Use of untargeted LC-MS metabolome to discriminate Italian mono-varietal red wines, produced in their different terroirs

Journal:	*Journal of Agricultural and Food Chemistry*
Manuscript ID	jf-2020-008799
Manuscript Type:	Article
Date Submitted by the Author:	08-Feb-2020
Complete List of Authors:	Arapitsas, Panagiotis; Fondazione Edmund Mach, Food Quality and Nutrition Ugliano, Maurizio; University of Verona, biotechnology Marangon, Matteo; Universita degli Studi di Padova, DAFNAE Piombino, Paola; Università degli Studi di Napoli Federico II, Dipartimento di Scienza degli Alimenti Rolle, Luca; University of Turin, DISAFA - Dipartimento di Scienze Agrarie, Forestali e Alimentari Gerbi, Vincenzo; delle Risorse Agroforestali, Dip.di Valorizzazione e Protenzione Versari, Andrea; University of Bologna, Food Science Mattivi, Fulvio; University of Trento, Center Agriculture Food Environment

Use of untargeted LC-MS metabolome to discriminate Italian mono-varietal red wines, produced in their different terroirs

Panagiotis Arapitsas[1*], Maurizio Ugliano[2], Matteo Marangon[3], Paola Piombino[4], Luca Rolle[5], Vincenzo Gerbi[5], Andrea Versari[6], Fulvio Mattivi[1,7].

[1]*Department of Food Quality and Nutrition, Research and Innovation Centre, Fondazione Edmund Mach (FEM), Via E. Mach 1, 38010 San Michele all´Adige, (Italy)*

[2]*Department of Biotechnology, University of Verona, (Italy)*

[3]*Department of Agronomy, Food, Natural Resources, Animals and Environment (DAFNAE), University of Padova, (Italy)*

[4]*Department of Agricultural Sciences, Division of Vine and Wine Sciences, University of Naples Federico II, Viale Italia, 83100 Avellino, (Italy)*

[5]*Dipartimento di Scienze Agrarie, Forestali e Alimentari, Università degli Studi di Torino, Largo Paolo Braccini 2, 10095 Grugliasco, TO, (Italy)*

[6]*Department of Agricultural and Food Sciences, University of Bologna, Piazza Goidanich 60, Cesena, FC 47521, (Italy)*

[7]*Department of Cellular, Computational and Integrative Biology - CIBIO, University of Trento, 38123 Povo, (Italy)*

Abstract

Aim of this project was to register for the first time in a LC-MS based untargeted single-batch analysis the metabolome of 11 single-cultivar Italian red wines (Aglianico, Cannonau, Corvina, Montepulciano, Nebbiolo, Nerello, Primitivo, Raboso, Sagrantino, Sangiovese and Teroldego) from 12 regions across Italy, each one produced in their terroirs under ad hoc legal frameworks to guarantee the quality and origin. The data provided indications about the similarity between the cultivars, and highlighted a rich list of putative Biomarkers of Origin Wines (pBOWs) characterizing each single combination cultivar-terroir, where Primitivo, Teroldego and Nebbiolo had the maximum number of unique pBOWs. The pBOWs included anthocyanins (Teroldego), flavanols (Aglianico, Sangiovese, Nerello and Nebbiolo), amino acids and N-containing metabolites (Primitivo), hydroxycinnamates (Cannonau) and flavonols (Sangiovese). The raw data generated in this study are publicly available, enabling the accessibility and reusability, and serve as a baseline dataset for future investigations.

Keywords: mass spectrometry, wine authenticity; bioinformatics; wine metabolomics; amines; polypehonols

Introduction

Italy is worldwide one of the most important countries in viticulture and oenology, with 705 thousand ha of vineyards (4th place), 8.6 million tons of grape production (2nd place), 54.8 million hl wine production (1st place) and 22.4 million hl of wine consumption according to OIV Focus for 2018.[1] Moreover, Italy is one of the richest countries in terms of number of grape cultivars, since according to the Italian National Catalogue of Grapevine Varieties, nowadays, over five hundred cultivars compose the Italian ampelographic platform.[2] Wine has a straight and tight correlation with the Italian culture already from the 2nd century BC, so all regions produce their own wine with their local cultivars, and according to the characteristics of the territory, the culinary habits, the tradition and human needs. During the centuries, the wine production of each region further evolved and differentiated from the others, creating the multi-oenological Italian culture of today, characterized by the presence of 525 **origin wines**, protected as intellectual property rights either as Denominazione di origine Controllata e Garantita (DOCG, n=74), or Denominazione di Origine Controllata (DOC, n=333), or Indicazione Geografica Tipica (IGT, n=118).[3]

In terms of grapes employed in wine production, Sangiovese is the major Italian cultivar with 54,000 ha all over Italy (including Tuscany and Romagna). From Sangiovese are produced famous Italian wines, like Brunello di Montalcino and Chianti Classico. Nebbiolo is mainly cultivated in Piedmont and from the 6,047 cultivated ha are produced iconic wines like Barolo and Barbaresco. Corvina grapes (6,695 ha) participate in the production of Amarone and Valpolicella in Veneto. In central and southern Italy, Montepulciano (27,434 ha) is the major red cultivar of Abruzzo, Primitivo (16,321 ha) of Puglia, Aglianico (9,947 ha) of Campania and Cannonau (6128 ha) of Sardinia.[1] Teroldego (627 ha), Raboso (~500 ha), Sagrantino (930 ha) and Nerello Mascalese (2,942 ha) are minor Italian cultivars, in term of volume of production, cultivated mainly in restricted areas of Trentino, Veneto, Umbria and Sicily, respectively.[1] In 2015, the above-mentioned cultivars accounted the 44% of the red grape vine-cultivated area of Italy, so they cover a representative portion of the Italian oenological biodiversity (Figure 1).

Wine, being the final product of a long and multistep process, has one of the richest and more complex metabolomic fingerprint. Several targeted protocols focused on the analysis of polyphenols, volatiles, lipids and etc. have been applied in order to find differences between wines coming from different grape cultivars, as well as understanding the chemical and sensorial character of mono-cultivar wines.[4–8] Over the last years, untargeted analytical approaches proved a valuable and powerful alternative for the study of wine metabolome.[9–12] Techniques such as LC-MS, GC-MS or direct injection FTICR-MS based metabolomics allowed identification of new wine metabolites,[13,14] discrimination of groups of wines,[14–17] elucidation of chemical reaction occurring during aging and storage[13,14,16,18,19] also in relationship to packaging[14], providing novel insights in wine history[20] and quality.[13,14,16,21,22] Some wines of the above mentioned Italian cultivars have been subject of untargeted LC-MS based analysis, alone or as groups together with other 2-3 cultivars, but the literature lacks of studies that combine a large part of the red Italian wines diversity. Historically, the most promising markers for the chemical characterization of varietal wines have been discovered trying to compare the presence of a few targeted metabolites in varietal wines. As an example, a pioneering study[23] based on the analysis of the variance of 20 organic acids and esters in six red wines, led to the discovery that shikimic acid was associated with the cultivar, and in particular useful to discriminate the Pinot noir wines. It is expected that the application of an untargeted method, capable to produce a semi-quantitative analysis of ca. 1000 metabolites, has the potential to support the discovery of several putative Biomarkers of Origin Wines (pBOWs).

Initially the aim of this project was to register for the first time the LC-MS metabolomic fingerprint of 11 mono-cultivar Italian red wines from 12 regions that representing a large portion of the Italian red wine production and biodiversity. Supplementary aim was to investigate the produced dataset in order to provide information about the metabolomic space similarity and dissimilarity between the studied wines, and extract pBOWs. Additional scope was to make the dataset public available with the intention to help other researchers.

Materials and Methods

Wine samples

A total of 110 Italian red wines, 100% mono-varietal, all vinified in 2016 from 11 diverse Italian grape varieties harvested in the corresponding main geographical areas of production (12 wine regions), were sampled directly from the producers. The wine sample set included: 11 Teroldego (TER) from Trentino-Alto Adige; 7 Corvina (COR) from Veneto; 10 Raboso Piave (RAB) from Veneto); 11 Nebbiolo (NEB) from Piedmont; 7 Sangiovese (SAT) from Tuscany; 12 Sangiovese (SAR) from Romagna; 10 Sagrantino (SAG) from Umbria; 9 Montepulciano (MON) from Abruzzo; 9 Cannonau (CAN) from Sardinia; 10 Aglianico (AGL) from Campania; 11 Primitivo (PRI) from Puglia; and 3 Nerello Mascalese (NER) from Sicily. The basic oenological information about the wine are in Supplementary material Table S1 and Figure S1. The mid-infrared spectroscopy data can be found in Parpinello et al.[24] Winemaking was carried out by each winery independently and according to their standard production practices. However, for each wine the following specifications were followed: a) wines had to be obtained from one single grape variety; b) wines should be fermented in stainless steel vats; c) fermentation should be run in industrial scale; d) the sampling should be preferentially made before malolactic fermentation; e) wines should not have any contact with oak; f) 50 mg/L of free SO_2 had to be added at the time of sampling, before bottling in dark glass bottles; g) Nomacorc Select Bio 500 (Nomacorc, France) closures had to be used. The sampling occurred in early 2017 and the wines were stored at 4 ^{o}C until analysis. All analysis were completed within a single batch, in 3 months after the sampling.

UPLC-QTOF MS analysis

Sample preparation followed a previously described protocol[11] and all steps until the LC/MS vial filling occurred under nitrogen atmosphere. Wines were uncorked and an aliquot was transferred into a 15 mL amber vial (filled to its capacity). Then a quality control (QC) pooled sample was prepared by pooling 1 mL of each wine. Then 1 mL of each wine sample/QC was diluted with 2 mL Milli-Q sonicated water and was finally filtered with 0.2 µm PTFE filters into a 2 mL amber vial (MS certificated) prior to LC-MS analysis.

The analysis followed a previously described protocol.[11,13] A Waters Acquity UPLC coupled via an electrospray ionization (ESI) interface to a Synapt HDMS QTOF MS (Waters, Manchester, UK) operating in W-mode and controlled by MassLynx 4.1 was used. The column was a reversed phase (RP) ACQUITY UPLC 1.8 µm 2.1 x 150 mm HSS T3 column (Waters); column manager was set at 40 ºC; the mobile phase flow rate was 0.28 mL/min; and the eluents was water and methanol both with 0.1% formic acid. The multistep linear gradient used was as follows: 0-1 min, 100% A isocratic; 1-3 min, 100-90 % A; 3-18 min, 90-60 % A; 18-21 min, 60-0 % A; 21-25.5 min, 0 % A isocratic; 25.5-25.6 min, 0-100 % A; 25.6-28 min 100% isocratic. Injection volume was 5 µL and the samples were kept at 4 ºC throughout the analysis. Mass spectrometry data were collected by separate runs in positive and negative ESI mode over a mass range of 50 to 2000 amu with scan duration of 0.4 s in centroid mode. The transfer collision energy and trap collision energy were set at 6 V and 4 V. The source parameters were set as follows: capillary 3 kV for positive scan and 2.5 kV for negative scan, sampling cone 25 V, extraction cone 3V, source temperature 150 ºC, desolvation temperature 500 ºC, desolvation gas flow 1000 L/h and nebulizer gas 50 L/h. External calibration of the instrument was performed at the beginning of each batch of analysis by direct infusion of a sodium formate solution (10 % formic acid/0.1 M NaOH/Acetonitrile at a ratio of 1/1/8), controlling the mass accuracy from 40 to 2000 m/z (less than 5 ppm) and mass resolution (over 14000 FWHM). LockMass calibration was applied using a solution of leucine enkephaline (0.5 mg/L, m/z 556.2771 for positive and 554.2620 for negative ion mode) at 0.1 mL/min. The QC samples were used for the LC-MS system initial equilibration (4-5 injections) and control at regular intervals (one QC sample injection every 6 real sample injections) during the sequence, according to the quality control flowchart.[11] In total, the public available database included 26 QC sample analysis for the ESI- and 24 analysis for the ESI+ mode (the system equilibration QC injections were excluded).

Data analysis

For quality control during the runs and data analysis, we used PCA (Principal Component Analysis) plots generated by Progenesis QI (Version 2.0.0.0.0, nonlinear Dynamics), checking the distribution/clustering of the QC injections.[11] Progenesis QI parameters used for alignment were

done on default mode by Progenesis QI with peak picking performed at maximum level; and the first minute and the last six minutes of the run excluded from data processing (only the range 1-22 min were used). Putative BOWs were considered the “compounds” that according to the Progenesis QI statistical analysis had max fold range ≥ 2 and Anova (p value ≤ 0.01). Progenesis QI views as “compound” a group of isotopic and adducts features belonging to the same metabolite.

Annotation was performed manually by comparing retention times and mass spectra accuracy with a mass tolerance of 5 ppm, based on the group’s previous experience with the specific instrumentation mass resolution,[25] and in accordance with the 4 levels described by Sumner et al.[26].

Known wine metabolites previously annotated by using the same protocol[11,13,14,27,28] were integrated semi-manually using the TargetLynx tools of Waters MassLynx 4.1 software (Milford, MA). The TargetLynx parameters were set at chromatogram mass window 0.08 Dalton; retention time window ± 0.2 min; smoothing iterations 1; and smoothing width 2. Further statistical analysis was performed on these integrated peaks by using MetaboAnalyst online platform version 4.0 (http://www.metaboanalyst.ca/,[29]) without normalization, missing values estimation and data transformation, by using Pareto scaling. For the Heatmap the Euclidean distance and Ward clustering algorithm were used.

Raw LC–MS data and other details will be made publicly available for download with the accession number MTBLS1443 from the MetaboLights public repository (http://www.ebi.ac.uk/metabolights/,[21,30]).

Result and Discussion

A central starting point of this study was to obtain a set of wine samples that was as representative as possible of the diversity of Italian red wine production both in terms of relevant varieties and areas of origin. As shown in Figure 1, the samples included regions of northern (Piedmont, Trentino and Veneto), central (Tuscany, Emilia-Romagna and Umbria), and southern (Campania

and Abruzzo) Italy, also with the two major islands (Sicily and Sardinia). In the case of Sangiovese, the most important red variety in Italy, two different production areas, namely Tuscany and Emilia Romagna, were considered. Wines were obtained from different wineries located in the production area, so that they could be considered true representations not only of the varietal characteristics but also of the winemaking practices commonly adopted in each area at winery level, and in agreement with the rules of the specific denomination of origin. In order to avoid possible differences deriving from aging and storage modalities, all samples were collected directly from the tank, without any previous contact with wood, and were bottled in the laboratory under the same conditions.

The applied LC-MS protocol proved several times in the past years its capability to register wine metabolome and generated various new hypothesis.[11,13,14,28] As stated by this protocol, one of the most crucial issues in untargeted LC-MS analysis is to inject all samples in a single batch. Due to this methodological constraint, in this project it was decided to analysed only the wines produced in one harvest. The number of biological replicates, i.e. different wines produced from different vineyards and/or different wineries, was in the range 7-12 (mean 9.7) for all the wine regions, with the sole exception of Nerello Mascalese from Sicily, for which only three suitable batches of wines were obtained.

According to the workflow, followed in our laboratory, before any further data analysis it is important to verify the quality of the dataset. Figure 2 shows the PCA plots of the sample injections distribution according to a multivariate and unsupervised principal component analysis. The PCA plot of the ESI+ analysis was performed using 11274 features and the ESI- 7397 features, and in both cases, the QC sample injections – injected all over the sequence - formed a tide cluster, proving the reliability of the measure, in term of absence of fluctuations for samples injected at different time points. According to this unsupervised analysis, it was possible to notice that Teroldego and Primitivo wine groups had a metabolomic fingerprint very different in respect to the other wines.

In order to investigate the metabolites that differentiated each wine group from the other we used supervised data analysis tools. By using the Anova tool of Progenesis QI, the metabolomic fingerprint of each wine group was compared against all the other groups, so a subgroup of features was created by using only the features with *p*-value ≤ 0.01 and fold change ≥ 2. The different lists were merged and created the Supplementary Tables S2-3. The ESI- analysis included 621 pBOWs and the ESI+ 1735 pBOWs. Figure 3 demonstrated the major outcome of this data analysis. For the ESI+ analysis it was possible to detect also pBOWs unique for each group of wine, while for ESI- that was not possible since Primitivo included all the pBOWs and did not have any unique. In fact, both ESI- and ESI+ shown that Primitivo had the highest number of pBOWs. This result was also in accordance both with the PCA plots (Figure 2) where Primitivo samples are separated from the other cultivars by PC1; and the hierarchical cluster analysis (Figure 4), where Primitivo sample is the first group of samples to slit from the others. In detail, Primitivo has 727 features pBOWs (226 of them unique) for ESI+ and 621 for ESI-. Teroldego and Nebbiolo also had a big number of pBOWs, and on the other hand, Montepulciano and Corvina had the smallest number of pBOWs.

The hierarchical cluster analysis (Figure 4) showed that the Primitivo group was the one differing the most for both the ESI- and ESI+ analysis. A second cluster in ESI+ included Nebbiolo, Corvina, Raboso and Sangiovese wines. Such behaviour should be attributes to the fact that these cultivars are known for their not very intense red colour[5] and because in ESI+ mode the positive charged anthocyanins give very good signal. Therefore, the here observed clustering was most probably strongly driven by the red coloured and positive charged anthocyanins. The result that Teroldego, a very rich cultivar in anthocyanins,[5] formed a cluster alone, supports this hypothesis. These findings indicated to us that we should investigate the anthocyanins and related pigments in detail. In ESI- Teroldego was the second more distant cluster, while Nebbiolo, Nerello and Sangiovese clustered again as nearest neighbours in the dendrogram (Figure 4).

The annotation process of the pBOWs showed that several of the metabolites belongs to the chemical classes of polyphenols, amino acids, dipeptides, tripeptides, bounded terpenoids, sugars and organic acids (Supplementary Tables S2-3). Therefore it was decided to take advantage of the annotation achieved previously using the same protocol in oenological projects and to study more in depth these groups of known metabolites.[11,13,14,21,25,28] With this aim, we turned back to the raw files and integrated a big number of metabolites. This integration process was independent to Progenesis QI workflow, therefore this was also a way to manually check the possible presence of false positive and false negatives markers. Then the integrated areas peak table was uploaded to the Metaboanalyst platform for further statistical analysis and data visualization.

Figures 5-7 show the (bio)synthetic pathway of several metabolites of oenological interest, annotated and detected as markers in this study. For each metabolite, data from the heatmap of Supplementary Figures S1 is also shown, in order to compare the relative concentration of each metabolite in the different wine groups. Concerning the amino acids included in Figure 5, Primitivo was the group with the highest amount of leucine, arginine, tyrosine, valine and phenylalanine. On the opposite end, the wine groups with the smallest amounts of the same amino acids were Nebbiolo and Sangiovese. We should take in consideration that yeasts could consume the majority of the amino acids during the alcoholic fermentation as a nitrogen source.[31] Thus common oenological practices, such as addition of inorganic and/or organic nitrogen to support yeast's growth would strongly affect the concentration of amino acids in wine.[31] Since the wines from each group originated from different wineries that followed different winemaking practices, we should not exclude that amino acids could be markers to discriminate wines originated from different cultivar. In the past, the amino acids profile has been proposed as a tool to wine discrimination.[31–33] Proline is the only amino acid not consumed by the yeast in anaerobic condition[31] and because of this characteristic it has been used in food frauds analysis.[34] According to our results, Primitivo wines showed relatively low concentration for this amino acid, with Teroldego showing the highest and Nerello the lowest.

Moreover, several di- and tri-peptides were tentatively annotated (3rd level annotation) as markers. According to the nitrogen rule/principle in mass spectrometry, odd *m/z* values indicate organic

compound with odd number of nitrogen (thus at least one) and even *m/z* values indicate organic compound with zero or even number of nitrogen. Of course, this rule is valid for organic compounds containing exclusively H, C, N, O, Si, P, S, and halogen, and for high resolution mass spectrometers it is more accurate for *m/z* values below 500. Primitivo wines pBOWs included several ions with odd *m/z* values (Supplementary Tables S2-3), had the highest concentrations in several amino acids (Figure 5), and the tentatively annotated compounds included di- and tri-peptide. If this issue is characteristic for Primitivo, further experiments are necessary to validate this hypothesis and better understand the composition of Primitivo wines, and the contribution of the cultivar and its terroir in determining this unusually richness in nitrogen compounds. Lately, Sherman et al.[35] discovered that wine sensorial quality was positively correlated with markers annotated as di- and tri-peptides. To validate the hypothesis that amino acid profile could be used to distinguish the cultivar in wines, bit analysis on wines produced in more than one harvest are necessary as well as the used of wines produced under the same winemaking conditions and under well controlled agronomical conditions. Indeed, it is well known that in addition to the cultivar, also the terroir (fertilization with nitrogen, grape maturity, climate and the sanitary status) can greatly influence the concentration in nitrogen containing compounds.[36]

Primitivo and Sagrantino were the richest wines in tryptophan, while Sangiovese and Raboso and Nebbiolo were the poorest. Conversely, Sangiovese wines were the richest in tryptophol, thus the Ehrlich reaction tryptophan product during the alcoholic formation, and Primitivo the poorest (Figure 5). This was an indication that tryptophan was used by the yeast during the alcoholic fermentation of Sangiovese wines.[31] The lower presence of tryptophol in Primitivo wines was expected, since the Ehrlich pathway is not a preferred way of nitrogen assimilation in presence of an abundant content in amino acids in the juice. Moreover, we found that Sangiovese wines were also the richest in sulfonated tryptophol (Supplementary Figure S1), which is a product of the sulfonation of tryptophol and it formation is favoured by oxygen and lower pH.[14,37,38] Primitivo wines were also the richest in two other N-containing metabolites, tryptophan products during the alcoholic fermentation, N-acetyl-tryptophan ethyl ester and tryptophan ethyl ester.[37] Apparently, tryptophan during Primitivo winemaking process turned to these two ethyl esters and not to the

279 fuse alcohol (tryptophol). As expected by our previous experience,[38] the same issue was also valid
280 for tyrosine (Supplementary Figure S1).[37,38]

281 In grapes, tryptophan is transformed to indole-lactic acid (ILA) and its glucosides (ILA-glu), and
282 later these two metabolites can react with SO_2 in wine and give the corresponding sulfonated
283 products (ILA-SO_3H and ILA-glu-SO_3H).[14,38] ILA and ILA-glu concentration depends on the cultivar
284 and climate, and in our experiment Montepulciano, Aglianico and Teroldego showed the highest
285 concentrations (Figure 5). The wines with the highest concentration of the sulfonated ILA-glu-SO_3H
286 were the Corvina wines, followed by Montepulciano and Raboso. The formation in wine of the
287 sulfonated indoles is strongly linked to the oxygen.[14,38]

288 Glutathione is a tripeptide, present in grapes that can also be added to the wine (mainly white
289 wines) as antioxidant to protect aromatic compounds.[31] Lately it was proven that in the presence of
290 SO_2 glutathione can produce its sulfonated analogue. The presence of oxygen can also favour this
291 reaction.[14] Corvina was the group of wines with the highest concentration of both glutathione and
292 its sulfonated analogue (Figure 5).

293 Through the phenylpropanoid pathway, grapevine is able to synthesize several polyphenols of
294 different families. One of the main families are the hydroxycinnamates, which include coutaric acid,
295 caftaric acid and fertaric acid. Sangiovese, Nerello, Raboso and Cannonau were the wines with the
296 highest concentration in mono-substituted (= one –OH to the aromatic ring) coutaric acid; while the
297 di-substituted caftaric acid, the sulfonated caftaric acid, caffeic acid and fertaric acid characterized
298 Cannonau (Figure 5). This is likely a character derived from the cultivar, since Cannonau grapes
299 belong to the family of Grenache/Garnacha grapes, known to be with the *Vitis vinifera* among the
300 richest in hydroxycinnamates.[39,40]

301 Primitivo showed the lowest concentrations of coutaric acid, medium concentrations of caftaric,
302 and the highest concentrations of fertaric acid. This could be a characteristic that genetically
303 distinguish the pathway that produce hydroxycinnamates in Primitivo, in respect to the other
304 cultivars analysed in this study. As far as the stilbenoids, which concentration depends on the

cultivar or to possible plant stress as a fungal infection,[41] Montepulciano showed the highest concentrations for the glucosidic forms.

Figure 6 summarizes another important branch of the general pathway for the synthesis of polyphenols, where the metabolites are divided based on the number of their B-ring substitutes. This figure includes the families of flavanonols (dihydroquercetin, dihyrokaempferol and dihydromirycetin), flavonols (quercetin, isorhamnetin, kaempferol, syringetin, myricetin and laricitrin), anthocyanins (cyanidin, peonidin, delphinidin, malvidin and petunidin), and flavanols (catechin, *epi*catechin, gallocatechin, etc). The kaempferol pathway have only one substitute, quercetin two and myricetin three. It is known that the ratio between these three chemical groups are genetically controlled and often used to distinguish cultivars.[4,5] Teroldego was characterized by the highest concentration in the tri-substitute families, thus to the derivatives of myricetin, delphinidin, petunidin and malvidin. Moreover, Teroldego wines appeared to those with the highest amount of all anthocyanins. Sangiovese wines were the richest in quercetin, followed by Nebbiolo and Nerello. These data are in agreement with a previous study on grapes, where all grapes vines were cultivated in the same vineyard and under the same condition.[5] According to Mattivi et al.[5], myricetin had the highest % between all the flavonols for Teroldego (74%) and Sagrantino (82%), while quercetin the highest for Sangiovese (67%) and Nebbiolo (70%). The same study, that included all the cultivars of the present project except Nerello Mascalese, is in agreement with our outcome about the richness of Teroldego in anthocyanins. In recent years, Sangiovese wines suffer from a problem of instability involving quercetin (and other flavonols), generating flakes floating in the wine that appears in bottled wine.[42] The chemical analysis demonstrated that the major component of these flakes is quercetin aglycon, and so it is believed that is occurring under high amount of quercetin in the wines.[42] As far as our knowledge is concerning this problem was not reported so far in Nebbiolo or Nerello wines, which according to our results had the highest concentration of quercetin after Sangiovese.

Nebbiolo was also the group of wines with the highest amount of isorhamnetin, which is the methylation product of quercetin and di-substituted in the B-ring. Also this result was in agreement with Mattivi et al.[5], where isorhamnetin represented the 15% of all flavonols for Nebbiolo. After

Teroldego, Raboso was the second group of wines in terms of the cyanidin and peonidin amount. For the tri-substitute anthocyanins, after Teroldego, Montepulciano and Sagrantino were the richest cultivars, followed by Aglianico and Cannonau.

For the monomeric flavanols, Aglianico was the richest group for catechin and *epi*catechin, followed by Sagrantino and Teroldego for epicatechin, and Sagrantino, Nerello, Nebbiolo and Corvina for catechin. Teroldego was the richest group for epicatechin gallate, followed by Sagrantino and Sangiovese from Romagna. Nerello was the richest in gallocatechin and Teroldego the richest for epigallocatechin. Finally, Sagrantino was also the richest for epigallocatechin gallate (Figure 6). Flavanols is an important family of polyphenols in wine since, between others, it influences wine astringency and bitterness. According to Cheynier et al.[43] *epi*catechin is more bitter that catechin, and the galloylation increases the astringency. Several other monomeric polyphenols inserted having a large variability have also been described to be sensory active, affecting the quality of bitterness and astringency of the red wines.[44]

Wine is not just a grape product, but includes a complex technological process (alcoholic fermentation, malolactic fermentation, etc.) and each step enriches and modifies the wine metabolomic fingerprint. Additionally wine metabolites continuously evolve during aging. Anthocyanins, which are the metabolites responsible for the red colour of the wines (and many other food and flowers), participate to a number of reaction during wine aging leading to the production of several classes of wine pigments. As Figure 7 depicted, Teroldego was the richest group in grape anthocyanins, but Aglianico was richest in direct linked and ethyl-bridged linked flavanols-anthocyanins, probably because of its higher content in epicatechin. Particularly rich in ethyl-bridged flavanols-anthocyanins were also Sagrantino, Cannonau and Primitivo. After Aglianico, the richest group in directed linked flavanols-anthocyanins were Sagrantino, Teroldego, Cannonau and Sangiovese. Cannonau, which was the richest in caftaric acid (Figure 5) was also the richest group for some pinotins which are condensation products of hydroxycinnamates with the anthocyanins (Figure 7). Finally, the product of the reaction between malvidin 3-glucoside and acetaldehyde, B-type vitisin was to found to characterize more Cannonau, Raboso and Aglianico;

while the product between malvidin 3-glucoside and pyruvic acid characterised the groups of Montepulciano, Aglianico, Sagrantino and Teroldego (Figure 7).

One central objective of this project was to study tannins of Italian red wines originated from the grapes, so all the wines were prepared without any tannin addition or contact with wooden barrels. Figure 8 demonstrates a comparison of the wine groups for different monomeric, dimeric, trimeric and tetrameric flavanols, and also included some monomeric sulfonated flavanols. Moreover, the metabolites were divided in 4 families based on the B-ring substitution: a) procyanidins, only constituted by the di-substitute catechin and epicatechin; b) proanthocyanidins, which have at least one tri-substituted gallocatechin or epigallocatechin, and one di-substituted catechin or epicatechin; c) prodelphinidins, only constituted by the tri-substituted gallocatechin and epigallocatechin; and d) gallates, that include at least one galloyl moiety. According to previous researches, the polymerization of tannins decreases the bitterness, and dimers, trimers and tetramers areperceived as more bitter than astringent. As the polymerization increases, initially astringency increases (oligomeric tannins), but as the polymerization further increases astringency decreases (polymeric tannins).[43]

Aglianico group was the richest in procyanidins type tannins, followed by Sagrantino and Nebbiolo. These three cultivars are known to produce wines with astringent character. Conversely, Cannonau, Corvina, Montepulciano, Raboso and Nerello showed the smallest amounts of procyanidins. Sagrantino wines were also the richest in mixed proanthocyanidins, followed by Nerello and Nebbiolo; while Primitivo, Corvina and Teroldego were the poorest. As far as concerns the prodelphinidins Sagrantino, Sangiovese, Nerello, Nebbiolo and Teroldego were the richest; and Primitivo, Corvina and Cannonau were the poorest. Sagrantino, Aglianico, Teroldego and Nebbiolo were the richest in galloylated flavanols; while Primitivo, Corvina, Cannonau and Nerello contained the lowest amounts. Raboso, Nerello, Sangiovese from Tuscany and Montepulciano were the wines with the highest concentration on sulfonated tannins (Figure 8).

Generally, this analytical survey on the untargeted metabolomic fingerprint of 11 Italian mono-cultivar red wines, all together for the first time, highlighted the huge diversity in the composition of these Italian Origin Wines, and generated hypothesis that will need to be validated in the future with targeted approaches. Primitivo was the wine group with the most distinctive metabolome, characterized by the highest amount in several amino acids (tyrosine, phenylalanine, arginine, valine, leucine and isoleucine), and the lowest levels of proline. In agreement with these findings, Primitivo wines were also characterized by a large number of N-containing metabolites. One additional characteristic of Primitivo was the increased level of methylation of both hydroxycinnamates and flavonols. Finally, Primitivo wines were poor in anthocyanins and oligomeric flavanols.

Teroldego was also a group of wine with a distinctive metabolomic fingerprint, characterised by the highest amount of anthocyanins, in particular three-substituted anthocyanins at the B-ring. Increased B-ring substitution in Teroldego was also observed for flavonols.

Nebbiolo wines were poor in amino acids, hydroxycinnamates, anthocyanins and their derivatives; but rich in kaempferol, isorhametin and quercetin (the 2nd richest group in quercetin after Sangiovese). Condensed tannins were detected in high concentration in Nebbiolo wines, as well as procyanidin gallates and gallic acid. This high galloylation could perhaps explain the astringent character of renowned Nebbiolo wines (Barolo, Barbaresco, etc).[45,46]

Aglianico wines were the richest in catechin, epicatechin, procyanidins, type A vitisin, type B vitisin, and the products of reaction between anthocyanins and flavanols (both ethyl-linked and direct-linked). Aglianico samples did not exhibit particularly high levels of anthocyanins, possibly due to the high rate of reaction with flavanols, resulting in the synthesis of stable anthocyanins adducts and therefore more stable color. The high content of monomeric and oligomeric procyanidins could be also responsible for the high astringent character of Aglianico wines.[47,48]

Sangiovese, which is the most widespread Italian cultivar, was close to Nebbiolo and Nerello in ESI- analysis, whereas for ESI+ it showed a metabolite profile close to Nebbiolo and Raboso. If we take into consideration all the wine groups, Sangiovese wines were characterized by the B-ring di-

substituted flavonols (quercetin derivatives) and anthocyanins (cyanidin 3-glucoside), and the di-substituted hydroxycinnamates (coutaric acid). The tannins of Sangiovese were rich in proanthocyanidins/prodelphinidins with tri-substituted flavanols (gallocatechin and/or epigallocatechin units), while the Sangiovese wines from Tuscany were rich in sulfonated oligomeric flavanols. Finally, Sangiovese wines were poor in amino acids and N-containing metabolites. Overall, Sangiovese wines from Tuscany and Romagna were close and had a very similar metabolome.

Cannonau wines were characterised by various caffeic acid metabolites (caftaric acid, caffeoyl derivatives, sulfonated caftaric acid, and pinotins). They were also rich in B-type vitisin, arginine and B-ring methylated flavonoids (syringetin, laricitrin and malvidin derivatives), while relatively poor in tannins.

Sagrantino wines showed the highest content of tryptophan, and had intermediate amounts for the other amino acids. Oligomeric tannins were generally high in Sagrantino, both direct-linked and ethyl-linked flavanol-anthocyanins, and the highest levels in proanthocyanidins and epigallocatechin gallate were also detected. Sagrantino wines were also characterized by the highest amounts in flavanonols (dihydroxykaempferol and dihydroxyquercetin), and for relatively high levels of coutaric acid than caftaric and fertaric.

Corvina wines were the less homogenous group, with generally low levels in polyphenols (except flavanonols), and highest content of sulfonated glutathione and sulfonated indole lactic acid glucoside. Raboso were characterised by the di-substituted anthocyanins, cyanidin 3-glucoside and peonidin 3- glucoside, and the sulfonated tannins. Montepulciano group was characterised by the acetylated anthocyanins, indole lactic acid and its glucoside, and ellagic acid.

In conclusion, the use of a robust untargeted LC-MS based analytical protocol together with a targeted sampling protocol covering a large portion of Italian enological biodiversity produced an interesting publicly available database. For the 11 mono-cultivar red wines investigated, Primitivo, Teroldego and Nebbiolo had the highest number of pBOWs; and a second group comprised Sangiovese, Aglianico, Cannonau and Raboso. Primitivo and Teroldego had the most

distinguished metabolomic fingerprint, while Sangiovese with Nebbiolo and Montepulciano with Cannonau had very similar metabolomes. Between the pBOWs were annotated several N-containing metabolites (amino acids, di- and tri-peptides, etc), showing that could be promising metabolites to understand and exploit wine diversity. Especially Primitivo wines were very rich in N-containing metabolites tentative markers. The wines with the richest metabolome in condensed tannins were Sagrantino, Nebbiolo and Aglianico. Teroldego was characterised by the highest amount in anthocyanins, followed by Raboso, Montepulciano, Sagrantino and Aglianico. Sangiovese, Nebbiolo, Nerello and Raboso were characterised by di-substituted flavonoids in the B-ring; and Primitivo, Teroldego, Aglianico, Cannonau and Montepulciano by tri-substituted. In parallel, mono-substituted hydroxycinnamates characterised Sangiovese, Nerello and Raboso, and di- and tri-substituted characterised Primitivo and Cannonau. As expected, the pathway of polyphenols offers many tools in order to understand the metabolomic diversity of the wines. Moreover, even if all wines had the same total SO_2, this wine preservative reacted in a different manner with the metabolites of each wine. In Corvina, Montepulciano and Raboso reacted with ILA-glu; in Teroldego, Corvina, Raboso and Primitivo with glutathione; and Nerello, Sangiovese and Raboso with flavanols. Both raw and analysed data are publicly available, in order to help other researchers in their aim to understand better the Italian oenological diversity and quality.

Abbreviations Used:

AGL, Aglianico; PRI, Primitivo; TER, Teroldego; NER, Nerello Mascalese; RAB, Raboso; COR, Corvina; CAN, Cannonau; MON, Montepulciano; SAG, Sagrantino; SAT, Sangiovese Tuscany; SAR, Sangiovese Romagna; NEB, Nebbiolo; QC, Quality Control; pBOWs, putative Biomarkers of Origin Wines; LC, Liquid Chromatography; MS, Mass Spectrometry; FTICR, Fourier-transform ion cyclotron resonance; UPLC-QTOF MS, Ultra-high Performance Liquid Chromatography-Quadrupole Time-of-Flight Mass Spectrometry; PCA, Principal component analysis.

Acknowledgements:

Authors acknowledge the other members of the D:Wines consortium: A. Curioni[3], A. Gambuti[4], S. Giacosa[5], D. Perenzoni[1], Luigi Moio[4], A. Rinaldi[4], S. Río Segade[5], G.P. Parpinello[6], A. Ricci[6], B. Simonato[2], G. Tornielli[2], D. Slaghenaufi[2], S. Vincenzi[3]

Funding sources:

The work was funded by the MIUR project N. 20157RN44Y

Conflict of interest

The authors declare that they have no conflict of interest.

Author contributions

All authors conceived and designed the experiment, collected the samples, and read and approved the manuscript; PA performed the LC-MS based metabolomics analysis and data analysis, interpreted the data, and prepared the Tables and Figures. PA, FM, MU, MM and PP wrote the manuscript.

AUTHOR INFORMATION

Corresponding Author

*Email addresses panagiotis.arapitsas@fmach.it and panagiotis.arapitsas@gmail.com

ORCID

Panagiotis Arapitsas 0000-0002-9378-9287

Maurizio Ugliano 0000-0002-6487-2866

Matteo Marangon 0000-0001-5358-2428

Paola Piombino 0000-0001-8705-9846

Luca Rolle 0000-0002-6075-079X

Vincenzo Gerbi 0000-0001-7748-120X

Andrea Versari 0000-0003-2960-0619

Fulvio Mattivi 0000-0003-4935-5876

Supporting Information description

Supplementary Table S1. Wine meta-information and basic oenological analysis

Supplementary Table S2. Putative markers list for the analysis in ESI-; including information about the annotation, annotation level, statistical data and the group(s) of wines that were markers.

Supplementary Table S3. Putative markers list for the analysis in ESI+; including information about the annotation, annotation level, statistical data and the group(s) of wines that were markers.

Supplementary Figure S1. Heatmap of all annotated metabolites used for the Figures 3-8.

References

(1) OIV. Organisation of Vine and Wine (OIV) Focus for 2018. 2018.

(2) Lacombe, T.; Audeguin, L.; Boselli, M.; Bucchetti, B.; Cabello, F.; Chatelet, P.; Crespan, M.; D'Onofrio, C.; Eiras Dias, J.; Ercisli, S.; et al. Grapevine European Catalogue: Towards a Comprehensive List. *VITIS* **2011**, *50* (2), 65–68.

(3) Federodoc. *I Vini Italiani a Denominazione d'origine 2016*; 2016.

(4) Flamini, R.; Mattivi, F.; De Rosso, M.; Arapitsas, P.; Bavaresco, L. Advanced Knowledge of Three Important Classes of Grape Phenolics: Anthocyanins, Stilbenes and Flavonols. *Int. J. Mol. Sci.* **2013**, *14* (10), 19651–19669. https://doi.org/10.3390/ijms141019651.

(5) Mattivi, F.; Guzzon, R.; Vrhovsek, U.; Stefanini, M.; Velasco, R. Metabolite Profiling of Grape: Flavonols and Anthocyanins. *J. Agric. Food Chem.* **2006**, *54* (20), 7692–7702.

https://doi.org/10.1021/jf061538c.

(6) de Villiers, A.; Venter, P.; Pasch, H. Recent Advances and Trends in the Liquid-Chromatography–Mass Spectrometry Analysis of Flavonoids. *J. Chromatogr. A* **2016**, *1430*, 16–78. https://doi.org/10.1016/J.CHROMA.2015.11.077.

(7) Black, C. A.; Parker, M.; Siebert, T. E.; Capone, D. L.; Francis, I. L. Terpenoids and Their Role in Wine Flavour: Recent Advances. *Aust. J. Grape Wine Res.* **2015**, *21*, 582–600. https://doi.org/10.1111/ajgw.12186.

(8) Gika, H.; Theodoridis, G.; Mattivi, F.; Vrhovsek, U.; Pappa-Louisi, A. Hydrophilic Interaction Ultra Performance Liquid Chromatography Retention Prediction under Gradient Elution. *Anal. Bioanal. Chem.* **2012**, *404* (3), 701–709. https://doi.org/10.1007/s00216-012-6015-6.

(9) Cozzolino, D. Metabolomics in Grape and Wine: Definition, Current Status and Future Prospects. *Food Anal. Methods* **2016**, *9* (11), 2986–2997. https://doi.org/10.1007/s12161-016-0502-x.

(10) Diamantidou, D.; Zotou, A.; Theodoridis, G. Wine and Grape Marc Spirits Metabolomics. *Metabolomics* **2018**, *14* (12), 159. https://doi.org/10.1007/s11306-018-1458-1.

(11) Arapitsas, P.; Mattivi, F. LC-MS Untargeted Protocol for the Analysis of Wine; Humana Press, New York, NY, 2018; pp 225–235. https://doi.org/10.1007/978-1-4939-7643-0_16.

(12) Pinu, F.; Pinu; R, F. Grape and Wine Metabolomics to Develop New Insights Using Untargeted and Targeted Approaches. *Fermentation* **2018**, *4* (4), 92. https://doi.org/10.3390/fermentation4040092.

(13) Arapitsas, P.; Speri, G.; Angeli, A.; Perenzoni, D.; Mattivi, F. The Influence of Storage on the "Chemical Age" of Red Wines. *Metabolomics* **2014**, *10* (5), 816–832. https://doi.org/10.1007/s11306-014-0638-x.

(14) Arapitsas, P.; Ugliano, M.; Perenzoni, D.; Angeli, A.; Pangrazzi, P.; Mattivi, F. Wine Metabolomics Reveals New Sulfonated Products in Bottled White Wines, Promoted by Small Amounts of Oxygen. *J. Chromatogr. A* **2016**, *1429*, 155–165.

https://doi.org/10.1016/j.chroma.2015.12.010.

(15) Carlin, S.; Vrhovsek, U.; Franceschi, P.; Lotti, C.; Bontempo, L.; Camin, F.; Toubiana, D.; Zottele, F.; Toller, G.; Fait, A.; et al. Regional Features of Northern Italian Sparkling Wines, Identified Using Solid-Phase Micro Extraction and Comprehensive Two-Dimensional Gas Chromatography Coupled with Time-of-Flight Mass Spectrometry. *Food Chem.* **2016**, *208*, 68–80. https://doi.org/10.1016/j.foodchem.2016.03.112.

(16) Roullier-Gall, C.; Hemmler, D.; Gonsior, M.; Li, Y.; Nikolantonaki, M.; Aron, A.; Coelho, C.; Gougeon, R. D.; Schmitt-Kopplin, P. Sulfites and the Wine Metabolome. *Food Chem.* **2017**, *237*, 106–113. https://doi.org/10.1016/J.FOODCHEM.2017.05.039.

(17) De Rosso, M.; Mayr, C. M.; Girardi, G.; Dalla Vedova, A.; Flamini, R. High-Resolution Mass Spectrometry Metabolomics of Grape Chemical Markers to Reveal Use of Not-Allowed Varieties in the Production of Amarone and Recioto Wines. *Metabolomics* **123AD**, *1*, 124. https://doi.org/10.1007/s11306-018-1415-z.

(18) Arapitsas, P.; Corte, A. D. A. Della; Gika, H.; Narduzzi, L.; Mattivi, F.; Theodoridis, G. Studying the Effect of Storage Conditions on the Metabolite Content of Red Wine Using HILIC LC-MS Based Metabolomics. *Food Chem.* **2016**, *197*, 1331–1340. https://doi.org/10.1016/j.foodchem.2015.09.084.

(19) Roullier-Gall, C.; Witting, M.; Moritz, F.; Gil, R. B.; Goffette, D.; Valade, M.; Schmitt-Kopplin, P.; Gougeon, R. D. Natural Oxygenation of Champagne Wine during Ageing on Lees: A Metabolomics Picture of Hormesis. *Food Chem.* **2016**, *203*, 207–215. https://doi.org/10.1016/j.foodchem.2016.02.043.

(20) Jeandet, P.; Heinzmann, S. S.; Roullier-Gall, C.; Cilindre, C.; Aron, A.; Deville, M. A.; Moritz, F.; Karbowiak, T.; Demarville, D.; Brun, C.; et al. Chemical Messages in 170-Year-Old Champagne Bottles from the Baltic Sea: Revealing Tastes from the Past. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (19), 5893–5898. https://doi.org/10.1073/pnas.1500783112.

(21) Franceschi, P.; Mylonas, R.; Shahaf, N.; Scholz, M.; Arapitsas, P.; Masuero, D.; Weingart,

G.; Carlin, S.; Vrhovsek, U.; Mattivi, F.; et al. MetaDB a Data Processing Workflow in Untargeted MS-Based Metabolomics Experiments. *Front. Bioeng. Biotechnol.* **2014**, *2* (DEC), 72. https://doi.org/10.3389/fbioe.2014.00072.

(22) Theodoridis, G.; Gika, H.; Franceschi, P.; Caputi, L.; Arapitsas, P.; Scholz, M.; Masuero, D.; Wehrens, R.; Vrhovsek, U.; Mattivi, F. LC-MS Based Global Metabolite Profiling of Grapes: Solvent Extraction Protocol Optimisation. *Metabolomics* **2012**, *8* (2), 175–185. https://doi.org/10.1007/s11306-011-0298-z.

(23) Etièvant, P.; Schlich, P.; Cantagrel, R.; Bertrand, M.; Bouvier, J.-C. Varietal and Geographic Classification of French Red Wines in Terms of Major Acids. *J. Sci. Food Agric.* **1989**, *46* (4), 421–438. https://doi.org/10.1002/jsfa.2740460405.

(24) Parpinello, G. P.; Ricci, A.; Arapitsas, P.; Curioni, A.; Moio, L.; Riosegade, S.; Ugliano, M.; VERSARI, A. Multivariate Characterisation of Italian Monovarietal Red Wines Using MIR Spectroscopy. *OENO One* **2019**, *53* (4). https://doi.org/10.20870/oeno-one.2019.53.4.2558.

(25) Shahaf, N.; Franceschi, P.; Arapitsas, P.; Rogachev, I.; Vrhovsek, U.; Wehrens, R. Constructing a Mass Measurement Error Surface to Improve Automatic Annotations in Liquid Chromatography/Mass Spectrometry Based Metabolomics. *Rapid Commun. Mass Spectrom.* **2013**, *27* (21), 2425–2431. https://doi.org/10.1002/rcm.6705.

(26) Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W.-M.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; et al. Proposed Minimum Reporting Standards for Chemical Analysis. *Metabolomics* **2007**, *3* (3), 211–221. https://doi.org/10.1007/s11306-007-0082-2.

(27) Moro, L.; Da Ros, A.; da Mota, R. V.; Purgatto, E.; Mattivi, F.; Arapitsas, P. LC–MS Untargeted Approach Showed That Methyl Jasmonate Application on Vitis Labrusca L. Grapes Increases Phenolics at Subtropical Brazilian Regions. *Metabolomics* **2020**, *16* (2), 18. https://doi.org/10.1007/s11306-020-1641-z.

(28) Arapitsas, P.; Scholz, M.; Vrhovsek, U.; Di Blasi, S.; Biondi Bartolini, A.; Masuero, D.;

593 Perenzoni, D.; Rigo, A.; Mattivi, F.; Blasi, S. Di; et al. A Metabolomic Approach to the Study
594 of Wine Micro-Oxygenation. *PLoS One* **2012**, *7* (5), e37783.
595 https://doi.org/10.1371/journal.pone.0037783.

596 (29) Chong, J.; Soufan, O.; Li, C.; Caraus, I.; Li, S.; Bourque, G.; Wishart, D. S.; Xia, J.
597 MetaboAnalyst 4.0: Towards More Transparent and Integrative Metabolomics Analysis.
598 *Nucleic Acids Res.* **2018**, *46* (W1), W486–W494. https://doi.org/10.1093/nar/gky310.

599 (30) Haug, K.; Salek, R. M.; Conesa, P.; Hastings, J.; de Matos, P.; Rijnbeek, M.; Mahendraker,
600 T.; Williams, M.; Neumann, S.; Rocca-Serra, P.; et al. MetaboLights—an Open-Access
601 General-Purpose Repository for Metabolomics Studies and Associated Meta-Data. *Nucleic*
602 *Acids Res.* **2013**, *41* (D1), D781–D786. https://doi.org/10.1093/nar/gks1004.

603 (31) Ribéreau-Gayon, P.; Glories, Y.; Maujean, A.; Duburdieu, D. *Trattato Di Enologia 2*, 4th
604 edition.; Edagricole: MIlano, 2018.

605 (32) Karasinski, J.; Elguera, J. C. T.; Ibarra, A. A. G.; Wrobel, K.; Bulska, E.; Wrobel, K.
606 Comparative Evaluation of Red Wine from Various European Regions Using Mass
607 Spectrometry Tools. *Anal. Lett.* **2018**, *51* (16), 2645–2659.
608 https://doi.org/10.1080/00032719.2018.1442472.

609 (33) Bouloumpasi, E.; Soufleros, E.; Tsarchopoulos, C.; Biliaderis, C. Primary Amino Acid
610 Composition and Its Use in Discrimination of Greek Red Wines with Regard to Variety and
611 Cultivation Region. *Vitis - J. Grapevine Res.* **2002**, *41* (4), 195–202.

612 (34) Paolini, M.; Ziller, L.; Bertoldi, D.; Bontempo, L.; Larcher, R.; Nicolini, G.; Camin, F. δ^{15} N
613 from Soil to Wine in Bulk Samples and Proline. *J. Mass Spectrom.* **2016**, *51* (9), 668–674.
614 https://doi.org/10.1002/jms.3824.

615 (35) Sherman, E.; Coe, M.; Grose, C.; Martin, D.; Villas-Bôas, S. G.; Greenwood, D. R. Can
616 Wine Composition Predict Quality? A Mebolomics Approach to Assessing Pinot Noir Wine
617 Quality as Rated by Experts. In *Oenolvas2019*; 2019; p 107.

618 (36) Würdig, G.; Woller, R.; Breitbach, K. *Chemie Des Weines*; Ulmer: Stuttgart, 1989.

(37) Álvarez-Fernández, M. A.; Fernandez-Cruz, E.; García Parrilla, M. C.; Troncoso, A. M.; Mattivi, F.; Vrhovsek, U.; Arapitsas, P. Saccharomyces Cerevisiae and Torulaspora Delbrueckii Intra- and Extra-Cellular Aromatic Amino Acids Metabolism. *J. Agric. Food Chem.* **2019**, *67* (28), 7942–7953. https://doi.org/10.1021/acs.jafc.9b01844.

(38) Arapitsas, P.; Guella, G.; Mattivi, F. The Impact of SO2 on Wine Flavanols and Indoles in Relation to Wine Style and Age. *Sci. Rep.* **2018**, *8* (1), 858. https://doi.org/10.1038/s41598-018-19185-5.

(39) Ong, B. Y.; NAGEL, C. W. Hydroxycinnamic Acid-Tartaric Acid Ester Content in Mature Grapes and During the Maturation of White Riesling Grapes. *Am. J. Enol. Vitic.* **1978**, *29*, 277–281.

(40) Singleton, V. L.; Zaya, J.; Trousdale, E. K. Caftaric and Coutaric Acids in Fruit of Vitis. *Phytochemistry* **1986**, *25* (9), 2127–2133. https://doi.org/10.1016/0031-9422(86)80078-4.

(41) Gatto, P.; Vrhovsek, U.; Muth, J.; Segala, C.; Romualdi, C.; Fontana, P.; Pruefer, D.; Stefanini, M.; Moser, C.; Mattivi, F.; et al. Ripening and Genotype Control Stilbene Accumulation in Healthy Grapes. *J. Agric. Food Chem.* **2008**, *56* (24), 11773–11785. https://doi.org/10.1021/jf8017707.

(42) Mattivi, F.; Arapitsas, P.; Perenzoni, D.; Guella, G.; Speri, G.; Angeli, A.; Perenzoni, D.; Mattivi, F. The Influence of Storage on the “Chemical Age” of Red Wines. *ACS Symp. Ser.* **2014**, *1203*, 29–49. https://doi.org/10.1007/s11306-014-0638-x.

(43) Cheynier, V.; Prieur, C.; Guyot, S.; Rigaud, J.; Moutounet, M. The Structures of Tannins in Grapes and Wines and Their Interactions with Proteins; 1997; pp 81–93. https://doi.org/10.1021/bk-1997-0661.ch008.

(44) Hufnagel, J. C.; Hofmann, T. Orosensory-Directed Identification of Astringent Mouthfeel and Bitter-Tasting Compounds in Red Wine. *J. Agric. Food Chem.* **2008**, *56* (4), 1376–1386. https://doi.org/10.1021/jf073031n.

(45) Ferrandino, A.; Carra, A.; Rolle, L.; Schneider, A.; Schubert, A. Profiling of

Hydroxycinnamoyl Tartrates and Acylated Anthocyanins in the Skin of 34 Vitis Vinifera Genotypes. *J. Agric. Food Chem.* **2012**, *60* (19), 4931–4945. https://doi.org/10.1021/jf2045608.

(46) Petrozziello, M.; Torchio, F.; Piano, F.; Giacosa, S.; Ugliano, M.; Bosso, A.; Rolle, L. Impact of Increasing Levels of Oxygen Consumption on the Evolution of Color, Phenolic, and Volatile Compounds of Nebbiolo Wines. *Front. Chem.* **2018**, *6*, 137. https://doi.org/10.3389/fchem.2018.00137.

(47) Gambuti, A.; Rinaldi, A.; Pessina, R.; Moio, L. Evaluation of Aglianico Grape Skin and Seed Polyphenol Astringency by SDS–PAGE Electrophoresis of Salivary Proteins after the Binding Reaction. *Food Chem.* **2006**, *97* (4), 614–620. https://doi.org/10.1016/J.FOODCHEM.2005.05.038.

(48) Rinaldi, A.; Jourdes, M.; Teissedre, P. L.; Moio, L. A Preliminary Characterization of Aglianico (Vitis Vinifera L. Cv.) Grape Proanthocyanidins and Evaluation of Their Reactivity towards Salivary Proteins. *Food Chem.* **2014**, *164*, 142–149. https://doi.org/10.1016/J.FOODCHEM.2014.05.050.

FIGURE CAPTIONS

Figure 1. Distribution of the wine sample set according to their cultivar (black) and region (red). The principal denomination of origin of each cultivar/region are also presented (light blue). The cultivation area refers to all Italy for each cultivar for the year 2015.[1]

Figure 2. PCA plots of all the wines in ESI+ (above) and ESI- (bellow). AGL, Aglianico; PRI, Primitivo; TER, Teroldego; NER, Nerello Mascalese; RAB, Raboso; COR, Corvina; CAN, Cannonau; MON, Montepulciano; SAG, Sagrantino; SAT, Sangiovese Tuscany; SAR, Sangiovese Romagna; NEB, Nebbiolo; QC, Quality Control.

Figure 3. Number of pBOW features for each cultivar in ESI+ and ESI-. Unique are the pBOWs that helps to discriminate the cultivar for all the others.

Figure 4. Clustering of the wines according to the markers in ESI+ and ESI-.

Figure 5. Biosynthesis and synthesis of N-containing metabolites, hydroxycinnamates and stilbenoids annotated in this study. Colours refers to the heat-map of Supplementary Figure S1 and represent a comparison of the concentration of each metabolite between the various mono-cultivar wine groups. The heatmap was build using Pareto scaling and Euclidean distance. AGL, Aglianico; PRI, Primitivo; TER, Teroldego; NER, Nerello Mascalese; RAB, Raboso; COR, Corvina; CAN, Cannonau; MON, Montepulciano; SAG, Sagrantino; SAT, Sangiovese Tuscany; SAR, Sangiovese Romagna; NEB, Nebbiolo.

Figure 6. General pattern for flavonoids biosynthesis, with the metabolites annotated in this study. Colours refers to the heat-map of Supplementary Figure S1 and represent a comparison of the concentration of each metabolite between the various mono-cultivar wine groups. The heatmap was build using Pareto scaling and Euclidean distance. AGL, Aglianico; PRI, Primitivo; TER, Teroldego; NER, Nerello Mascalese; RAB, Raboso; COR, Corvina; CAN, Cannonau; MON, Montepulciano; SAG, Sagrantino; SAT, Sangiovese Tuscany; SAR, Sangiovese Romagna; NEB, Nebbiolo.

688 **Figure 7.** Generic diagram with the major reaction that anthocyanins take part in wine. Colours
689 refers to the heatmap of Supplementary Figure S1 and represent a comparison of the
690 concentration of each metabolite between the various mono-cultivar wine groups. The heatmap
691 was build using Pareto scaling and Euclidean distance. AGL, Aglianico; PRI, Primitivo; TER,
692 Teroldego; NER, Nerello Mascalese; RAB, Raboso; COR, Corvina; CAN, Cannonau; MON,
693 Montepulciano; SAG, Sagrantino; SAT, Sangiovese Tuscany; SAR, Sangiovese Romagna; NEB,
694 Nebbiolo.

695 **Figure 8.** Variation of the annotated monomeric and oligomeric flavanols according to the various
696 mono-cultivar wine groups. The separation is based on the B-ring substitution. Colours refers to
697 the heat-map of Supplementary Figure S1 and represent a comparison of the average
698 concentration of each metabolite within each of the various mono-cultivar wine groups. The
699 heatmap was build using Pareto scaling and Euclidean distance. AGL, Aglianico; PRI, Primitivo;
700 TER, Teroldego; NER, Nerello Mascalese; RAB, Raboso; COR, Corvina; CAN, Cannonau; MON,
701 Montepulciano; SAG, Sagrantino; SAT, Sangiovese Tuscany; SAR, Sangiovese Romagna; NEB,
702 Nebbiolo. [a]Two di-substituted and one tri-substituted block; [b]One di-substituted and two tri-
703 substituted block.

FIGURES

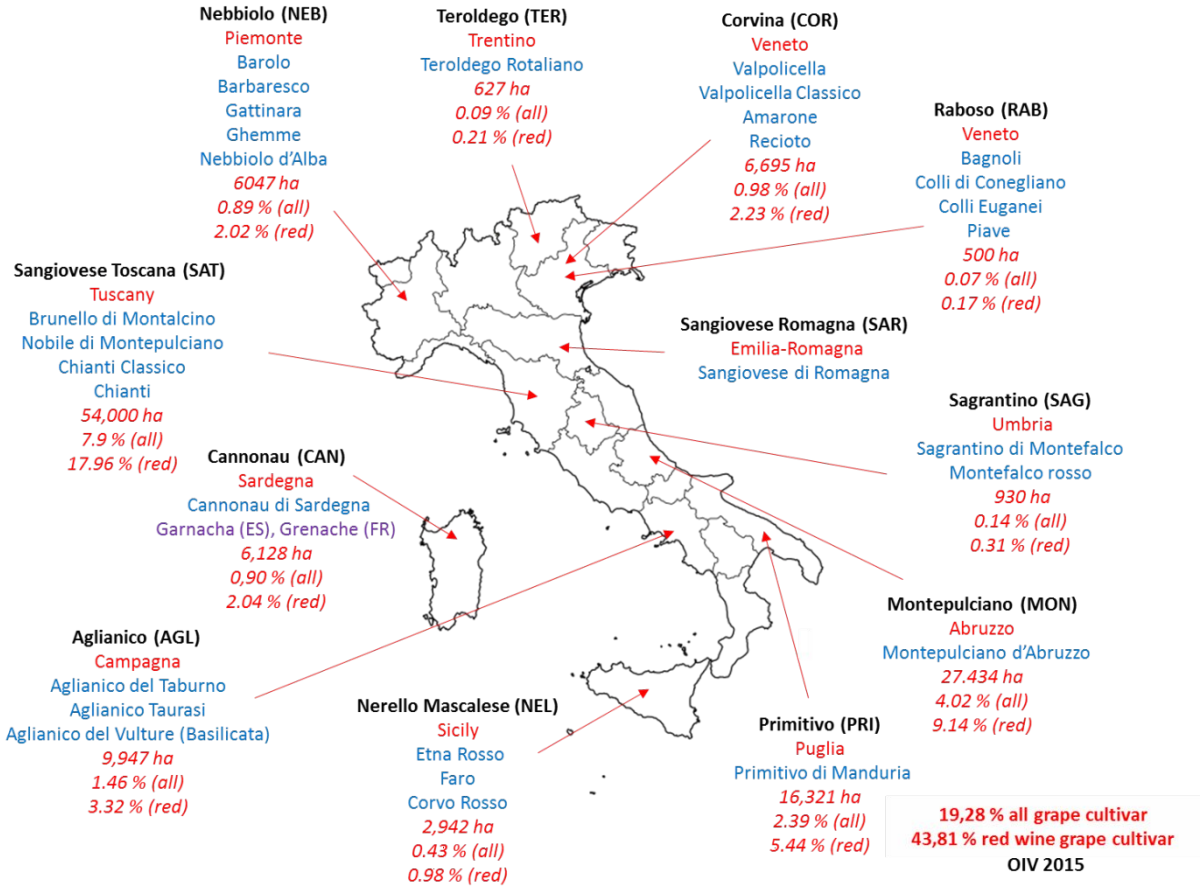


Figure 1

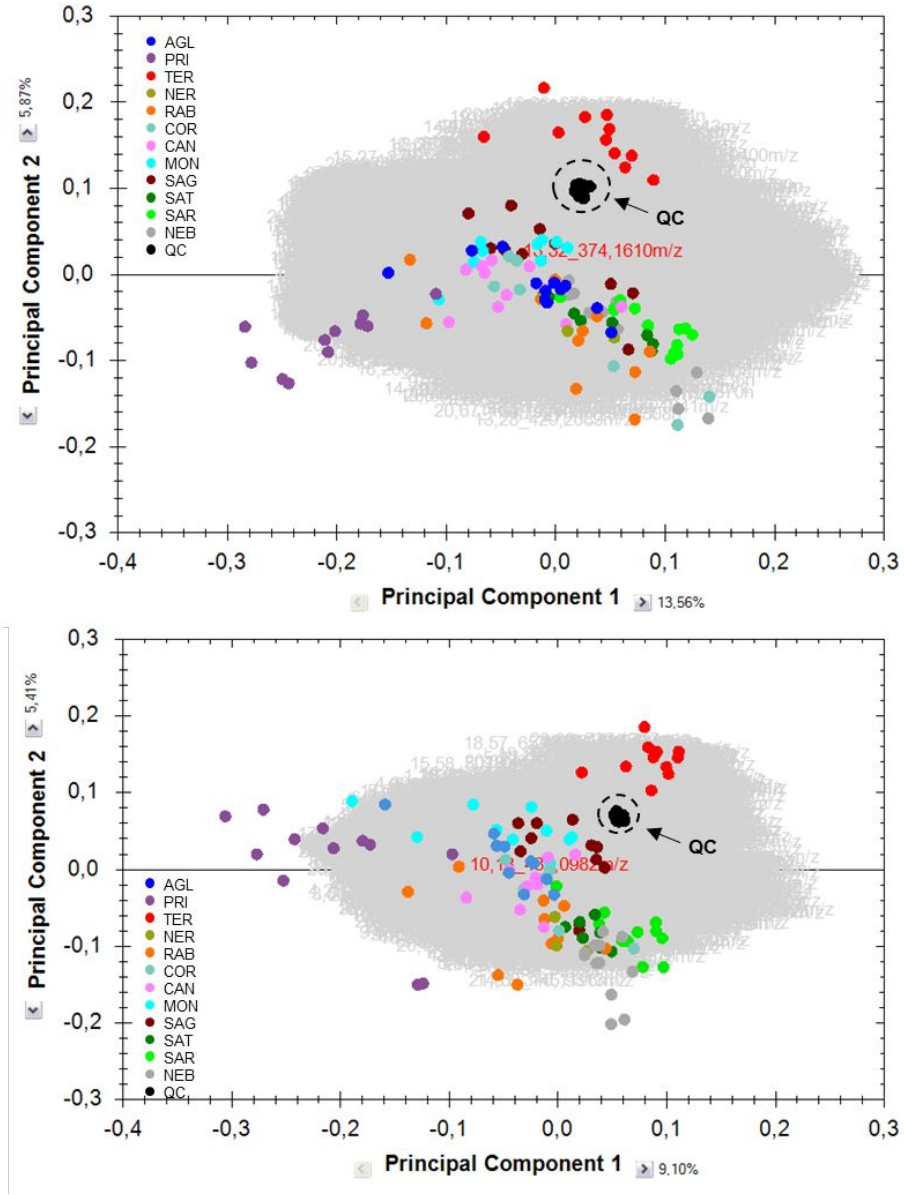
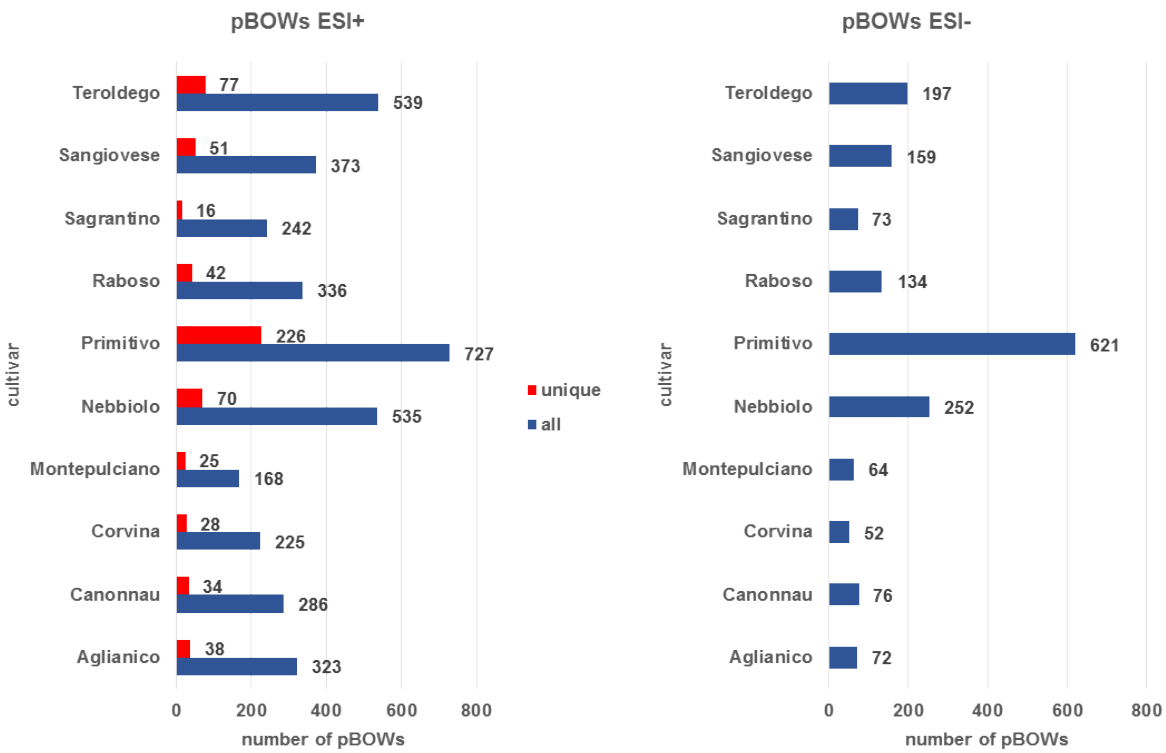

Principal Component 2
5,87%
Principal Component 1
13,56%
AGL
PRI
TER
NER
RAB
COR
CAN
MON
SAG
SAT
SAR
NEB
QC
QC
Principal Component 2
5,41%
Principal Component 1
9,10%

Figure 2



710

711 Figure 3

712

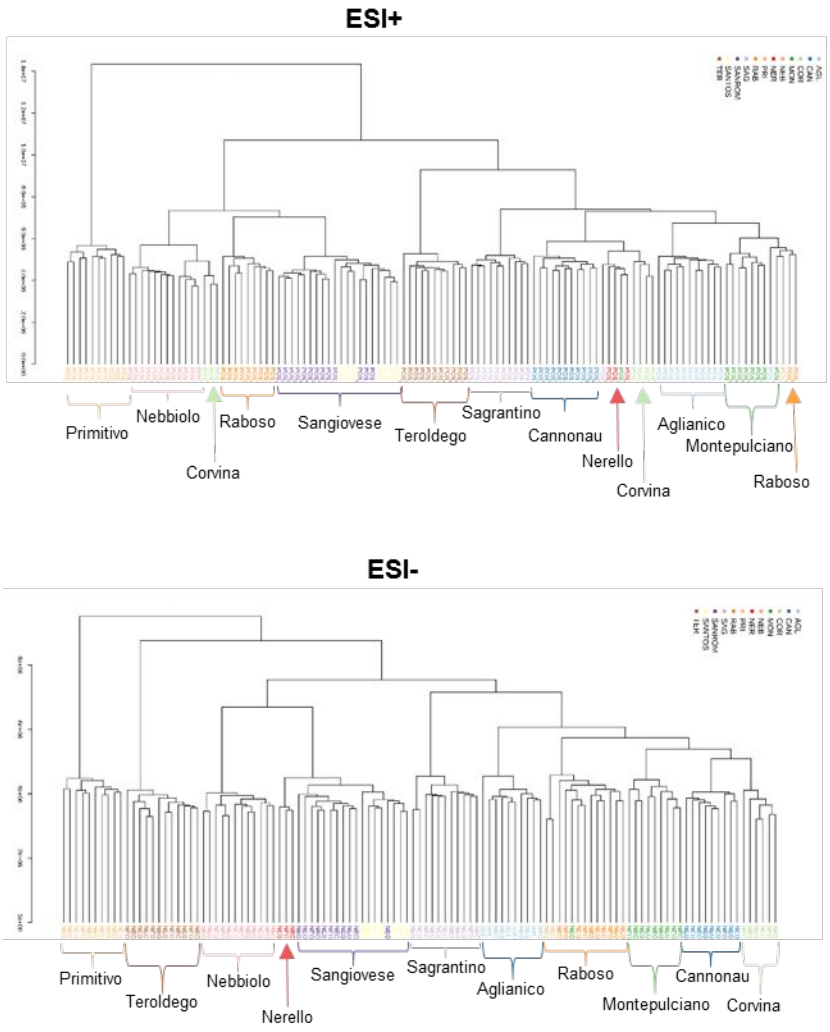
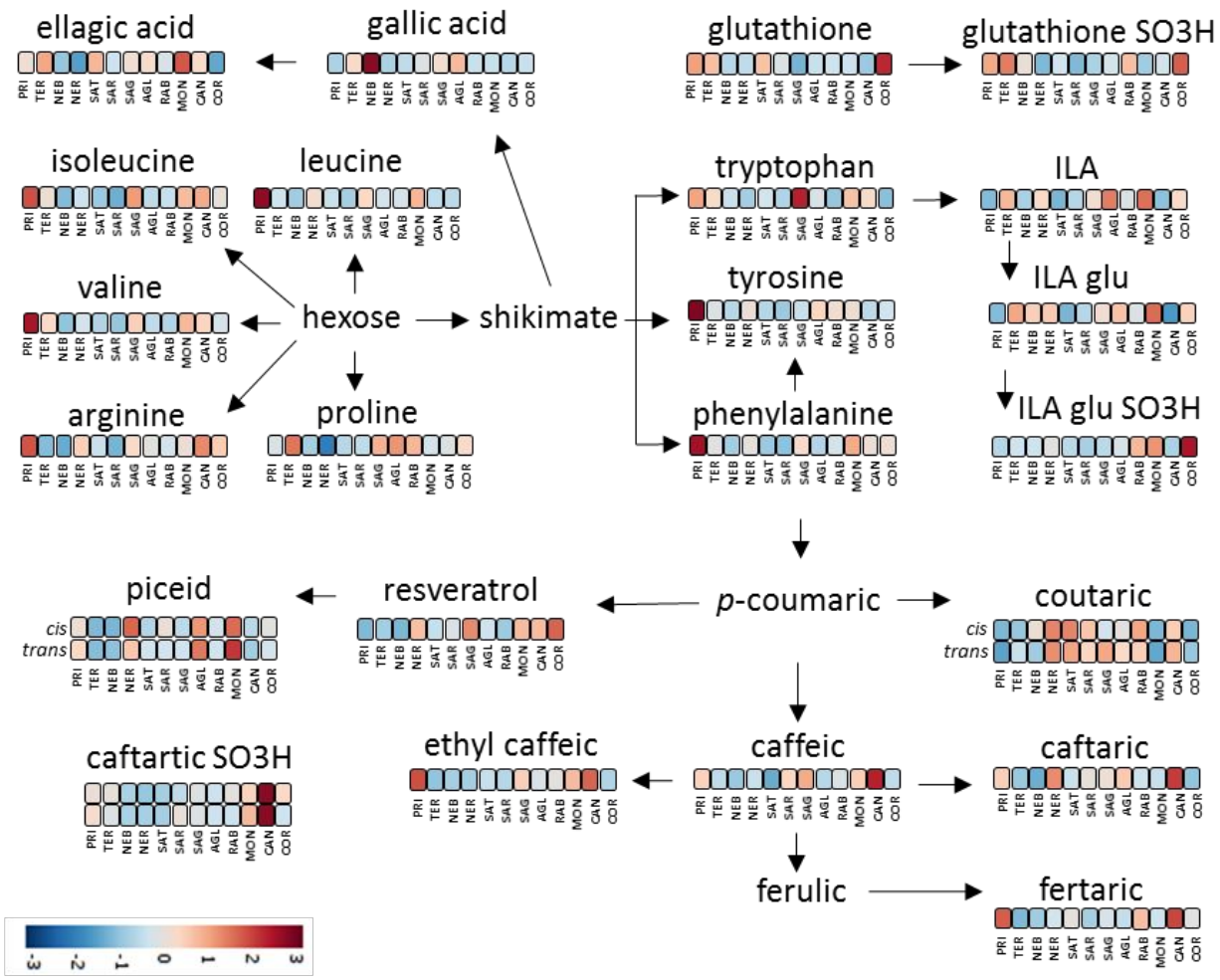

ESI+
Primitivo
Nebbiolo
Raboso
Corvina
Sangiovese
Teroldego
Sagrantino
Cannonau
Nerello
Corvina
Aglianico
Montepulciano
Raboso
ESI-
Primitivo
Teroldego
Nebbiolo
Nerello
Sangiovese
Sagrantino
Aglianico
Raboso
Montepulciano
Cannonau
Corvina

Figure 4



717

718

719 Figure 5

720

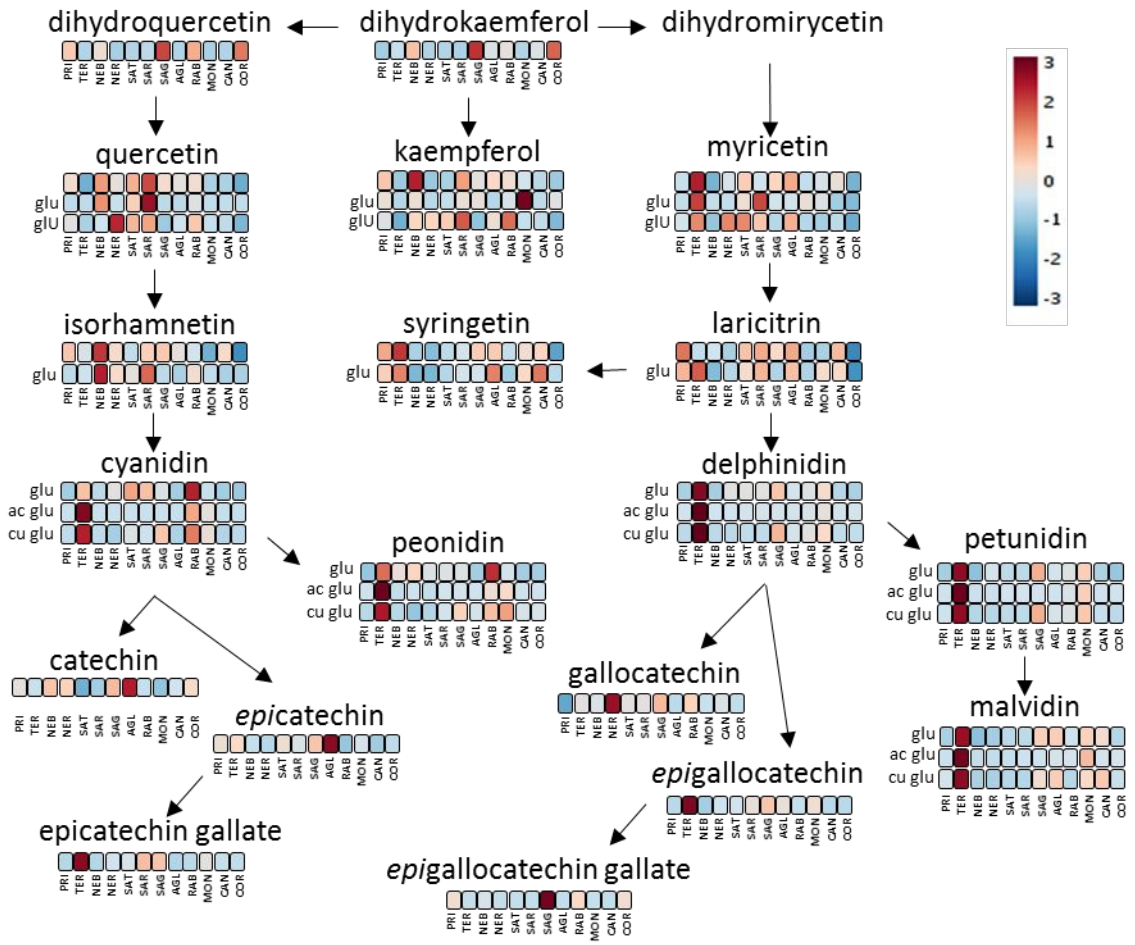

dihydroquercetin
dihydrokaempferol
dihydromyricetin
quercetin
kaempferol
myricetin
isorhamnetin
syringetin
laricitrin
cyanidin
delphinidin
peonidin
petunidin
catechin
epicatechin
gallocatechin
malvidin
epicatechin gallate
epigallocatechin
epigallocatechin gallate
glu
ac glu
cu glu
PRI TER NEB NER SAT SAR SAG AGL RAB MON CAN COR
3
2
1
0
-1
-2
-3

Figure 6

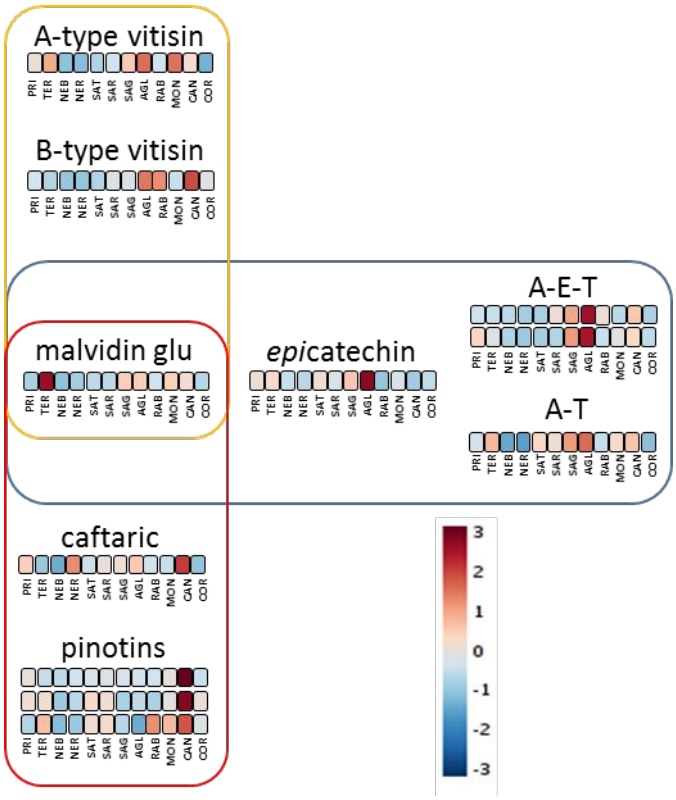
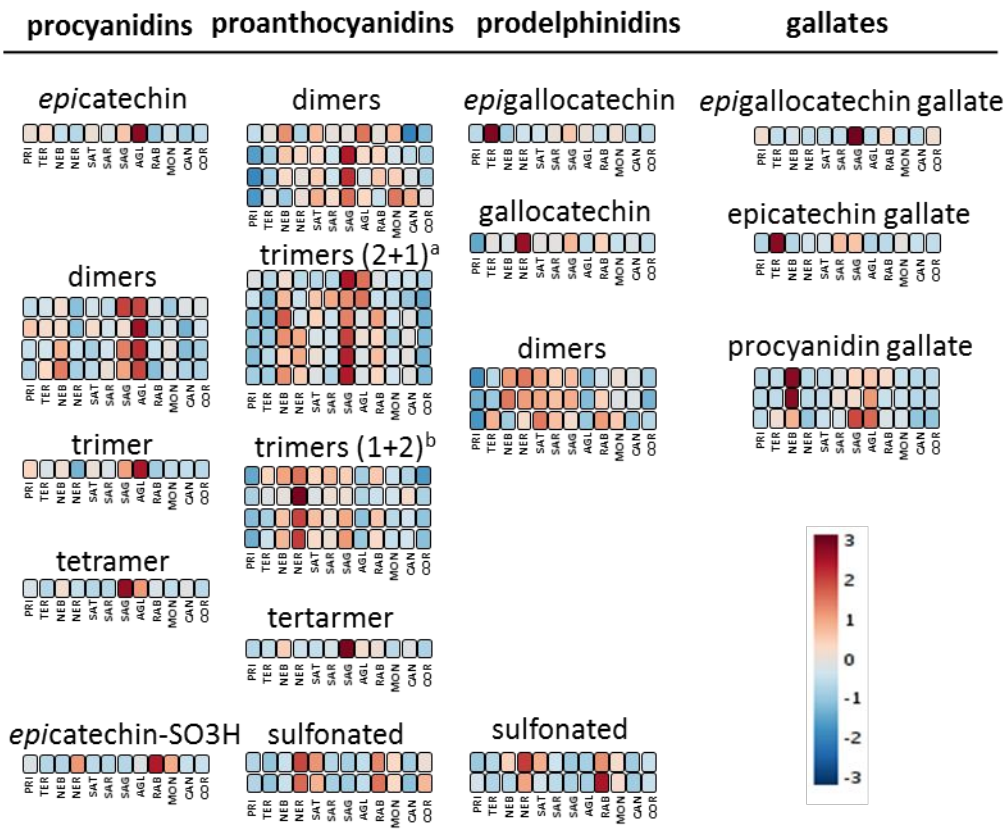

A-type vitisin
B-type vitisin
malvidin glu
epicatechin
A-E-T
A-T
caftaric
pinotins
PRI TER NEB NER SAT SAR SAG AGL RAB MON CAN COR
3
2
1
0
-1
-2
-3

Figure 7

728



729

730 Figure 8