

# Comparative study on the use of anthocyanin profile, color image analysis and near-infrared hyperspectral imaging as tools to discriminate between four autochthonous red grape cultivars from La Rioja (Spain)

Julio Nogales-Bueno, Francisco José Rodríguez-Pulido, Francisco José Heredia, José Miguel Hernández-Hierro\*

Food Colour & Quality Laboratory, Department of Nutrition & Food Science, Universidad de Sevilla, Facultad de Farmacia, 41012 Sevilla, Spain

---

## A B S T R A C T

Three independent methodologies were investigated to achieve the differentiation of red grapes from different grape varieties (Garnacha, Graciano, Mazuelo and Tempranillo) collected from five vineyards located in the D.O.Ca. Rioja. Anthocyanin chromatographic analysis, color image analysis and near infrared hyperspectral imaging were carried out for the grapes. Then, a Stepwise Linear Discriminant Analysis (SLDA) was developed for each data set in order to discriminate grapes according to their grape variety. As a result, using anthocyanin profile, color image analysis and near infrared hyperspectral imaging respectively, 88%, 54% and 100% of the samples were correctly classified in the internal validation process and 86%, 52% and 86% were correctly classified in the leave-one-out cross-validation process.

### Keywords:

Anthocyanin profile  
Color image analysis  
Near infrared hyperspectral imaging  
Stepwise linear discriminant analysis  
Grapes  
Chemometrics

---

## 1. Introduction

The major factors impacting on wine quality are related to winemaking process and cultivar features. Grape composition – which depends on climate, soil where grapevines are grown and grape varieties – mainly influences wine composition and therefore wine sensory parameters such as aroma, color and flavor. Consumers value the quality characteristics of wine produced from high-quality grapes enough to pay higher prices, so for each harvest, grape price are closely linked to grape quality [1].

In a cellar it is really important to know the characteristics of grapes that are taken by the vine growers. Grape variety, maturity or sugar content are typically analyzed in order to determine grape quality, set grape price and classify grapes for the various wines produced. The grape maturity and sugar content are usually determined using analytical methods recommended by the O.I.V. [2], which are simple and reliable methods. However, in cellars, grape variety is typically determined by means of visual methods based on the staff experience and knowledge. Furthermore, in situ

vine analyses, such as ampelometry are also used. However, it is not possible to have a panel to classify the berries according to their variety, essentially because this is expensive, there are many samples to be classified, and there is insufficient time to carry out the abovementioned task at harvest time. It would be appropriate to have rapid and inexpensive analytical methods to classify grapes according to their variety.

There are several studies that try to classify grape and wine samples according to their grape variety. A number of them use polyphenol content or polyphenol profiles to classify or authenticate samples [3–7]. In addition, some studies only use anthocyanin profile to carry out these classifications [8–10]. These studies generally achieve very accurate results although they are destructive and time consuming analyses that also require the use of chemical reagents. Other authors have classified wine and grape samples using spectroscopic tools in the ultraviolet–visible and near-infrared regions in order to discriminate samples according to their grape variety and origin [11–13]. Urbano et al. [11] use ultraviolet–visible spectroscopy in order to discriminate wine samples according to their origin, grape variety and ageing process. The classification rates obtained by these authors were 90%, 75% and 75%, respectively. Ferrer-Gallego et al. [12] classify grape samples using NIR spectroscopy according to their origin

obtaining a prediction rate of entire grapes correctly classified of about 93%. Using MIR/NIR traditional spectroscopy, up to seven different grape variety wines are discriminated and over 95% of the samples were correctly classified [14]. Spectroscopy achieves good results in a fast and reasonably inexpensive way. Furthermore, the use of appropriate chemometric tools is needed to classify grape and wine samples according to their grape variety [14].

The area where the present study was conducted involving the wine-producing region of D.O.Ca. Rioja, in Northern Spain, which has vineyards extending over 63,593 ha. *Vitis vinifera* L. cv Tempranillo is the most widely grown red grape variety in Spain and it is cultivated in 75% of the D.O.Ca. Rioja. This variety is capable of producing wines with a good balance of alcohol content, color and acidity, and a fruity mouthfeel that turns velvety as it ages. Graciano, which is an autochthonous cultivar to the aforesaid region, is less known, but it is used as an integral component of many Rioja wines because it is considered to contribute significantly to the quality of the wine. It offers wines with a marked acidity and polyphenolic content, ideal for ageing, with a unique aroma. Furthermore, cultivars such as Garnacha and Mazuelo are also grown in this area. Garnacha is a high-yielding grape that produces vigorous wines of great texture, body and color. Finally, although short on flavors, Mazuelo produces wines with abundant tannins, high acidity and stable color [15,16].

Taking into account the importance of these grape varieties in the D.O.Ca. Rioja and the differences between wines produced from these varieties, it would be very useful to have a fast and inexpensive method to discriminate between these four autochthonous red grape cultivars from La Rioja.

As state above, hyperspectral imaging and digital imaging may provide fast, non destructive and green chemistry determinations compared with traditional methods. Furthermore, the number of samples that might be evaluated using these methods is higher than using the traditional ones and also the determination could be automated. Moreover the use of hyperspectral imaging over conventional spectroscopy in samples having small size, such as individual berries, provides the individual spectrum of each berry using the whole grape without any contact. This may be an advantage in order to use this device at line or in line at winery belts. Nevertheless, the availability of these new devices in conventional laboratories at that time due to their high cost could be a drawback. Digital imaging is a surrogate to human inspection in order to provide an objective analysis. Moreover, the digital image device is cheaper than the hyperspectral device and should be tested in order to find the cheapest way to perform the analysis.

The main aim of this work is to compare the use of anthocyanin profile, color image analysis and near infrared hyperspectral imaging as tools to discriminate between Tempranillo, Graciano, Garnacha and Mazuelo grapes in order to select the more useful and accurate tool. In a second strand, a second level data fusion has also been tried in order to improve the results. To our knowledge, this is the first time that near infrared hyperspectral imaging has been applied to red grapes to discriminate between cultivars. Furthermore, this is also the first time that a comparative study on the use of the three aforesaid analytical tools has been performed.

## 2. Material and methods

### 2.1. Samples

*Vitis vinifera* L. cv. Tempranillo (two different vineyards), Graciano, Garnacha and Mazuelo were collected from five vineyards located in the D.O.Ca. Rioja. Red grapes were collected at two

different developmental stages during berry maturity in the 2012 vintage: harvest time and over-ripening. Harvest time and over-ripening dates were selected in agreement with the cellar staff. Individual grapes were randomly separated from the top, middle and bottom of the cluster. Five berries were taken into account for each cultivar and developmental stage with the exception of Tempranillo. In the case of Tempranillo, 20 berries were used since two different vineyards were taken into account (i.e. 5 berries  $\times$  2 vineyards  $\times$  2 developmental stages). A total of 50 samples were used. Berries were immediately frozen and stored at  $-20\text{ }^{\circ}\text{C}$  until analyses were performed.

### 2.2. Image analysis

An individual image for each grape was taken. For acquiring images, the DigiEye<sup>®</sup> system based upon a calibrated digital camera was used [17]. This system includes an illumination box specially designed by VeriVide Ltd. (Leicester, UK) to illuminate the samples consistently and a digital camera connected to a computer. The digital camera used for image acquisition was a 10.2-MP Nikon<sup>®</sup> D80 with Nikkor<sup>®</sup> 35 mm f/2D objective. In order to calibrate the digital camera, a color chart (DigiTizer, VeriVide Ltd., Leicester, UK) is used to characterize its response by relating its RGB signals to CIE specifications. The cabinet is equipped with two fluorescent tubes that emulate the CIE standard illuminant D65 and offer stable lighting conditions. They were switched on at least 10 min before being used, according to manufacturer indications, to stabilize them.

Image processing was carried out using MATLAB (The Mathworks, 2009). For obtaining CIELAB coordinates from RGB color space, the software DigiFood<sup>®</sup> [18] was used. The obtained CIELAB coordinates were used in the subsequent analyses.

### 2.3. Near-infrared hyperspectral imaging

Hyperspectral imaging device (Infaimon S.L., Barcelona, Spain) comprised a Xenics<sup>®</sup> XEVA-USB InGaAs camera (320  $\times$  256 pixel; Xenics Infrared Solutions, Inc., Leuven, Belgium), a spectrograph (Specim ImSpector N17E Enhanced; Spectral Imaging Ltd., Oulu, Finland) covering the spectral range between 900 and 1700 nm (spectral resolution of 3.25 nm). The images were recorded using the abovementioned mirror scanner pushbroom device, a 50 Hz frame rate, an exposure time of 9 ms and the instrument acquisition software SpectralDAQ v. 3.62 (Spectral Imaging Ltd., Oulu, Finland). The samples were thawed and tempered at room temperature and individual hyperspectral image of each grape was recorded. Equipment and procedure used to image recording are described in detail elsewhere in Hernández-Hierro et al. [19].

After calibration and segmentation processes, the average spectral profile for each individual grape was saved. Noisy wavebands at both extremes of the spectra range were removed and only spectral data in the resulting effective wavelength 950–1650 nm regions were used in data analysis due to reduced efficiency outside this range in the used device. Principal component analysis (PCA) was applied to reduce the dimension of the aforementioned spectral matrix. The software used for PCA analysis was Win ISI<sup>®</sup> (v1.50) (Infrasoft International, LLC, Port. Matilda, PA, USA). Overall, the spectral variability explained was 99.9% using 8 principal components. The obtained scores were used in the subsequent analyses.

### 2.4. Anthocyanins extraction and chromatographic analysis

After image acquisition process, grape skins were separated manually from the whole grapes. Individual grape skins were macerated twice at  $4\text{ }^{\circ}\text{C}$  in 10 mL of methanol containing 0.1% of

12 M HCl until colorless of skins. In order to assist the extraction, the methanolic mixtures were sonicated (JP Selecta, Barcelona, Spain) for 30 min prior to the 24 h maceration step. Methanolic phases were centrifuged (3000 rpm, 10 min), successively pooled and finally made up to 25 mL with the extraction solvent. One aliquot of 1 mL was taken and a few microliters of 0.1 M HCl was added and the extract was concentrated under vacuum at 30 °C (Eppendorf Concentrator Plus, Germany) until methanol was removed and finally made up to 1 mL with 0.1 M HCl. The aqueous extract was filtered through 0.45 µm pore-size filters and directly injected into the chromatographic system to determine the anthocyanins. Anthocyanin chromatographic analysis was carried out following a modification of García-Marino et al. [20] as described elsewhere in Hernández-Hierro et al. [19]. All analyses were performed in duplicate. Results were expressed as normalized chromatographic areas. Up to 11 anthocyanins were identified and their sum was expressed as total anthocyanins, which was used to normalize the anthocyanins areas. Moreover, taking into account their basic structures anthocyanins were grouped as acylated or non-acylated anthocyanins. Individual, acylated and non-acylated anthocyanin normalized areas were used in the subsequent analyses.

### 2.5. Supervised pattern recognition analysis

Supervised pattern recognition methods usually indicate whether samples fall into pre-defined groups, how well, and what causes this separation. The supervised pattern recognition method used in the present study was linear discriminant analysis (LDA). Stepwise feature selection was used to select the most significant variables for the discrimination between classes using *F*-statistic to test the significance of the change in Wilks' Lambda by adding or removing a variable. Prior probabilities of classification were used in this analysis taking into account each group size. The prediction ability was estimated considering the percentage of samples correctly classified by the rules developed with the training set using a leave-one-out cross-validation procedure. This approach has the advantage of being utilized when the number of total samples is low. The variables used were anthocyanin normalized areas, colorimetric parameters and the scores from the PCA performed on the hyperspectral data set. Moreover, a second level data fusion has also been tried in order to improve the results.

All discriminant analyses were carried out via SPSS 22.0 for Windows software package (SPSS, Inc., Chicago, IL, USA).

## 3. Results and discussion

### 3.1. Chromatographic data

Fig. 1a shows the average normalized area for each grape variety of the eleven chromatographic peaks. The first five peaks are nonacylated anthocyanins. This group accounts for almost 80% of total anthocyanins in red grape skins. Taking into account these pseudo chromatograms, some differences can be found between the studied varieties. Nevertheless, the use of appropriate chemometric tools is needed in order to classify grape samples according to their grape cultivar.

In order to discriminate between these four red grape cultivars, a stepwise linear discriminant analysis (SLDA) was performed as a supervised pattern recognition method using individual, acylated and non-acylated anthocyanin normalized areas to allocate the grape samples to their grape variety group. Three variables were retained that allowed 88% of the samples were correctly classified in the internal validation process and 86% were correctly classified in the leave-one-out cross-validation process. The three normalized

areas retained correspond to peaks 2, 6 and 8 (cyanidin-3-*O*-glucoside, malvidin-3-*O*-(6''-acetyl) glucoside and petunidin-3-*O*-(6''-*p*-coumaroyl) glucoside (*trans*) respectively). Fig. 1b compares normalized areas for these peaks and it is possible to observe some differences between the four varieties in this figure. Similar differences have been reported by other authors between Garnacha, Graciano and Tempranillo varieties [21,22]. However, no comparative study has been found for the minority variety Mazuelo.

Table 1 includes the results obtained for the discrimination of La Rioja grape varieties according to its chromatographic profile. The results are expected, since it is well known the relationship between grape varieties and their chromatographic profiles [23].

### 3.2. Color data

Fig. 2 shows the CIELAB coordinates for all grape samples ( $L^*$ ,  $a^*$  and  $b^*$ ), polar coordinates ( $C^*_{ab}$  and  $h_{ab}$ ) are also readily available from this figure. Majority of Garnacha, Mazuelo and Tempranillo samples present positive values of  $a^*$  coordinate, and Graciano samples present the fewest values of chroma. Moreover, all samples present very low values for  $a^*$  and  $b^*$  coordinates and the four varieties are overlapped in the CIELAB diagram.

In order to achieve the classification goal using discriminate variables more easily obtained, a new SLDA model was developed using the aforesaid color data.

After develop the stepwise discriminant analysis only  $C^*_{ab}$  and  $h_{ab}$  were retained and this allowed 54% of the samples were correctly classified in the internal validation process and 52% were correctly classified in the leave-one-out cross-validation process. Consequently, image analysis, although is a rapid and inexpensive tool, has lower discriminant power in red grape samples than chromatographic analysis. This can be due to the sample nature with low values of  $a^*$  and  $b^*$  coordinates which lead to low signal to noise ratio.

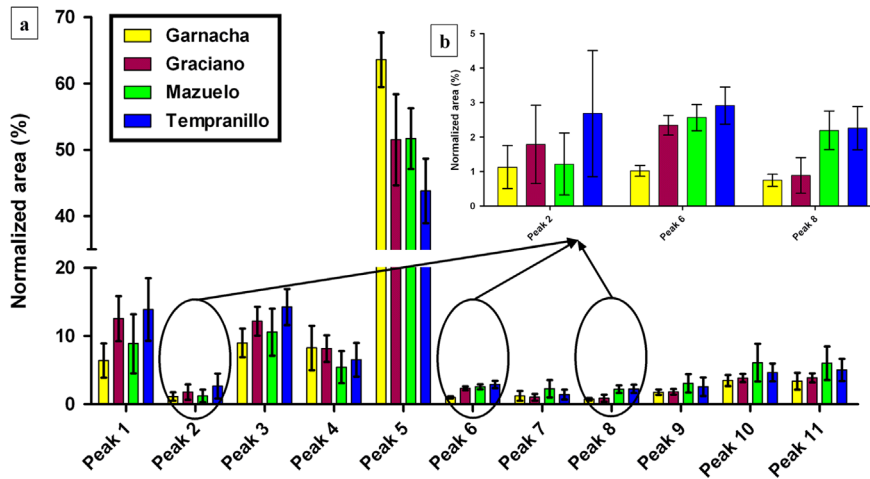
### 3.3. Near-infrared hyperspectral data

Fig. 3 shows the scores of the first and second principal components, the four varieties are completely overlapped in this space. The same discriminant procedure (i.e. a new SLDA model) was applied to the dataset using the scores of the 8 first principal components obtained from near infrared hyperspectral data. In this case all variables were retained and 100% of the samples were correctly classified in the internal validation process and 86% were correctly classified in the leave-one-out cross-validation process. These results are better in the internal validation and similar in the leave-one-out cross-validation than the results obtained from the chromatographic data. In addition, hyperspectral imaging is a non-destructive and faster than the chromatographic technique, which highlights the importance of these results. Fig. 4 represents the grape samples in the space defined by the three obtained discriminant functions. In this space, grape variety groups are almost completely separated.

### 3.4. Data fusion

In order to improve the results obtained using the hyperspectral data set a second level data fusion was carried out. The aim of second level data fusion includes performing high-level inferences and identifying significant activities and events (patterns in general) [24]. Color and hyperspectral data were fused; they are non-destructive and fast tools. Hyperspectral data have shown potential classifying these grape varieties and color data could be linked with visible spectra and could add some discriminant information.

A SLDA was performed with this new data set. Only near infrared hyperspectral image variables were retained and, as a



**Fig. 1.** (a) Average normalized areas for the 11 anthocyanins detected for Garnacha, Graciano, Mazuelo and Tempranillo grape varieties. In order of elution: delphinidin-3-*O*-glucoside, cyanidin-3-*O*-glucoside, petunidin-3-*O*-glucoside, peonidin-3-*O*-glucoside, malvidin-3-*O*-glucoside, malvidin-3-*O*-(6'-acetyl) glucoside, cyanidin-3-*O*-(6'-*p*-coumaroyl) glucoside (*trans*), petunidin-3-*O*-(6'-*p*-coumaroyl) glucoside (*trans*), malvidin-3-*O*-(6'-*p*-coumaroyl) glucoside (*cis*), peonidin-3-*O*-(6'-*p*-coumaroyl) glucoside (*trans*) and malvidin-3-*O*-(6'-*p*-coumaroyl) glucoside (*trans*). (b) Average normalized areas for the three anthocyanins selected by the stepwise linear discriminant analysis: cyanidin-3-*O*-glucoside, malvidin-3-*O*-(6'-acetyl) glucoside and petunidin-3-*O*-(6'-*p*-coumaroyl) glucoside (*trans*).

**Table 1**

Results of the different stepwise linear discriminant analyses developed to classify Garnacha, Graciano, Mazuelo and Tempranillo grapes.

Data set	Retained variables	Internal validation		Leave-one-out cross-validation	
		Samples correctly classified	% of Samples correctly classified	Samples correctly classified	% of Samples correctly classified
Chromatographic data <sup>a</sup>	Cyg <sup>b</sup> , Mvac <sup>c</sup> , Ptcp <sup>d</sup>	44/50	88	43/50	86
Color data <sup>e</sup>	$C_{ab}^*$ , $h_{ab}^*$	27/50	54	26/50	52
NIR Hyperspectral data <sup>f</sup>	NIR Hyperspectral data <sup>f</sup>	50/50	100	43/50	86
Data fusion (NIR+ color)	NIR Hyperspectral data <sup>f</sup>	50/50	100	43/50	86

<sup>a</sup> Chromatographic data: individual, acylated and non-acylated anthocyanin normalized areas.

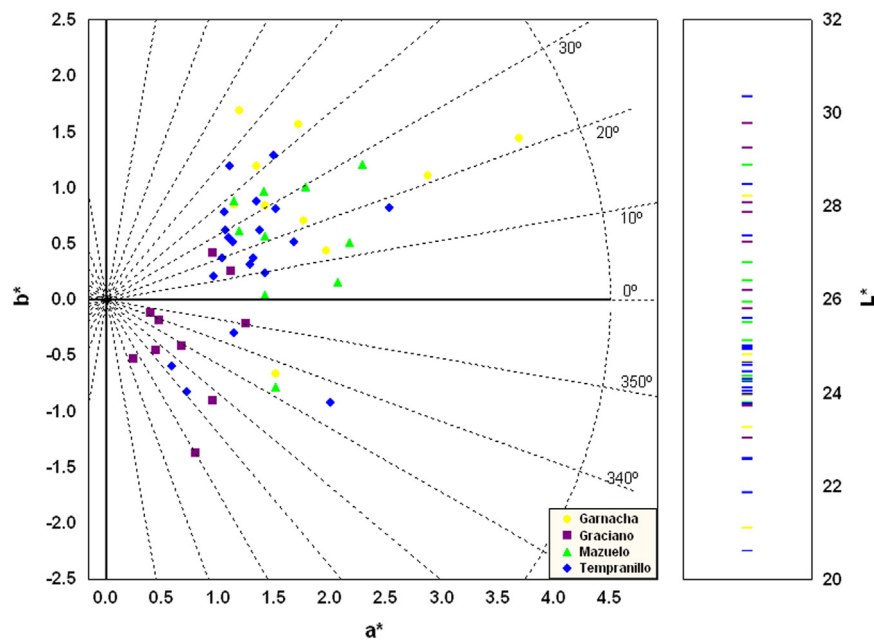
<sup>b</sup> Cyg: average normalized area of cyanidin-3-*O*-glucoside.

<sup>c</sup> Mvac: average normalized area of malvidin-3-*O*-(6'-acetyl) glucoside.

<sup>d</sup> Ptcp: average normalized area of petunidin-3-*O*-(6'-*p*-coumaroyl) glucoside (*trans*).

<sup>e</sup> Color data: CIELAB coordinates ( $L^*$ ,  $a^*$  and  $b^*$ ) and polar coordinates ( $C_{ab}^*$  and  $h_{ab}^*$ ) for all grape samples.

<sup>f</sup> NIR Hyperspectral data: scores of the eight first principal components obtained from the near infrared hyperspectral data.



**Fig. 2.** CIELAB coordinates for Garnacha, Graciano, Mazuelo and Tempranillo grape samples.

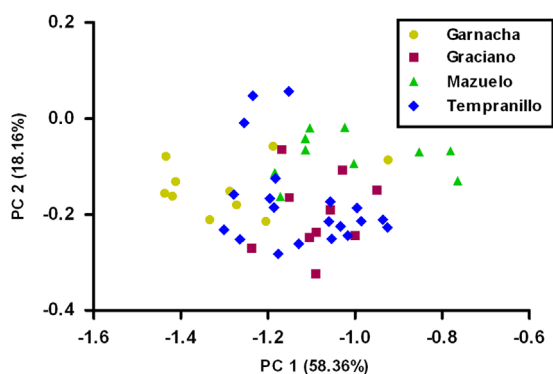


Fig. 3. Score plot of grape samples in the space defined by the first and second principal component codified according to grape variety.

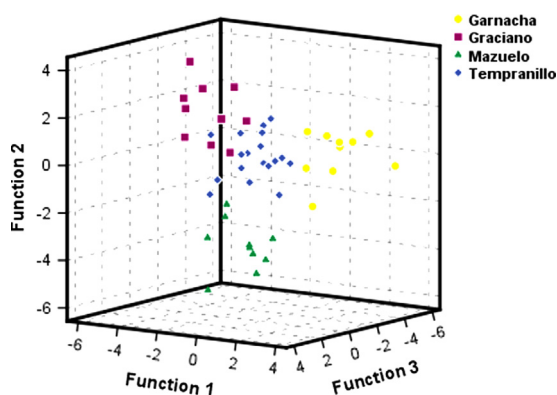


Fig. 4. Representation of the grape samples in the space defined by the first, second and third discriminant functions obtained from the hyperspectral data.

result, the same discriminant functions were obtained with the same discriminant power. Taking that into account, color data do not improve the results obtained from near infrared hyperspectral image. Table 1 resumes data fusion results and the results obtained in the previous stepwise linear discriminant analyses.

#### 4. Conclusions

Chromatographic analysis is a widespread and reliable tool to classify red grapes according to grape variety and the results obtained in this study confirm it. Color image analysis is a rapid and inexpensive tool that has lower discriminant power for red grape samples than chromatographic analysis. Near-infrared hyperspectral analysis is a non-destructive and fast technique that shows better or similar results than the chromatographic technique. A second level data fusion between color data and near infrared hyperspectral data does not improve the results obtained from near infrared hyperspectral image. Nonetheless, a comprehensive study

should be made in order to evaluate several abiotic factors, including vintage, watering, sun exposure, wine age, soil, ripening etc in the complete development of these models.

#### Acknowledgements

The Spanish MINECO is thanked for J. Nogales-Bueno FPI grant (BES-2012-060192), J.M. Hernández-Hierro Juan de la Cierva contract (JCI-2011-09201) and project AGL2011-30254-C02. Junta de Andalucía is also thanked for financial support (project P10-AGR6331). The authors also thank Bodegas RODA (La Rioja, Spain) for supplying the grape samples and to technical staff of Biology Service (SGI, Universidad de Sevilla) for the technical assistance.

#### References

- [1] A. Golan, H. Shalit, *J. Agric. Econ.* 44 (1993) 311–321.
- [2] OIV. *Recueil de methods internationaux d'Analyse des vins. Caracteristiques chromatiques*, 1990. Paris, OIV.
- [3] A. Hosu, V.M. Cristea, C. Cimpoi, *Food Chem.* 150 (2014) 113–118.
- [4] A. de Villiers, P. Majek, F. Lynen, A. Crouch, H. Lauer, P. Sandra, *Eur. Food Res. Technol.* 221 (2005) 520–528.
- [5] F. Galgano, M. Caruso, G. Perretti, F. Favati, *Eur. Food Res. Technol.* 232 (2011) 889–897.
- [6] I. Hermosin-Gutierrez, N. Castillo-Munoz, S. Gomez-Alonso, E. Garcia-Romero, *Abstr. Pap. Am. Chem. Soc.* 239 (2010).
- [7] L. Jaitz, K. Siegl, R. Eder, G. Rak, L. Abranko, G. Koellensperger, S. Hann, *Food Chem.* 122 (2010) 366–372.
- [8] D. von Baer, M. Rentzsch, M.A. Hitschfeld, C. Mardones, C. Vergara, P. Winterhalter, *Anal. Chim. Acta* 621 (2008) 52–56.
- [9] A. de Villiers, G. Vanhoenacker, P. Majek, P. Sandra, *J. Chromatogr. A* 1054 (2004) 195–204.
- [10] M.L. Gonzalez San Jose, C. Diez, *Agrochimica* 37 (1993) 86–92.
- [11] M. Urbano, M.D. Luque de Castro, P.M. Perez, J. Garcia-Olmo, M.A. Gomez-Nieto, *Food Chem.* 97 (2006) 166–175.
- [12] R. Ferrer-Gallego, J.M. Hernández-Hierro, J.C. Rivas-Gonzalo, M.T. Escribano-Bailón, *J. Sci. Food Agric.* 93 (2013) 967–972.
- [13] C.J. Bevin, R.G. Damberg, A.J. Fergusson, D. Cozzolino, *Anal. Chim. Acta* 621 (2008) 19–23.
- [14] A. Versari, V. Laurie, A. Ricci, L. Laghi, G.P. Parpinello, *Food Res. Int.* 60 (2014) 2–18.
- [15] Denominación de Origen Calificada Rioja. Denominación de Origen Calificada Rioja, 2014. URL: (<http://es.riojawine.com/en/5-learn-about-rioja.html>), Accessed 23.04.13.
- [16] ICEX España Exportación e Inversiones. Vinos de España. 2014. URL: (<http://www.winesfromspain.com/>) Accessed 23.04.13.
- [17] M.R. Luo, G.H. Cui, L. C. British Patent entitled apparatus and method for measuring colour (DigiEye System), Derby University Enterprises Limited. [0124683.4], 4-10-2001.
- [18] F.J. Heredia, M.L. González-Miret, C. Álvarez, A. Ramírez. DigiFood, [Registration No. Se-01298], 2006.
- [19] J.M. Hernández-Hierro, J. Nogales-Bueno, F.J. Rodríguez-Pulido, F.J. Heredia, *J. Agric. Food Chem.* 61 (2013) 9804–9809.
- [20] M. García-Marino, J.M. Hernández-Hierro, J.C. Rivas-Gonzalo, M.T. Escribano-Bailón, *Anal. Chim. Acta* 660 (2010).
- [21] E. Revilla, E. García-Beneytez, F. Cabello, G. Martín-Ortega, J.M. Ryan, *J. Chromatogr. A* 915 (2001) 53–60.
- [22] I. Arozarena, B. Ayestarán, M. Cantalejo, M. Navararro, M. Vera, I. Abril, A. Casp, *Eur. Food Res. Technol.* 214 (2002) 303–309.
- [23] A.L. Waterhouse, *Wine Phenolics*, New York Acad Sciences, New York, 2002.
- [24] F. Castanedo, *Sci. World J.* 10 (2013) 1–19.