
387 NMR data from Libournais (red, n=36) versus Médoc (blue, n=47) samples (validation
388 parameters of the model: $Q^2 = 0.68$ and $R^2Y = 0.77$).

389

390 **Figure 4.** OSC-PLS-DA (A) score plot, (B) loading plot, and (C) one way ANOVA
391 performed with normalised data on 12 most discriminant parameters based on q-NMR
392 normalized data from young vintages (red, 2013 to 2016, n=28) versus old ones (blue, 2004 to
393 2007, n=37) of Bordeaux red wines (validation parameters of the model: $Q^2 = 0.75$ and $R^2Y =$
394 0.78).

395

396 **Figure 5.** OSC-PLS-DA (A) score plot, (B) loading plot, and (C) one way ANOVA
397 performed with normalised data on 11 most discriminant parameters based on qNMR
398 normalized data from 2010 (red, n=10) versus 2011 vintage (blue, n=15) of Bordeaux red
399 wines (validation parameters of the model: $Q^2 = 0.83$ and $R^2Y = 0.81$).

References

Ali, K., Maltese, F., Toepfer, R., Choi, Y. H., & Verpoorte, R. (2011). Metabolic characterization of Palatinate German white wines according to sensory attributes, varieties, and vintages using NMR spectroscopy and multivariate data analyses. *Journal of Biomolecular NMR, 49*(3-4), 255-266.

Amargianitaki, M., & Spyros, A. (2017). NMR-based metabolomics in wine quality control and authentication. *Chemical and Biological Technologies in Agriculture, 4*(1), 9.

Anastasiadi, M., Zira, A., Magiatis, P., Haroutounian, S. A., Skaltsounis, A. L., & Mikros, E. (2009). 1H NMR-Based Metabonomics for the Classification of Greek Wines According to Variety, Region, and Vintage. Comparison with HPLC Data. *Journal of Agricultural and Food Chemistry, 57*(23), 11067-11074.

Bharti, S. K., & Roy, R. (2012). Quantitative 1H NMR spectroscopy. *TRAC Trends in Analytical Chemistry, 35*, 5-26.

Cassino, C., Tsolakis, C., Bonello, F., Gianotti, V., & Osella, D. (2018). Wine evolution during bottle aging, studied by 1H NMR spectroscopy and multivariate statistical analysis. *Food Research International*, doi:10.1016/j.foodres.2018.1008.1075.

Cobas, C., Seoane, F., Domínguez, S., Sykora, S., & Davies, A. N. (2011). A new approach to improving automated analysis of proton NMR spectra through Global Spectral Deconvolution (GSD). *Spectroscopy Europe, 23*(1), 26-30.

Coulter, A. D., Godden, P. W., & Pretorius, I. S. (2004). Succinic acid-how is it formed, what is its eff ect on titratable acidity, and what factors infl uence its concentration in wine? *Australian and New Zealand Wine Industry Journal, 19*, 16-25.

De Pascali, S. A., Coletta, A., Del Coco, L., Basile, T., Gambacorta, G., & Fanizzi, F. P. (2014). Viticultural practice and winemaking effects on metabolic profile of Negroamaro. *Food Chemistry, 161*(Supplement C), 112-119.

Etschmann, M. M. W., Sell, D., & Schrader, J. (2003). Screening of yeasts for the production of the aroma compound 2-phenylethanol in a molasses-based medium. *Biotechnology Letters, 25*(7), 531-536.

Fan, S., Zhong, Q., Fauhl-Hassek, C., Pfister, M. K. H., Horn, B., & Huang, Z. (2018). Classification of Chinese wine varieties using 1H NMR spectroscopy combined with multivariate statistical analysis. *Food Control, 88*, 113-122.

Forina, M., Lanteri, S., Casolino, C., & Oliveri, P. C. I. H. o. F. A. e. e., pp. 429–457). Leo M.L. Nollet. (2004). Chemometrics. In L. M. L. Nollet (Ed.), *Handbook of Food Analysis (2e ed.)*, vol. 3 (pp. 1757-1804). New York: Marcel Dekker, Inc.

Gavaghan, C. L., Wilson, I. D., & Nicholson, J. K. (2002). Physiological variation in metabolic phenotyping and functional genomic studies: use of orthogonal signal correction and PLS-DA. *FEBS Letters, 530*(1-3), 191-196.

Geana, E. I., Popescu, R., Costinel, D., Dinca, O. R., Ionete, R. E., Stefanescu, I., Artem, V., & Bala, C. (2016). Classification of red wines using suitable markers coupled with multivariate statistic analysis. *Food Chemistry, 192*, 1015-1024.

Godelmann, R., Fang, F., Humpfer, E., Schütz, B., Bansbach, M., Schäfer, H., & Spraul, M. (2013). Targeted and Nontargeted Wine Analysis by 1H NMR Spectroscopy Combined with Multivariate Statistical Analysis. Differentiation of Important Parameters: Grape Variety, Geographical Origin, Year of Vintage. *Journal of Agricultural and Food Chemistry, 61*(23), 5610-5619.

Gougeon, L., Da Costa, G., Le Mao, I., Ma, W., Teissedre, P.-L., Guyon, F., & Richard, T. (2018). Wine Analysis and Authenticity Using 1H-NMR Metabolomics Data: Application to Chinese Wines. *Food Analytical Methods, 11*(12), 3425-3434.

Gougeon, L., Da Costa, G., Richard, T., & Guyon, F. (2019). Wine Authenticity by Quantitative 1H NMR Versus Multitechnique Analysis: a Case Study. *Food Analytical Methods*, doi: 10.1007/s12161-12018-01425-z.

Hong, Y.-S. (2011). NMR-based metabolomics in wine science. *Magnetic Resonance in Chemistry, 49*(S1), S13-S21.

Hu, B., Yue, Y., Zhu, Y., Wen, W., Zhang, F., & Hardie, J. W. (2015). Proton Nuclear Magnetic Resonance-Spectroscopic Discrimination of Wines Reflects Genetic Homology of Several Different Grape (V. vinifera L.) Cultivars. *PloS One, 10*(12), e0142840.

Huang, Z., & Ough, C. S. (1991). Amino Acid Profiles of Commercial Grape Juices and Wines. *American Journal of Enology and Viticulture, 42*(3), 261-267.

Lonvaud-Funel, A. (1999). Lactic acid bacteria in the quality improvement and depreciation of wine. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology, 76*(1-4), 317-331.

López-Rituerto, E., Cabredo, S., López, M., Avenoza, A., Busto, J. H., & Peregrina, J. M. (2009). A thorough study on the use of quantitative 1H NMR in Rioja red wine fermentation processes. *Journal of Agricultural and Food Chemistry, 57*(6), 2112-2118.

López-Rituerto, E., Savorani, F., Avenoza, A., Busto, J. H., Peregrina, J. M., & Engelsen, S. B. (2012). Investigations of la Rioja terroir for wine production using 1H NMR metabolomics. *Journal of Agricultural and Food Chemistry, 60*(13), 3452-3461.

Mardones, C., Hitschfeld, A., Contreras, A., Lepe, K., Gutiérrez, L., & von Baer, D. (2005). Comparison of shikimic acid determination by capillary zone electrophoresis with direct and indirect detection with liquid chromatography for varietal differentiation of red wines. *Journal of Chromatography A, 1085*(2), 285-292.

Mazzei, P., Spaccini, R., Francesca, N., Moschetti, G., & Piccolo, A. (2013). Metabolomic by 1H NMR Spectroscopy Differentiates “Fiano Di Avellino” White Wines Obtained with Different Yeast Strains. *Journal of Agricultural and Food Chemistry, 61*(45), 10816-10822.

Médina, B., Salagoïty, M. H., Guyon, F., Gaye, J., Hubert, P., & Guillaume, F. (2013). 8 - Using new analytical approaches to verify the origin of wine. In P. Brereton (Ed.), *New Analytical Approaches for Verifying the Origin of Food*, (pp. 149-188): Woodhead Publishing.

Müller-Maatsch, J., Bencivenni, M., Caligiani, A., Tedeschi, T., Bruggeman, G., Bosch, M., Petrusan, J., Van Droogenbroeck, B., Elst, K., & Sforza, S. (2016). Pectin content and composition from different food waste streams. *Food Chemistry, 201*, 37-45.

Papotti, G., Bertelli, D., Graziosi, R., Silvestri, M., Bertacchini, L., Durante, C., & Plessi, M. (2013). Application of One- and Two-Dimensional NMR Spectroscopy for the Characterization of Protected Designation of Origin Lambrusco Wines of Modena. *Journal of Agricultural and Food Chemistry, 61*(8), 1741-1746.

Pauli, G. F., Gödecke, T., Jaki, B. U., & Lankin, D. C. (2012). Quantitative 1H NMR. Development and potential of an analytical method: an update. *Journal of Natural Products, 75*(4), 834-851.

Pereira, G. E., Gaudillère, J. P., Van Leeuwen, C., Hilbert, G., Maucourt, M., Deborde, C., Moing, A., & Rolin, D. (2007). 1 H-NMR metabolic profiling of wines from three cultivars, three soil types and two contrasting vintages. *Journal International des Sciences de la Vigne et du Vin, 41*(2), 103-109.

Revilla, I., & González-SanJosé, M. L. (1998). Methanol release during fermentation of red grapes treated with pectolytic enzymes. *Food Chemistry, 63*(3), 307-312.

Riedl, J., Esslinger, S., & Fauhl-Hassek, C. (2015). Review of validation and reporting of non-targeted fingerprinting approaches for food authentication. *Analytica Chimica Acta, 885*, 17-32.

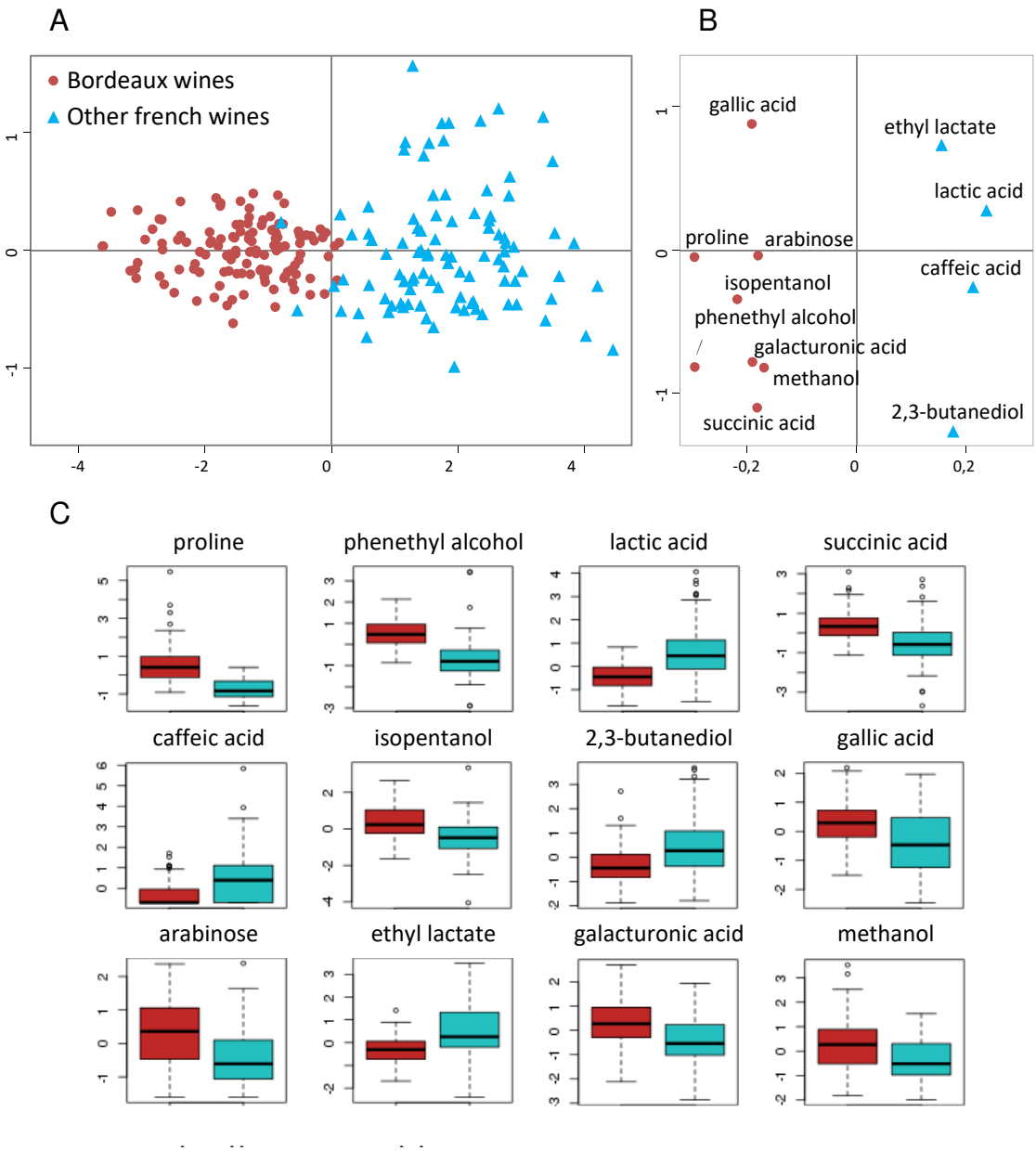
Son, H.-S., Hwang, G.-S., Kim, K. M., Ahn, H.-J., Park, W.-M., Van Den Berg, F., Hong, Y.-S., & Lee, C.-H. (2009). Metabolomic Studies on Geographical Grapes and Their Wines Using 1H NMR Analysis Coupled with Multivariate Statistics. *Journal of Agricultural and Food Chemistry, 57*(4), 1481-1490.

Son, H.-S., Kim, K. M., van den Berg, F., Hwang, G.-S., Park, W.-M., Lee, C.-H., & Hong, Y.-S. (2008). 1H Nuclear Magnetic Resonance-Based Metabolomic Characterization of Wines by Grape Varieties and Production Areas. *Journal of Agricultural and Food Chemistry, 56*(17), 8007-8016.

Wehrens, R. (2011). Multivariate Regression. In *Chemometrics with R*, (pp. 103-144). Berlin, Heidelberg: Springer.

Worley, B., & Powers, R. (2013). Multivariate Analysis in Metabolomics. *Current Metabolomics, 1*(1), 92-107.

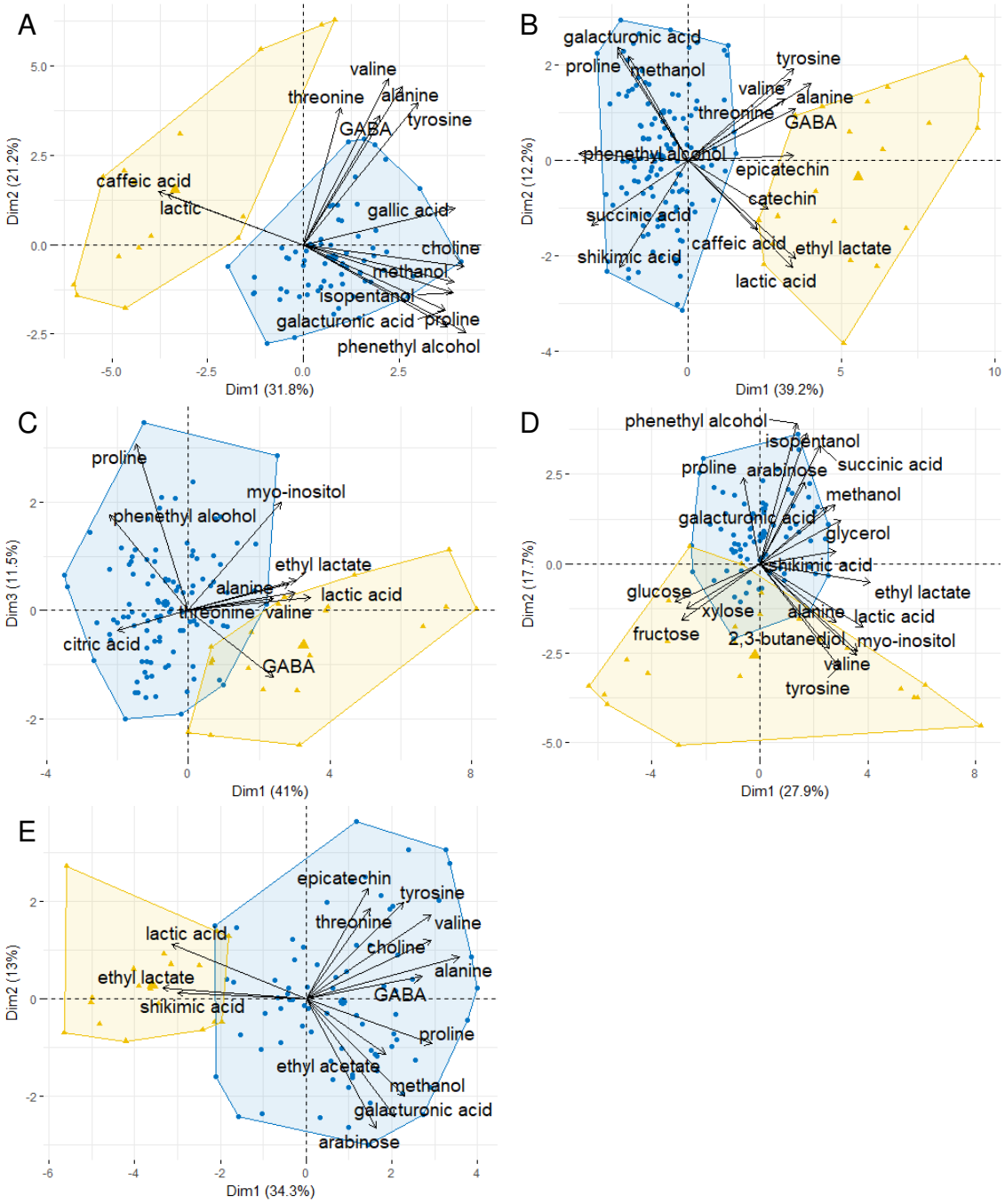
Zhu, J., Hu, B., Lu, J., & Xu, S. (2018). Analysis of Metabolites in Cabernet Sauvignon and Shiraz Dry Red Wines from Shanxi by 1H NMR Spectroscopy Combined with Pattern Recognition Analysis. In *Open Chemistry*, vol. 16 (pp. 446).



510

511

512 **Figure 1.**



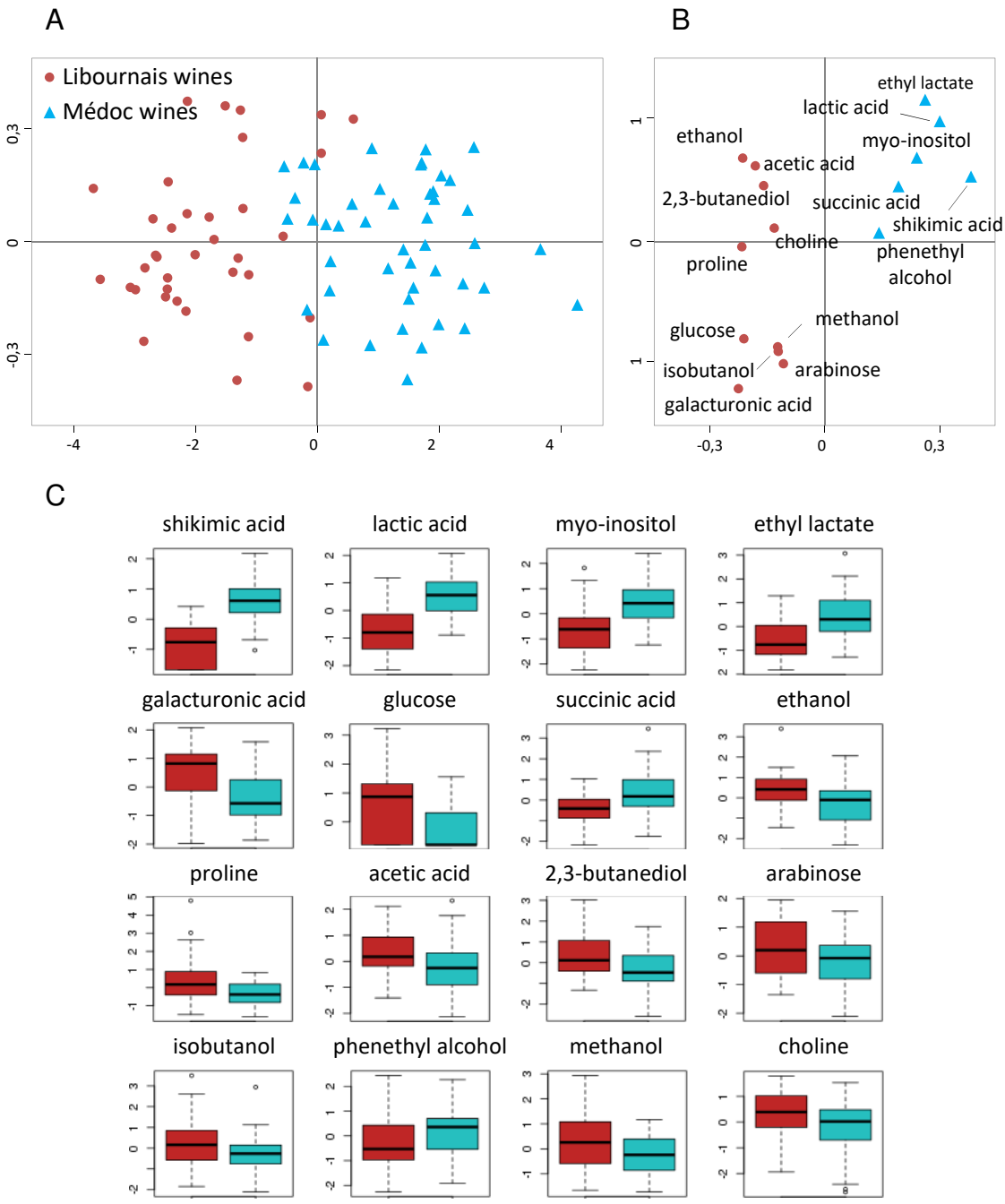
513

514

515 **Figure 2.**

516

517



518

519

520 **Figure 3.**

521

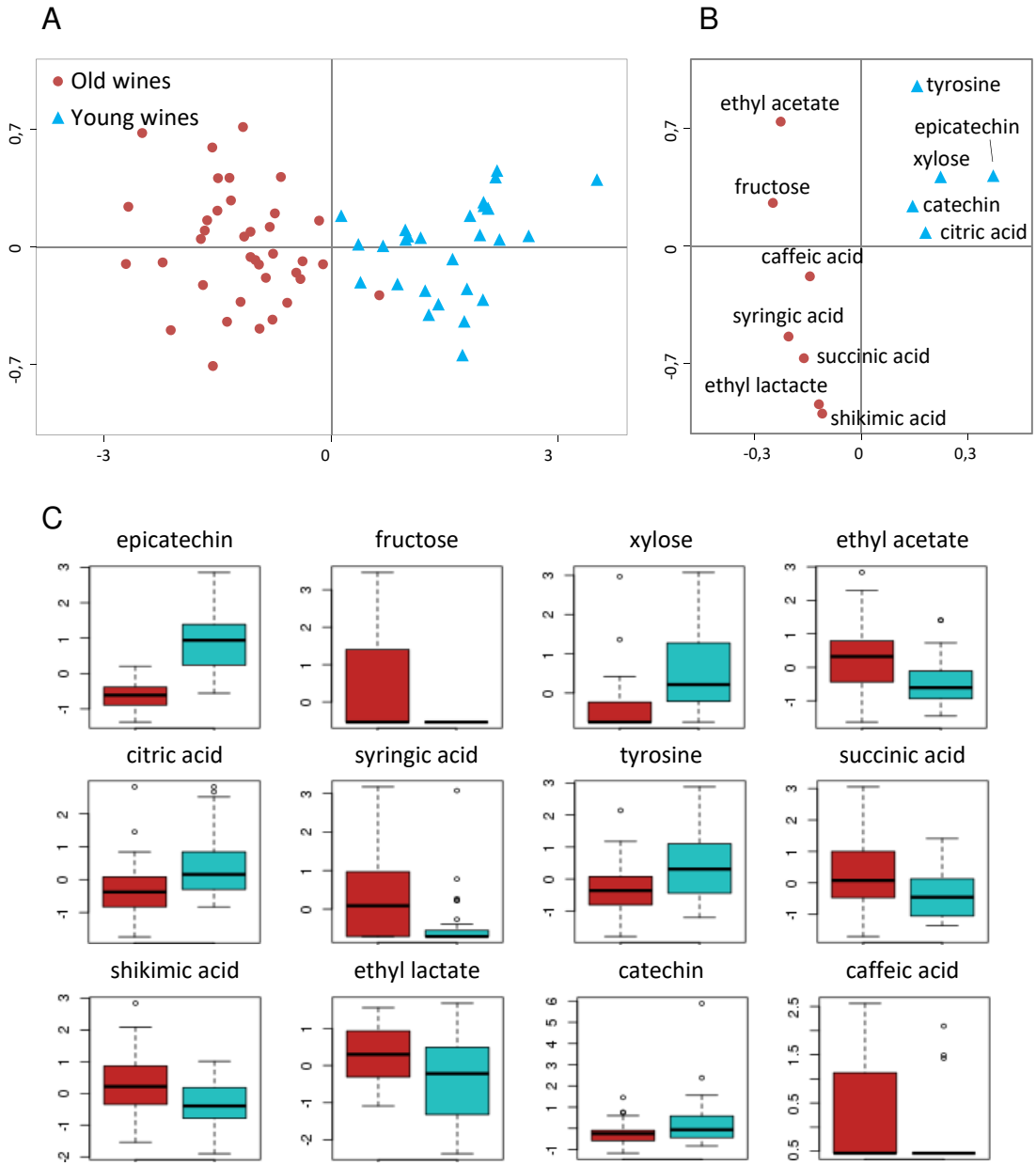

A
Old wines
Young wines
B
ethyl acetate
tyrosine
epicatechin
xylose
fructose
catechin
citric acid
caffeic acid
syringic acid
succinic acid
ethyl lactacte
shikimic acid
C
epicatechin
fructose
xylose
ethyl acetate
citric acid
syringic acid
tyrosine
succinic acid
shikimic acid
ethyl lactate
catechin
caffeic acid

Figure 4.

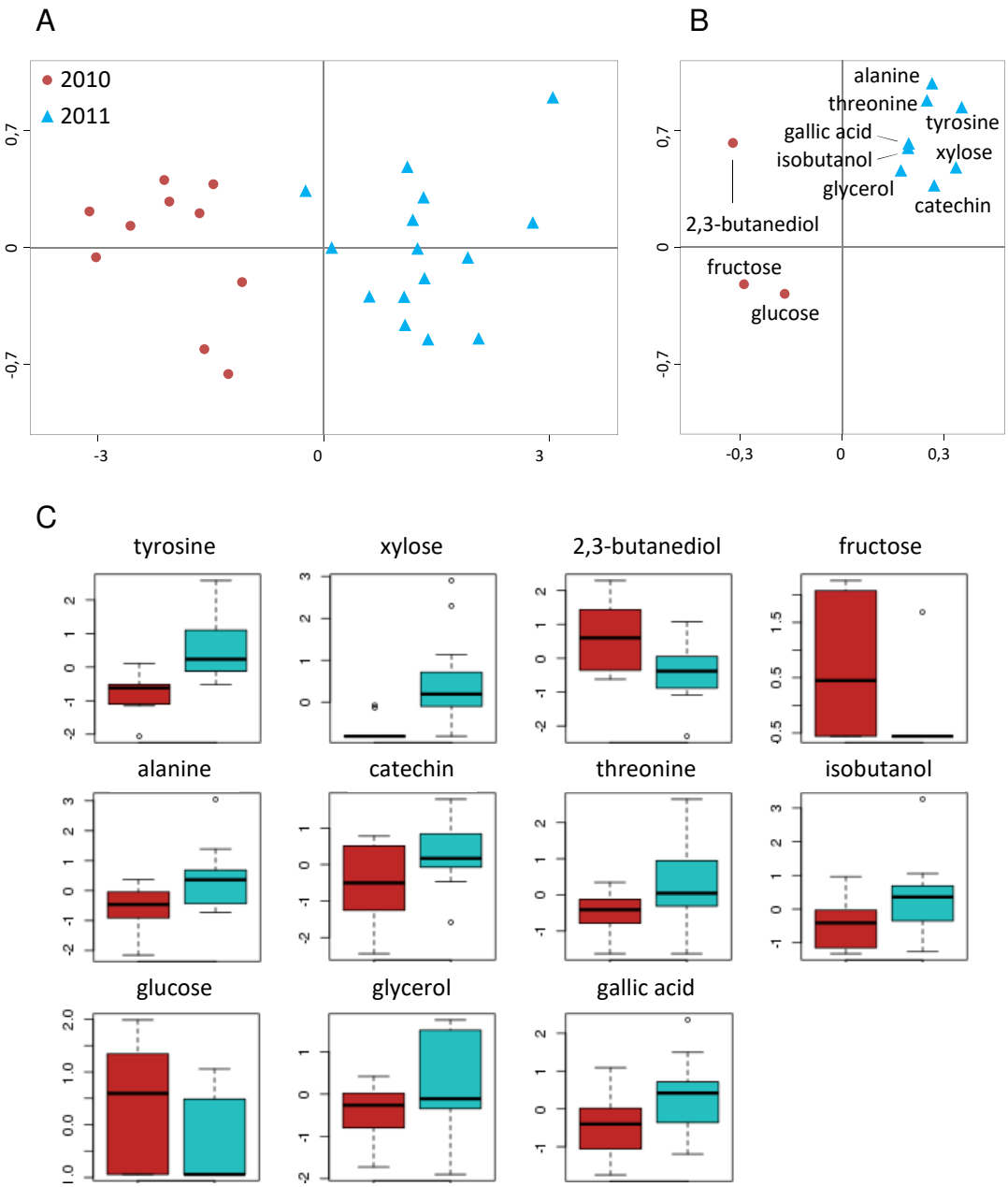

A
2010
2011
B
alanine
threonine
tyrosine
gallic acid
xylose
isobutanol
glycerol
catechin
2,3-butanediol
fructose
glucose
C
tyrosine
xylose
2,3-butanediol
fructose
alanine
catechin
threonine
isobutanol
glucose
glycerol
gallic acid

Figure 5.