

Article

Liquid Chromatographic Approach for the Discrimination and Classification of Cava Samples Based on the Phenolic Composition Using Chemometric Methods

Anaïs Izquierdo-Llopart¹ and Javier Saurina^{1,2,*} 

¹ Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, Martí i Franquès, 1-11, E-08028 Barcelona, Spain; anaïs.izquierdo.llopart@gmail.com

² Research Institute in Food Nutrition and Food Safety, University of Barcelona, Recinte Torribera, Av. Prat de la Riba 171, Edifici de Recerca (Gaudí), Santa Coloma de Gramenet, E-08921 Barcelona, Spain

* Correspondence: xavi.saurina@ub.edu

Received: 26 June 2020; Accepted: 28 August 2020; Published: 1 September 2020



Abstract: Phenolic profiles obtained by liquid chromatography with UV/vis detection were here exploited to classify cava samples from the protected designation of origin Cava. Wine samples belonging to various classes which differed in grape varieties, blends and fermentation processes were studied based on profiling and fingerprinting approaches. Hence, concentrations of relevant phenolic acids and chromatograms registered at 310 nm were preliminarily examined by Principal Component Analysis (PCA) to extract information on cava classes. It was found that various hydroxybenzoic and hydroxycinnamic acids such as gallic, gentisic, caffeic or caftaric acids were up- or down-expressed depending on the wine varieties. Additionally, Partial Least Squares Discriminant Analysis (PLS-DA) was applied to classify the cava samples according to varietal origins and blends. The classification models were established using well-known wines as the calibration standards. Subsequently, models were applied to assign unknown samples to their corresponding classes. Excellent classification rates were obtained thus proving the potentiality of the proposed approach for characterization and authentication purposes.

Keywords: cava (sparkling wine); liquid chromatography; phenolic acids; principal component analysis; classification methods; partial least square-discriminant analysis

1. Introduction

Cava is a highly popular sparkling wine with Protected Designation of Origin (PDO) produced in various Spanish regions following the traditional Champenoise method. White cavas are just elaborated from the pressed must of white grapes, rosé cavas are subjected to a must maceration with its own skins and peas for a given period of time, and “blanc de noirs” cavas are produced with musts coming from black skin grapes but with no maceration process. After that, enzymes and SO₂ are added to clarify and avoid some unwanted processes such as the oxidation of polyphenols. The alcoholic fermentation is carried out after the must clarification inside stainless steel tanks in which selected yeasts are added to transform the sugars into ethanol and CO₂. Once the alcoholic fermentation is completed, base wines can be subjected to the malolactic fermentation (MLF) to convert malic acid into lactic acid. MLF provides a notable change in acidity features, leading to a more pleasant and bulky mouthfeel [1]. At the end of this step, base wines are ready to be mixed to create the desired coupages or blends. Wines are then stabilized, bottled with a small volume of the *tirage* liquor (mixture of sugars, yeasts and wine) to perform the second fermentation inside the own bottle. Finally, cavas are aged in

the cellar for a minimum of 9 months. A total of 200 million liters of cava are yearly commercialized, representing more than EUR 1100 M and being the most exported Spanish wine. Owing to the great social and economic impact around the cava trade, mechanisms of control to ensure the standards of quality and to prevent potential fraudulent practices are needed, especially those based on rapid, simple and inexpensive analytical methods suitable for large series of samples [2–4].

The classification, authentication and traceability of wines and related products is a hot topic of agricultural and food science [5,6]. In many cases, chemometric methods for data analysis are required to help to extract the underlying information and provide feasible descriptive and classificatory models. In this regard, principal component analysis (PCA) is typically used for exploratory studies of wines to deduce trends related to oenological features. Further classification surveys can be based on linear discriminant analysis (LDA), partial least square-discriminant analysis (PLA-DA), cluster analysis (CA), soft independent modeling of class analogy (SIMCA) analysis, etc. For additional details, reference [7] reviewed the possibilities of multivariate methods for the assessment of geographical origins of wines.

The number of publications on wine characterization has increased dramatically in the last years and some basic considerations on the analytical possibilities and working strategies have been established [8]. Nowadays, near-infrared spectroscopy (NIR) is one of the most successful approaches to carry out authentication studies from a fingerprinting approach, taking advantage of features such as the simplicity and reduced instrumental cost [9]. An interesting revision facing the possibilities of NIR was published by Wang et al. [10], including representative examples of alcoholic beverages and other food products. In parallel, elemental and isotopic analysis has demonstrated its great performance, especially for dealing with geographical issues, as the elemental composition of grapes and wines is highly dependent on terroir factors [11]. Other alternative instrumental techniques often applied to food authentication include NMR and MS, providing excellent discriminatory performance despite the higher degree of instrumental sophistication and cost [12,13]. UV-vis and fluorescence spectroscopies have also been used in the study of food features [14]. The potentiality of joining data from different instrumental sources has been critically evaluated by Borrás et al. [15] showing the advantages in the sample description from the richer data fusion systems.

Some recent wine characterization and authentication applications rely on a more specific phenolic profiling approach using separation techniques, mainly liquid chromatography (HPLC) with UV detection, for a feasible and accurate quantification of targeted compounds. For instance, red wines of Rias Baixas and Ribeira Sacra have been analyzed by HPLC-UV and the resulting compositional profiles have been used for their classification [16]. Other authors have applied phenolic profiling by HPLC-UV to discriminate among Malbec wines produced in Mendoza (Argentina) and California (USA) regions [17]. The authentication of Czech Republic wines from different terroirs has also accomplished by HPLC profiling based on hydroxycinnamic and flavonoid components [18]. Another comprehensive survey for the characterization and classification of Spanish wines or cavas from Penedès, Rioja and Ribera designations of origin has also been based on HPLC with UV-vis multi-way and fluorescence detection to provide more sensitive and selective data [19]. In order to go into detail in the characterization of potential markers of wines classes defined by origins, varieties, blends, vintages, etc., HPLC with mass spectrometry (MS) detection is able to generate a rich set of discriminant features that could further submitted to structural elucidation [20,21]. Apart from HPLC, other separation techniques have been used such as thin-layer chromatography [22], GC-MS [23] and capillary zone electrophoresis [24,25].

Alternative approaches to separation methods have also been proposed in the literature such as electroanalytical assays carbon nanotube-modified electrodes for the determination of the phenolic acid and flavonoid content of white wines [26]. Electronic tongue devices based on enzymatic biosensors have also been applied to the quantification of this type of analytes towards a fast and simple characterization of wines [27]. In some cases, phenolic data have been complemented with information from other families of compounds to increase the descriptive performance. For instance, phenolic indexes, phenolic acid and flavonoid contents have been combined with concentrations of

organic acids, sugars and biogenic amine to try to recognize autochthonous Calabrian red wines from other similar [28].

This work is focused on the study the possibilities of phenolic acid profiles and fingerprints obtained by HPLC with UV detection to discriminate and classify cava made from different grape varieties and blends. The chromatographic method has been established previously and applied to the determination of the most important hydroxybenzoic and hydroxycinnamic acid, and flavonoids components in white wines and cava samples [29,30]. Cava samples belonging to various classes have been analyzed and the resulting data have been treated with chemometric methods to extract the relevant information. The study of the sample distribution by PCA has identified potential discriminating factors among cava. Further studies by PLS-DA have accomplished the classification of samples according to their phenolic composition and new (unknown) samples have been successfully assigned to their corresponding classes.

2. Materials and Methods

2.1. Chemicals and Solutions

Gallic, homogentisic, protocatechuic, caftaric, gentisic, vanillic, caffeic, syringic, ferulic, and *p*-coumaric acids, and (+)-catechin, (–)-epicatechin, ethyl gallate, resveratrol, rutin, myricetin, quercetin, kaempferol and apigenin were purchased from Sigma-Aldrich (St. Louis, MO). Pure stock standard solutions of each phenolic acid were prepared at 5000 mg L⁻¹ in methanol (UHPLC-Supergradient, Panreac ApplyChem, Castellar del Vallès) and stock standard solutions of resveratrol and flavonoids were prepared at 1000 mg L⁻¹ in dimethyl sulfoxide (for analysis, 99.9%, Panreac ApplyChem). Formic acid (>96%, Merck KGaA, Darmstadt, Germany), methanol and water (Elix3, Millipore, Bedford, MA) were the components of the mobile phase used in the chromatographic method. Working standard solutions consisting of a mixture of analytes at concentrations ranging from 0.05 to 20 mg L⁻¹ were prepared in methanol:water (1:1, *v:v*) as the solvent. These standard solutions were stored at 4 °C in the refrigerator for no more than 1 week.

2.2. Samples

Cava samples were kindly provided by Raventós Codorníu Group. The set of analyzed samples come from the harvest of 2016–2017 and consisted of white and rosé cava produced in Penedès and Costers del Segre regions (both from Catalonia, Spain). Different white grape varieties—Chardonnay (Ch), Macabeu (Ma), Xarel·lo (Xa), Parellada (Pa)—, and the black ones—Blanc de Noirs (BN, from PN), Pinot Noir (PN), Monastrell (Mo), Garnacha negra (Ga) and Trepal (Tr)—have been studied. Base wines resulting from the first fermentation were conveniently blended to produce the following cava combinations according to their aging period as shown in Table 1. The total of 75 samples under study consisted of 5 classes of monovarietal cava (10 samples in each class) and 5 types of blended cava (5 samples in each class). It should be pointed out that despite class C is not monovarietal, the combination of Macabeu, Xarel·lo and Parellada has been considered as a main class since it is the traditional coupage of Catalan Cava. Samples were degasified and filtered through a nylon membrane (0.45 µm pore size) prior HPLC-UV analysis.

A quality control (QC) solution was prepared mixing 50 µL of each cava sample. The QC was used to evaluate the reproducibility of the chromatographic method and the significance of the PCA models. Cava samples were analyzed randomly, and the QC was repeatedly injected every 10 samples.

Table 1. Composition of the studied cava blends.

Class/Blend	Composition	Number of Samples
C ¹	100% Macabeu, Xarel·lo, Parellada	10
G ²	100% Chardonnay	10
I ³	100% Blanc de Noirs	10
P	100% Pinot Noir	10
W	100% Blanc de Noirs	10
A	70% Macabeu, Xarel·lo, Parellada/30% Chardonnay	5
E	85% Macabeu, Xarel·lo, Parellada/15% Chardonnay	5
S	50% Macabeu, Xarel·lo/50% Chardonnay	5
T	70% Pinot Noir/30% Chardonnay	5
V	Monastrell/Garnatxa negra/Trepat	5

⁽¹⁾ The combination of Macabeu, Xarel·lo and Parellada wines is considered the traditional coupage of Catalan Cavas; this is why this blend is here is treated as a single class. ⁽²⁾ Blend G: 15–30 months of aging period; 9–15 months of aging period for the other blends. ⁽³⁾ not subjected to malolactic fermentation.

2.3. Chromatographic Method

The chromatographic system was composed of an Agilent Series 1100 HPLC Chromatograph (Agilent Technologies, Palo Alto, CA, USA) equipped with a quaternary pump (G1311A), a degasser (G1322A), an automatic injection system (G1392A) and a diode array detector (G1315B). An Agilent ChemStation for LC 3D (Rev. A. 10.02) software was used for instrument control and data processing. The chromatographic method was previously optimized and validated [29,30]. It was based on reversed-phase mode using core-shell column (Kinetex, 100 mm × 4.6 mm I.D., 2.6 µm particle size from Phenomenex, Torrance, CA, USA) and 0.1% formic acid aqueous solution and methanol as the components of the mobile phase. The elution gradient was as follows: 0 to 20 min, 15–60% methanol (lineal increase); 20 to 22 min, 60–90% methanol (lineal increase); 22 to 27 min, 90% methanol (cleaning step); 27 to 27.5 min, 90–15% methanol (lineal decrease); 27.5 to 30, 15% methanol (conditioning step). The flow rate was 1 mL min⁻¹ and the injection volume was 10 µL. Chromatograms were acquired at 280, 310 and 370 nm.

2.4. Data Analysis

Exploratory and classification studies were carried out using PLS-Toolbox (Eigenvector Research, Inc. Manson, WA, USA) working with MATLAB (MatWorks Natick, MA, USA). X-data matrices corresponding to either concentrations (profiling approach) or chromatographic fingerprints (fingerprinting approach) were preliminarily evaluated by principal component analysis (PCA) to try to find patterns for sample characterization. Further discriminatory studies to assign unknown cava samples to their corresponding classes relied on partial least square-discriminant analysis (PLS-DA). Apart from the profiling or fingerprinting X-matrices, the belonging of each sample to a predefined class was conveniently coded in the Y-matrix. The number of latent variables to build the classification models was established by cross validation under Venetian bind approach. Once the models were validated, they were used to predict the membership of the new unknown cava samples.

3. Results and Discussion

Cava samples were analyzed according to a HPLC-UV/vis method previously established for the determination of the most abundant phenolic acids and flavonoids in cava and white wines [29,30]. The optimized chromatographic method was able to separate successfully the target compounds in standard mixtures. In the case of cava samples, however, additional peaks of unknown species were found which may overlap with the analytes. Anyway, the occurrence of coeluting peaks was irrelevant in fingerprinting while, in profiling, data from partially resolved peaks may be more imprecise.

Although chromatograms were recorded at 3 different wavelengths (280, 310 and 370 nm), the richest fingerprints in terms of intensity of absorbing compounds were obtained at 280 nm

which mainly corresponded to benzoic acids. Alternatively, detection at 310 nm provided clearer chromatograms with an easier identification of peaks of hydroxycinnamic acids and their esters with tartaric acid. In contrast, detection at 370 nm displayed poor information since flavonoid concentrations were low in this kind of samples. Detection at 310 nm was finally preferred as a compromise among sensitivity and selectivity. Figure 1 shows the structures of the most relevant phenolic acids considered in this paper as tentative wine descriptors.

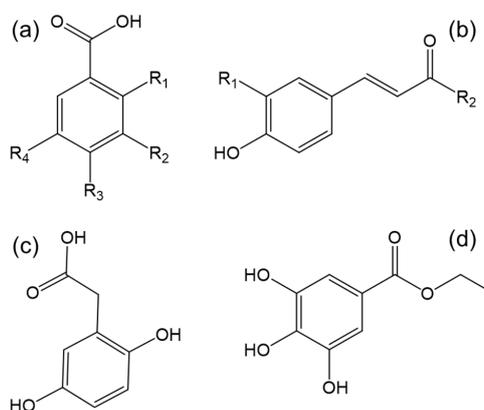


Figure 1. Structures of the principal phenolic acids found in cava samples. (a) General structure of benzoic acids: Gallic acid ($R_1 = H$ and $R_2, R_3, R_4 = OH$); Gentisic acid ($R_2, R_3 = H$ and $R_1, R_4 = OH$); Protocatechuic acid ($R_1, R_4 = H$ and $R_2, R_3 = OH$); Vanillic acid ($R_1, R_4 = H, R_3 = OH$ and $R_2 = OCH_3$); Syringic acid ($R_1 = H, R_3 = OH$ and $R_2, R_4 = OCH_3$). (b) General structure of hydroxycinnamic acids: Caffeic acid ($R_1 = OH$ and $R_2 = H$); Ferulic acid ($R_1 = OCH_3$ and $R_2 = H$); *p*-Coumaric acid ($R_1, R_2 = H$); Caftaric acid ($R_1 = OH$ and $R_2 = \text{tartaric acid-HO}_2\text{CCH(OH)CH(OH)CO}_2\text{H}$). (c) homogentisic acid. (d) ethyl gallate.

When comparing some featured monovarietal and blend cava, the visual inspection of chromatograms depicted in Figure 2 showed differences regarding phenolic acid composition and levels. Some relevant peaks found in the chromatograms were identified from the comparison of spectral and chromatographic features with those of pure standards. In this way, compounds such as gallic, caftaric, gentisic, vanillic, caffeic or *p*-coumaric acid were detected. Anyway, some unknown signals are still pending of assignment. The quantification of compounds was carried out using 10 different standard mixtures with concentrations ranging from 0.05 to 20 mg L⁻¹ for each analyte to establish the corresponding linear regression models.

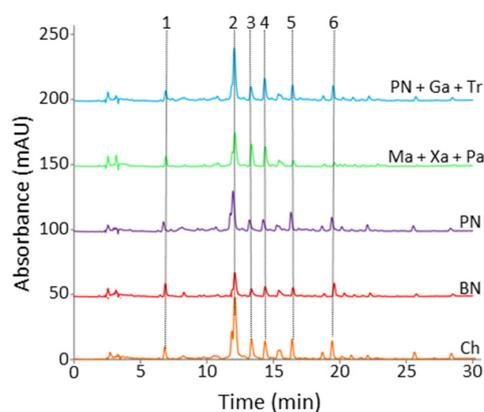


Figure 2. Chromatograms of representative white and Rosé cava wine recorded at 310 nm. Sample assignment: Ch, Chardonnay; BN, Blanc de Noirs; PN, Pinot Noir; Ma + Xa + Pa, Macabeu + Xarel-lo + Parellada; PN + Ga + Tr, Pinot Noir + Garnacha + Trepat. Peak identification: (1) gallic acid, (2) caftaric acid, (3) gentisic acid, (4) vanillic acid, (5) caffeic acid, (6) *p*-coumaric acid.

Quantitative results indicated that monovarietal Ch samples displayed, in general, the highest concentrations for most of the analytes. Rosé cavas such as the monovarietal PN and the blend of Mo, Ga, and Tr were also rich in phenolic compounds. On the other hand, the classical blend of Ma, Xa and Pa showed lower concentrations, except for gentisic and caftaric acids. Finally, the BN class was the one exhibiting the lowest levels of phenolic compounds. As can be seen below, these general differences in composition were exploited in further discrimination and classification studies as a function of grape varieties, blends and winemaking processes.

3.1. Exploratory Studies by PCA

3.1.1. Study of Chromatographic Fingerprints

Chromatograms acquired by UV spectroscopy at several working wavelengths were used as the data to be treated by PCA for a preliminary characterization of cavas according to class varieties. As detailed in the experimental section, fingerprint datasets consisted of absorbance values taken over a selected time window from the set of samples under study. QCs were also included to assess the soundness of the model. The performance of the PCA models built with chromatograms at 280 and 310 nm was first compared. Although the results were quite similar in terms of sample description ability, conclusions extracted from the data set at 310 nm seemed to be more solid as the experimental variability and the compactness of the group of QCs were better. For this reason, we have restricted the description of the PCA results to the best model at 310 nm.

Data were taken within the time window 6.30–27.30 min containing most of the peak features while the initial chromatographic and column cleaning ranges were excluded. The resulting X–matrix was composed of 84 samples (75 cava wines and 9 QC replicates) and 2700 time variables. The PCA model captured more than 75% of data variance with 3 principal components (PCs) (52.6, 17.0 and 6.6% for PC1, PC2 and PC3, respectively). It was deduced that PC1 discriminated among rosé (right side) and white (left side) cavas based on differences in the distribution of the major phenolic signals (see Figure 2). PC2 was mainly focused on the modeling the chromatographic variability associated to baseline drifts and peak shifting so that, in this case, it was less relevant from the point of view of the wine description. On the contrary, PC3 dealt with the overall phenolic content, with samples rich in phenolic compounds at the top and those with lower concentrations at the bottom.

QC samples clustered in a compact group in the central area of the scores plot, thus confirming the reproducibility of the chromatographic data and the robustness of the PCA model. Besides, samples belonging to each cava class were closely grouped (see Figure 3) because of their analogous features and phenolic composition. In more detail, the four rosé cava types were on the right side, white wines including the classical blend and Ch were located at the top left part, and BN samples were at the bottom left part (divided in two groups, with and without MLF, respectively). As a result, the classical blend of Ma, Xa and Pa and the monovarietal PN, BN and Ch were quite separated from each other, in agreement with differences in composition. The evolution of class groups as a function of the Ch percentage was also evidenced from the scores plot. For instance, blends with increasing amounts of Ch (e.g., E, A and S classes) were found in intermediate positions between pure Ch (class G) and Ma, Xa and Pa (blend C).

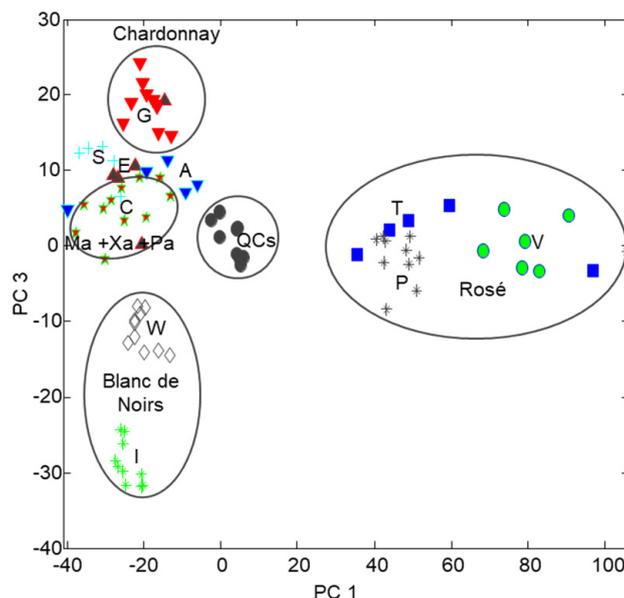


Figure 3. Principal component analysis of the fingerprinting dataset consisting of chromatograms recorded at 310 nm within the working range 6.30–27.30 min. Plot of scores of PC1 versus PC3. Coupage code: C = Ma, Xa and Pa; E = Ma, Xa and Pa with a small percentage of Ch; A = Ma, Xa and Pa (30%) and Ch (70%); G = 100% Ch; S = Ma and Xa (50%) with Ch (50%); T = Ch (30%) and PN (70%); V = PN, Ga and Tr; W = BN (MLF); I = BN (non MLF).

3.1.2. Study of Compositional Profiles

Another PCA model was carried out using concentrations or peak areas of various relevant components, including well-known phenolic compounds, flavonoids and other related species. The study was initially conceived considering the list of 19 compounds given in the experimental section (see Section 2.1), which were chosen as they could be present in white wines. From this list, phenolic acids such as gallic, gentisic, vanillic, caftaric, caffeic, coumaric acids (see structures in Figure 1) as well as various unknown related species were detected at 310 nm. In contrast, flavonoids such as rutin, myricetin, quercetin, kaempferol and apigenin absorbing at 370 were not found; also, flavanols monitored at 280 nm, such as catechin and epicatechin, and components such as resveratrol were undetectable.

For this PCA study, data from those samples belonging to the same class were averaged to provide a simplified and more robust description of cava behavior. Hence, the X-matrix was composed of 10 sample classes and 10 compounds as the variables (see Table 2). The autoscaled PCA model retained 68.3%, 17.0%, 9.3% of data variance on PC1, PC2 and PC3, respectively. Plots of scores and loadings of PC1 vs. PC3 are depicted in Figure 4. It was first concluded that PC1 explained the overall concentration of phenolic compounds which increased from left to right. PC3 described differences in the distribution of components. Ch, PN and BN with and without MLF were in extreme positions in the plot of scores (right, down, left and top, respectively); the classical blend of Ma, Xa and Pa was in the left part of the graph. The gradation in the Ch percentage was also observed in the classical and PN cases.

Table 2. Average composition of phenolic acids in the cava samples.

Class	Composition	Gallic Ac. (mg L ⁻¹)	Caftaric Ac. (mg L ⁻¹)	Gentisic Ac. (mg L ⁻¹)	Vanillic Ac. (mg L ⁻¹)	Caffeic Ac. (mg L ⁻¹)	Coumaric Ac. (mg L ⁻¹)	Unknown1 (mg L ⁻¹) *	Unknown2 (mg L ⁻¹) *	Unknown3 (mg L ⁻¹) *	Unknown4 (mg L ⁻¹) *
C	100% Macabeu, Xarel·lo, Parellada	3.58	8.09	39.12	0.87	1.09	0.47	0.22	5.42	0.27	0.17
G	100% Chardonnay	3.83	14.62	32.70	1.44	2.18	1.01	0.12	0.00	0.31	0.16
I	100% Blanc de Noirs	3.90	8.64	38.45	0.12	1.49	0.65	0.33	4.50	0.74	0.06
P	100% Pinot Noir	5.29	13.68	35.65	0.22	1.98	0.83	0.93	0.86	1.22	0.10
W	100% Blanc de Noirs	4.39	10.15	31.35	0.82	1.30	0.56	0.47	2.87	0.86	0.10
A	70% Macabeu, Xarel·lo, Parellada/30% Chardonnay	3.79	12.46	32.02	1.30	0.84	0.40	0.13	0.80	0.66	0.13
E	85% Macabeu, Xarel·lo, Parellada/15% Chardonnay	3.61	8.36	39.05	1.21	1.18	0.47	0.12	3.91	0.56	0.10
S	50% Macabeu, Xarel·lo/50% Chardonnay	3.84	10.66	28.35	1.17	1.18	0.89	0.15	1.48	0.20	0.13
T	70% Pinot Noir/30% Chardonnay	4.47	8.24	33.11	0.16	1.85	0.79	0.67	4.06	1.03	0.09
V	Monastrell/Garnatxa negra/Trepat	4.79	11.55	33.61	0.37	1.80	0.83	0.72	1.91	0.98	0.10
QC	Representative mixture of all of them	3.89	9.49	34.03	0.90	1.32	0.59	0.26	2.64	0.47	0.10

* expressed as caffeic acid equivalents, i.e., mg mg L⁻¹ caffeic acid.

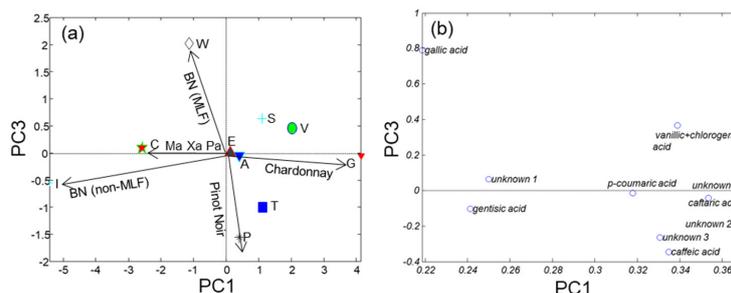


Figure 4. Principal component analysis of the profiling dataset consisting of concentrations and peak areas of relevant compounds. (a) Plot of scores of PC1 versus PC3. (b) Plot of loadings of PC1 versus PC3. Coupage code: see Figure 2.

Variables nearby in the loadings plot were positively correlated such as, in the right side, caftaric and caffeic acids ($r = 0.73$), *p*-coumaric and caffeic acids ($r = 0.90$) or caftaric acid and unknown compound 2 ($r = 0.87$), possibly because they belong to the same family of hydroxycinnamic derivatives. On the left side, gentisic acid and unknown 1 were found pretty close ($r = 0.84$). In contrast, gallic and vanillic acids showed a quite independent behavior as they were uncorrelated with any other variable.

Despite the lack of specificity, the most characteristic markers of each cava type could be deduced from the simultaneous analysis of scores and loadings. In agreement with the previous model, it was concluded that Ch is rich in caftaric acid, *p*-coumaric acid, vanillic acid and caffeic acid, whereas, gentisic acid is more abundant in the classical blend. On the other hand, rosé samples have a different phenolic pattern. Finally, in the case of BN samples, increased levels of gentisic acid without MLF and gallic acid with MLF were encountered.

3.2. Classification Studies by PLS-DA

This section is focused on the assessment of classification models by PLS-DA as an initial survey towards cava authentication. As a proof of concept, various classification cases were designed to assign cava samples to pre-established classes. As a general strategy, the set of cava samples was divided into two subgroups for calibration and prediction which accounted ca. 70% and 30% of the samples (randomly selected), respectively. The calibration set contained well-known cava samples belonging to the different classes which were used to create a classification model using chromatographic fingerprints as the data. Samples included in the prediction set were considered as unknowns to be assigned. As detailed below, Blanc de Noirs vs. non-Blanc de Noirs and Chardonnay vs. non-Chardonnay cases were selected as model examples to evaluate the classification performance of PLS-DA based on phenolic information.

3.2.1. Blanc de Noirs Versus Other Cavas

This model intended the recognition of monovarietal BN cava samples from the rest of blends including the other varieties, so, two classes were defined accordingly: (i) BN, both subjected or not to MLF and (ii) non-BN including the remainder of cava samples from the other blends. The calibration set was composed of 54 samples, including 9 QC replicates. X-data consisted of chromatograms taken at 310 nm within the working time window (6.3 to 27.3 min). Y-data described the belonging to BN and non-BN classes with the code 0 and 1, respectively. The prediction set consisted of 25 “unknown” samples to be identified as BN and non-BN members.

The estimation of the optimum number of latent variables by cross validation according to a Venetian blind approach indicated that the minimum of the root mean standard error in cross validation (RMSECV) was achieved at 4 LVs, so this number was used in both calibration and prediction steps by PLS-DA. The assignment plot in Figure 5 showed that all the samples used for calibration were correctly located, down and up of the threshold line those belonging to the Blanc de Noirs class and non-Blanc de Noirs class, respectively. Analogously, “unknown” samples of the prediction set were

also assigned correctly to BN and non-BN classes. It should be noted that this classification model grouped both BN samples with and without MLF. Differences in the phenolic composition due to MLF processes were less noticeable than those found with respect to the other classes.

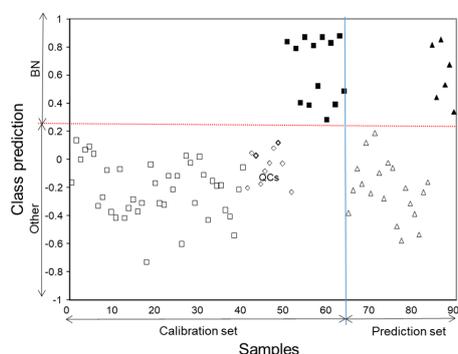


Figure 5. Classification plot of cava samples into Blanc de Noirs (top) and non- Blanc de Noirs (bottom) classes by PLS-DA using 4 LVs. Horizontal dotted line indicates the classification threshold separating the two classes. Vertical solid line separated calibration (left) and prediction (right) sets.

3.2.2. Chardonnay Versus Other Cavas

This example aims at identifying pure (monovarietal) Ch cavas from other classes, including those elaborated with other varieties and blends containing different proportions of Ch. This model presented high mathematical complexity, as cava with high percentages of Ch could be confounded with the pure ones and vice versa. Two classification classes were defined accordingly, referred to as Ch (pure Chardonnay) and non-Ch (other cava wines). In total, 64 samples belonging to the different blends (including 9 QC replicates) and 24 “unknown” samples were used in the calibration and prediction sets, respectively. As in the previous case, chromatographic data taken at 310 nm within the time window 6.30 to 27.30 min were used in the X-matrix. The Y-data matrix coded the belonging to Ch and non-Ch classes with 1 and 0, respectively.

The optimum number of LVs, corresponding to the minimum of RMSECV function, was 6. Results of both calibration and prediction steps are summarized in the assignment plot of Figure 6. As it can be seen, all the samples used for calibration were correctly predicted (just one Ch cava was on the edge of the threshold line separating the two classes). In a similar way, “unknown” samples could also be identified successfully. Besides, it should be mentioned that non-Ch samples located next to the decision boundary line consisted of blends containing high percentages of Ch base wine in their composition. Since their chemical composition was quite similar to pure Ch, they appeared very close to the Ch samples.

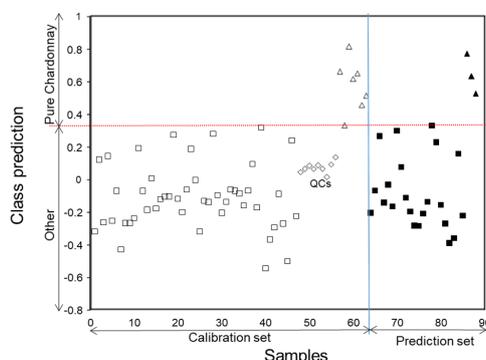


Figure 6. Classification plot of cava samples into pure Chardonnay (top) and others (bottom) classes by PLS-DA using 6 LVs. Horizontal dotted line indicates the classification threshold separating the two classes. Vertical solid line separated the calibration (left) and the prediction (right) set.

4. Conclusions

This work proposed a new strategy to identify and assign cava samples to predefined classes using phenolic data as the source of information. It was found that the composition of hydroxybenzoic, hydroxycinnamic acid depended on the cava varieties and blends so such data can be exploited to extract relevant patterns about the behavior of samples with the assistance of chemometric methods. First, exploratory studies by PCA revealed that cavas could be distinguished as a function of blends and phenolic compounds were good descriptors in terms of MLF. Subsequently, the discrimination of cavas by PLS-DA was evaluated in two cases of different complexity. The first one dealt with the identification of monovarietal Blanc de Noirs samples, with respect to other classes that did not contain this variety. The second one dealt with the detection of monovarietal Chardonnay cavas in front of other ones. In any case, excellent results in terms of selectivity and sensitivity were achieved with classification rates of 100%.

To summarize, phenolic data combined with chemometric analysis resulted in an excellent approach to classify cava wines according to wine varieties and blends and winemaking process in terms of MLF. On the other hand, it should be emphasized that the method is not limited to the cases presented here but it can be adapted to more complex studies. In this regard, we believe that phenolic profiles will result to be a great source of information for sample authentication to detect adulterations and labeling frauds.

Author Contributions: Conceptualization, A.I.-L. and J.S.; methodology, A.I.-L. and J.S.; formal analysis, A.I.-L. and J.S.; investigation, A.I.-L.; resources, J.S.; data curation, J.S.; writing—original draft preparation, A.I.-L.; writing—review and editing, J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors gratefully acknowledge Codorníu Raventós winery to provide the samples.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Institut Del Cava. Available online: <http://www.institutdelcava.com/en/> (accessed on 20 June 2020).
2. Saurina, J. Characterization of wines using compositional profiles and chemometrics. *Trend Anal. Chem.* **2010**, *29*, 234–245. [[CrossRef](#)]
3. Versari, A.; Laurie, V.F.; Ricci, A.; Laghi, L.; Parpinello, G.P. Progress in authentication, typification and traceability of grapes and wines by chemometric approaches. *Food Res. Int.* **2014**, *60*, 2–18. [[CrossRef](#)]
4. Hong, E.; Lee, S.Y.; Jeong, J.Y.; Park, J.M.; Kim, B.H.; Kwon, K.; Chun, H.S. Modern analytical methods for the detection of food fraud and adulteration by food category. *J. Sci. Food Agric.* **2017**, *97*, 3877–3896. [[CrossRef](#)] [[PubMed](#)]
5. Kamiloglu, S. Authenticity and traceability in beverages. *Food Chem.* **2019**, *277*, 12–24. [[CrossRef](#)]
6. Palade, L.M.; Popa, M.E. Polyphenol Fingerprinting Approaches in Wine Traceability and Authenticity: Assessment and Implications of Red Wines. *Beverages* **2018**, *4*, 75. [[CrossRef](#)]
7. Romisch, U.; Jager, H.; Capron, X.; Lanteri, S.; Forina, M.; Smeyers-Verbeke, J. Characterization and determination of the geographical origin of wines. Part III: Multivariate discrimination and classification methods. *Eur. Food Res. Technol* **2009**, *230*, 31–45. [[CrossRef](#)]
8. Tzouros, N.E.; Arvanitoyannis, I.S. Agricultural produces: Synopsis of employed quality control methods for the authentication of foods and application of chemometrics for the classification of foods according to their variety or geographical origin. *Crit. Rev. Food Sci. Nutr* **2001**, *41*, 287–319. [[CrossRef](#)]
9. Cozzolino, D.; Cynkar, W.; Shah, N.; Smith, P. Technical solutions for analysis of grape juice, must, and wine: The role of infrared spectroscopy and chemometrics. *Anal. Bioanal. Chem.* **2011**, *401*, 1475–1484. [[CrossRef](#)]
10. Wang, L.; Sun, D.W.; Pu, H.B.; Cheng, J.H. Quality analysis, classification, and authentication of liquid foods by near-infrared spectroscopy: A review of recent research developments. *Crit. Rev. Food Sci. Nutr.* **2017**, *57*, 1524–1538. [[CrossRef](#)]
11. Drivelos, S.A.; Georgiou, C.A. Multi-element and multi-isotope-ratio analysis to determine the geographical origin of foods in the European Union. *Trend Anal. Chem.* **2012**, *40*, 38–51. [[CrossRef](#)]

12. Marcone, M.F.; Wang, S.A.; Albabish, W.; Nie, S.P.; Somnarain, D.; Hill, A. Diverse food-based applications of nuclear magnetic resonance (NMR) technology. *Food Res. Int.* **2013**, *51*, 729–747. [[CrossRef](#)]
13. Ogrinc, N.; Kosir, I.J.; Spangenberg, J.E.; Kidric, J. The application of NMR and MS methods for detection of adulteration of wine, fruit juices, and olive oil. A review. *Anal. Bioanal. Chem.* **2003**, *376*, 424–430. [[CrossRef](#)] [[PubMed](#)]
14. Urickova, V.; Sadecka, J. Determination of geographical origin of alcoholic beverages using ultraviolet, visible and infrared spectroscopy: A review. *Spectrochim. Acta A* **2015**, *148*, 131–137. [[CrossRef](#)] [[PubMed](#)]
15. Borràs, E.; Ferré, J.; Boqué, R.; Mestres, M.; Aceña, L.; Busto, O. Data fusion methodologies for food and beverage authentication and quality assessment—A review. *Anal. Chim. Acta* **2015**, *891*, 1–14. [[CrossRef](#)]
16. Martelo-Vidal, M.J.; Vazquez, M. Polyphenolic Profile of Red Wines for the Discrimination of Controlled Designation of Origin. *Food Anal. Methods* **2015**, *9*, 332–341. [[CrossRef](#)]
17. Buscema, F.; Boulton, R.B. Phenolic Composition of Malbec: A Comparative Study of Research-Scale Wines between Argentina and the United States. *Am. J. Enol. Viticult.* **2015**, *66*, 30–36. [[CrossRef](#)]
18. Pavlousek, P.; Kumsta, M. Authentication of Riesling Wines from the Czech Republic on the Basis of the Non-flavonoid Phenolic Compounds. *Czech J. Food Sci.* **2013**, *31*, 474–482. [[CrossRef](#)]
19. Serrano-Lourido, D.; Saurina, J.; Hernandez-Cassou, S.; Checa, A. Classification and characterisation of Spanish red wines according to their appellation of origin based on chromatographic profiles and chemometric data analysis. *Food Chem.* **2012**, *135*, 1425–1431. [[CrossRef](#)]
20. Jaitz, L.; Siegl, K.; Eder, R.; Rak, G.; Abranko, L.; Koellensperger, G.; Hann, S. LC-MS/MS analysis of phenols for classification of red wine according to geographic origin, grape variety and vintage. *Food Chem.* **2010**, *122*, 366–372. [[CrossRef](#)]
21. Diaz, R.; Gallart-Ayala, H.; Sancho, J.V.; Nunez, O.; Zamora, T.; Martins, C.P.B.; Hernandez, F.; Hernandez-Cassou, S.; Saurina, J.; Checa, A. Told through the wine: A liquid chromatography-mass spectrometry interplatform comparison reveals the influence of the global approach on the final annotated metabolites in non-targeted metabolomics. *J. Chromatogr. A* **2016**, *1433*, 90–97. [[CrossRef](#)]
22. Agatonovic-Kustrin, S.; Milojkovic-Opsenica, D.; Morton, D.W.; Ristivojevic, P. Chemometric characterization of wines according to their HPTLC fingerprints. *Eur. Food Res. Technol.* **2017**, *243*, 659–667. [[CrossRef](#)]
23. Ivanovic, M.; Razborsek, M.I.; Kolar, M. Chemometric Characterization of Slovenian Red Wines. *Acta Chim. Slov.* **2016**, *63*, 661–669. [[CrossRef](#)] [[PubMed](#)]
24. Arribas, A.S.; Martinez-Fernandez, M.; Moreno, M.; Bermejo, E.; Zapardiel, A.; Chicharro, M. Classification of Spanish white wines using their electrophoretic profiles obtained by capillary zone electrophoresis with amperometric detection. *Electrophoresis* **2014**, *35*, 1693–1700. [[CrossRef](#)] [[PubMed](#)]
25. Franquet-Griell, H.; Checa, A.; Nunez, O.; Saurina, J.; Hernandez-Cassou, S.; Puignou, L. Determination of Polyphenols in Spanish Wines by Capillary Zone Electrophoresis. Application to Wine Characterization by Using Chemometrics. *J. Agric. Food Chem.* **2012**, *60*, 8340–8349. [[CrossRef](#)] [[PubMed](#)]
26. Arribas, A.S.; Moreno, M.; Moreno, G.A.; Bermejo, E.; Zapardiel, A.; Chicharro, M. Characterization of White Wines by Electrochemical Indexes Obtained Using Carbon Nanotube-modified Electrodes. *Electroanalysis* **2018**, *30*, 1461–1471. [[CrossRef](#)]
27. Ceto, X.; Cespedes, F.; del Valle, M. BioElectronic Tongue for the quantification of total polyphenol content in win. *Talanta* **2012**, *99*, 544–551. [[CrossRef](#)]
28. Restuccia, D.; Sicari, V.; Pellicano, T.M.; Spizzirri, U.G.; Loizzo, M.R. The impact of cultivar on polyphenol and biogenic amine profiles in Calabrian red grapes during winemaking. *Food Res. Int.* **2017**, *102*, 303–312. [[CrossRef](#)]
29. Larrauri, A.; Nunez, O.; Hernandez-Cassou, S.; Saurina, J. Determination of Polyphenols in White Wines by Liquid Chromatography: Application to the Characterization of Alella (Catalonia, Spain) Wines Using Chemometric Methods. *J. AOAC Int.* **2017**, *100*, 323–329. [[CrossRef](#)]
30. Izquierdo-Llopart, A.; Saurina, J. Characterization of Sparkling Wines According to Polyphenolic Profiles Obtained by HPLC-UV/Vis and Principal Component Analysis. *Foods* **2019**, *8*, 22. [[CrossRef](#)]

