UNIVERSITÀ DEGLI STUDI DI PADOVA

Università degli Studi di Padova

Padua Research Archive - Institutional Repository

Use of Untargeted Liquid Chromatography–Mass Spectrometry Metabolome To Discriminate Italian Monovarietal Red Wines, Produced in Their Different Terroirs

(Article begins on next page)

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

SCHOLARONE™
Manuscripts

1 **Use of untargeted LC-MS metabolome to discriminate Italian mono-varietal red**
2 **wines, produced in their different terroirs**
3
4

5 Panagiotis Arapitsas[1*], Maurizio Ugliano[2], Matteo Marangon[3], Paola Piombino[4], Luca Rolle[5],

6 Vincenzo Gerbi[5], Andrea Versari[6], Fulvio Mattivi[1,7].

7

8 [1]*Department of Food Quality and Nutrition, Research and Innovation Centre, Fondazione Edmund*

9 *Mach (FEM), Via E. Mach 1, 38010 San Michele all´Adige, (Italy)*

10 [2]*Department of Biotechnology, University of Verona, (Italy)*

11 [3]*Department of Agronomy, Food, Natural Resources, Animals and Environment (DAFNAE),*

12 *University of Padova, (Italy)*

13 [4]*Department of Agricultural Sciences, Division of Vine and Wine Sciences, University of Naples*

14 *Federico II, Viale Italia, 83100 Avellino, (Italy)*

15 [5]*Dipartimento di Scienze Agrarie, Forestali e Alimentari, Università degli Studi di Torino, Largo Paolo*

16 *Braccini 2, 10095 Grugliasco, TO, (Italy)*

17 [6]*Department of Agricultural and Food Sciences, University of Bologna, Piazza Goidanich 60,*

18 *Cesena, FC 47521, (Italy)*

19 [7]*Department of Cellular, Computational and Integrative Biology - CIBIO, University of Trento, 38123*

20 *Povo, (Italy)*

21   **Abstract**

22   Aim of this project was to register for the first time in a LC-MS based untargeted single-batch analysis

23   the metabolome of 11 single-cultivar Italian red wines (Aglianico, Cannonau, Corvina,

24   Montepulciano, Nebbiolo, Nerello, Primitivo, Raboso, Sagrantino, Sangiovese and Teroldego) from

25   12 regions across Italy, each one produced in their terroirs under ad hoc legal frameworks to

26   guarantee the quality and origin. The data provided indications about the similarity between the

27   cultivars, and highlighted a rich list of putative Biomarkers of Origin Wines (pBOWs) characterizing

28   each single combination cultivar-terroir, where Primitivo, Teroldego and Nebbiolo had the maximum

29   number of unique pBOWs. The pBOWs included anthocyanins (Teroldego), flavanols (Aglianico,

30   Sangiovese, Nerello and Nebbiolo), amino acids and N-containing metabolites (Primitivo),

31   hydroxycinnamates (Cannonau) and flavonols (Sangiovese). The raw data generated in this study

32   are publicly available, enabling the accessibility and reusability, and serve as a baseline dataset for

33   future investigations.

34

35   **Keywords**: mass spectrometry, wine authenticity; bioinformatics; wine metabolomics; amines;

36   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).

63 Wine, being the final product of a long and multistep process, has one of the richest and more

64 complex metabolomic fingerprint. Several targeted protocols focused on the analysis of polyphenols,

65 volatiles, lipids and etc. have been applied in order to find differences between wines coming from

66 different grape cultivars, as well as understanding the chemical and sensorial character of mono-

67 cultivar wines.[4–8] Over the last years, untargeted analytical approaches proved a valuable and

68 powerful alternative for the study of wine metabolome.[9–12] Techniques such as LC-MS, GC-MS or

69 direct injection FTICR-MS based metabolomics allowed identification of new wine metabolites,[13,14]

70 discrimination of groups of wines,[14–17] elucidation of chemical reaction occurring during aging and

71 storage[13,14,16,18,19] also in relationship to packaging[14], providing novel insights in wine history[20] and

72 quality.[13,14,16,21,22] Some wines of the above mentioned Italian cultivars have been subject of

73 untargeted LC-MS based analysis, alone or as groups together with other 2-3 cultivars, but the

74 literature lacks of studies that combine a large part of the red Italian wines diversity. Historically, the

75 most promising markers for the chemical characterization of varietal wines have been discovered

76 trying to compare the presence of a few targeted metabolites in varietal wines. As an example, a

77 pioneering study[23] based on the analysis of the variance of 20 organic acids and esters in six red

78 wines, led to the discovery that shikimic acid was associated with the cultivar, and in particular useful

79 to discriminate the Pinot noir wines. It is expected that the application of an untargeted method,

80 capable to produce a semi-quantitative analysis of ca. 1000 metabolites, has the potential to support

81 the discovery of several putative Biomarkers of Origin Wines (pBOWs).

82 Initially the aim of this project was to register for the first time the LC-MS metabolomic fingerprint of

83 11 mono-cultivar Italian red wines from 12 regions that representing a large portion of the Italian red

84 wine production and biodiversity. Supplementary aim was to investigate the produced dataset in

85 order to provide information about the metabolomic space similarity and dissimilarity between the

86 studied wines, and extract pBOWs. Additional scope was to make the dataset public available with

87 the intention to help other researchers.

88

89 **Materials and Methods**

90   Wine samples

91   A total of 110 Italian red wines, 100% mono-varietal, all vinified in 2016 from 11 diverse Italian

92   grape varieties harvested in the corresponding main geographical areas of production (12 wine

93   regions), were sampled directly from the producers. The wine sample set included: 11 Teroldego

94   (TER) from Trentino-Alto Adige; 7 Corvina (COR) from Veneto; 10 Raboso Piave (RAB) from

95   Veneto); 11 Nebbiolo (NEB) from Piedmont; 7 Sangiovese (SAT) from Tuscany; 12 Sangiovese

96   (SAR) from Romagna; 10 Sagrantino (SAG) from Umbria; 9 Montepulciano (MON) from Abruzzo; 9

97   Cannonau (CAN) from Sardinia; 10 Aglianico (AGL) from Campania; 11 Primitivo (PRI) from

98   Puglia; and 3 Nerello Mascalese (NER) from Sicily. The basic oenological information about the

99   wine are in Supplementary material Table S1 and Figure S1. The mid-infrared spectroscopy data

100   can be found in Parpinello et al.[24] Winemaking was carried out by each winery independently and

101   according to their standard production practices. However, for each wine the following

102   specifications were followed: a) wines had to be obtained from one single grape variety; b) wines

103   should be fermented in stainless steel vats; c) fermentation should be run in industrial scale; d) the

104   sampling should be preferentially made before malolactic fermentation; e) wines should not have

105   any contact with oak; f) 50 mg/L of free SO_2 had to be added at the time of sampling, before

106   bottling in dark glass bottles; g) Nomacorc Select Bio 500 (Nomacorc, France) closures had to be

107   used. The sampling occurred in early 2017 and the wines were stored at 4 ^{o}C until analysis. All

108   analysis were completed within a single batch, in 3 months after the sampling.

109   UPLC-QTOF MS analysis

110   Sample preparation followed a previously described protocol[11] and all steps until the LC/MS vial

111   filling occurred under nitrogen atmosphere. Wines were uncorked and an aliquot was transferred

112   into a 15 mL amber vial (filled to its capacity). Then a quality control (QC) pooled sample was

113   prepared by pooling 1 mL of each wine. Then 1 mL of each wine sample/QC was diluted with 2 mL

114   Milli-Q sonicated water and was finally filtered with 0.2 µm PTFE filters into a 2 mL amber vial (MS

115   certificated) prior to LC-MS analysis.

116   The analysis followed a previously described protocol.[11,13] A Waters Acquity UPLC coupled via an

117   electrospray ionization (ESI) interface to a Synapt HDMS QTOF MS (Waters, Manchester, UK)

118   operating in W-mode and controlled by MassLynx 4.1 was used. The column was a reversed

119   phase (RP) ACQUITY UPLC 1.8 μm 2.1 x 150 mm HSS T3 column (Waters); column manager

120   was set at 40 °C; the mobile phase flow rate was 0.28 mL/min; and the eluents was water and

121   methanol both with 0.1% formic acid. The multistep linear gradient used was as follows: 0-1 min,

122   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

123   % A isocratic; 25.5-25.6 min, 0-100 % A; 25.6-28 min 100% isocratic. Injection volume was 5 μL

124   and the samples were kept at 4 °C throughout the analysis. Mass spectrometry data were collected

125   by separate runs in positive and negative ESI mode over a mass range of 50 to 2000 amu with

126   scan duration of 0.4 s in centroid mode. The transfer collision energy and trap collision energy

127   were set at 6 V and 4 V. The source parameters were set as follows: capillary 3 kV for positive

128   scan and 2.5 kV for negative scan, sampling cone 25 V, extraction cone 3V, source temperature

129   150 °C, desolvation temperature 500 °C, desolvation gas flow 1000 L/h and nebulizer gas 50 L/h.

130   External calibration of the instrument was performed at the beginning of each batch of analysis by

131   direct infusion of a sodium formate solution (10 % formic acid/0.1 M NaOH/Acetonitrile at a ratio of

132   1/1/8), controlling the mass accuracy from 40 to 2000 m/z (less than 5 ppm) and mass resolution

133   (over 14000 FWHM). LockMass calibration was applied using a solution of leucine enkephaline

134   (0.5 mg/L, m/z 556.2771 for positive and 554.2620 for negative ion mode) at 0.1 mL/min. The QC

135   samples were used for the LC-MS system initial equilibration (4-5 injections) and control at regular

136   intervals (one QC sample injection every 6 real sample injections) during the sequence, according

137   to the quality control flowchart.[11] In total, the public available database included 26 QC sample

138   analysis for the ESI- and 24 analysis for the ESI+ mode (the system equilibration QC injections

139   were excluded).

140   Data analysis

141   For quality control during the runs and data analysis, we used PCA (Principal Component Analysis)

142   plots generated by Progenesis QI (Version 2.0.0.0.0, nonlinear Dynamics), checking the

143   distribution/clustering of the QC injections.[11] Progenesis QI parameters used for alignment were

144   done on default mode by Progenesis QI with peak picking performed at maximum level; and the

145   first minute and the last six minutes of the run excluded from data processing (only the range 1-22

146   min were used). Putative BOWs were considered the "compounds" that according to the

147   Progenesis QI statistical analysis had max fold range ≥ 2 and Anova (p value ≤ 0.01). Progenesis

148   QI views as "compound" a group of isotopic and adducts features belonging to the same

149   metabolite.

150   Annotation was performed manually by comparing retention times and mass spectra accuracy with

151   a mass tolerance of 5 ppm, based on the group's previous experience with the specific

152   instrumentation mass resolution,[25] and in accordance with the 4 levels described by Sumner et

153   al.[26].

154   Known wine metabolites previously annotated by using the same protocol[11,13,14,27,28] were

155   integrated semi-manually using the TargetLynx tools of Waters MassLynx 4.1 software (Milford,

156   MA). The TargetLynx parameters were set at chromatogram mass window 0.08 Dalton; retention

157   time window ± 0.2 min; smoothing iterations 1; and smoothing width 2. Further statistical analysis

158   was performed on these integrated peaks by using MetaboAnalyst online platform version 4.0

159   (http://www.metaboanalyst.ca/,[29]) without normalization, missing values estimation and data

160   transformation, by using Pareto scaling. For the Heatmap the Euclidean distance and Ward

161   clustering algorithm were used.

162   Raw LC–MS data and other details will be made publicly available for download with the accession

163   number MTBLS1443 from the MetaboLights public repository

164   (http://www.ebi.ac.uk/metabolights/,[21,30]).

165

166   **Result and Discussion**

167   A central starting point of this study was to obtain a set of wine samples that was as representative

168   as possible of the diversity of Italian red wine production both in terms of relevant varieties and

169   areas of origin. As shown in Figure 1, the samples included regions of northern (Piedmont,

170   Trentino and Veneto), central (Tuscany, Emilia-Romagna and Umbria), and southern (Campania

7

171 and Abruzzo) Italy, also with the two major islands (Sicily and Sardinia). In the case of Sangiovese,

172 the most important red variety in Italy, two different production areas, namely Tuscany and Emilia

173 Romagna, were considered. Wines were obtained from different wineries located in the production

174 area, so that they could be considered true representations not only of the varietal characteristics

175 but also of the winemaking practices commonly adopted in each area at winery level, and in

176 agreement with the rules of the specific denomination of origin. In order to avoid possible

177 differences deriving from aging and storage modalities, all samples were collected directly from the

178 tank, without any previous contact with wood, and were bottled in the laboratory under the same

179 conditions.

180 The applied LC-MS protocol proved several times in the past years its capability to register wine

181 metabolome and generated various new hypothesis.[11,13,14,28] As stated by this protocol, one of the

182 most crucial issues in untargeted LC-MS analysis is to inject all samples in a single batch. Due to

183 this methodological constraint, in this project it was decided to analysed only the wines produced in

184 one harvest. The number of biological replicates, i.e. different wines produced from different

185 vineyards and/or different wineries, was in the range 7-12 (mean 9.7) for all the wine regions, with

186 the sole exception of Nerello Mascalese from Sicily, for which only three suitable batches of wines

187 were obtained.

188 According to the workflow, followed in our laboratory, before any further data analysis it is

189 important to verify the quality of the dataset. Figure 2 shows the PCA plots of the sample injections

190 distribution according to a multivariate and unsupervised principal component analysis. The PCA

191 plot of the ESI+ analysis was performed using 11274 features and the ESI- 7397 features, and in

192 both cases, the QC sample injections – injected all over the sequence - formed a tide cluster,

193 proving the reliability of the measure, in term of absence of fluctuations for samples injected at

194 different time points. According to this unsupervised analysis, it was possible to notice that

195 Teroldego and Primitivo wine groups had a metabolomic fingerprint very different in respect to the

196 other wines.

197   In order to investigate the metabolites that differentiated each wine group from the other we used

198   supervised data analysis tools. By using the Anova tool of Progenesis QI, the metabolomic

199   fingerprint of each wine group was compared against all the other groups, so a subgroup of

200   features was created by using only the features with p-value ≤ 0.01 and fold change ≥ 2. The

201   different lists were merged and created the Supplementary Tables S2-3. The ESI- analysis

202   included 621 pBOWs and the ESI+ 1735 pBOWs. Figure 3 demonstrated the major outcome of

203   this data analysis. For the ESI+ analysis it was possible to detect also pBOWs unique for each

204   group of wine, while for ESI- that was not possible since Primitivo included all the pBOWs and did

205   not have any unique. In fact, both ESI- and ESI+ shown that Primitivo had the highest number of

206   pBOWs. This result was also in accordance both with the PCA plots (Figure 2) where Primitivo

207   samples are separated from the other cultivars by PC1; and the hierarchical cluster analysis

208   (Figure 4), where Primitivo sample is the first group of samples to slit from the others. In detail,

209   Primitivo has 727 features pBOWs (226 of them unique) for ESI+ and 621 for ESI-. Teroldego and

210   Nebbiolo also had a big number of pBOWs, and on the other hand, Montepulciano and Corvina

211   had the smallest number of pBOWs.

212   The hierarchical cluster analysis (Figure 4) showed that the Primitivo group was the one differing

213   the most for both the ESI- and ESI+ analysis. A second cluster in ESI+ included Nebbiolo, Corvina,

214   Raboso and Sangiovese wines. Such behaviour should be attributes to the fact that these cultivars

215   are known for their not very intense red colour[5] and because in ESI+ mode the positive charged

216   anthocyanins give very good signal. Therefore, the here observed clustering was most probably

217   strongly driven by the red coloured and positive charged anthocyanins. The result that Teroldego,

218   a very rich cultivar in anthocyanins,[5] formed a cluster alone, supports this hypothesis. These

219   findings indicated to us that we should investigate the anthocyanins and related pigments in detail.

220   In ESI- Teroldego was the second more distant cluster, while Nebbiolo, Nerello and Sangiovese

221   clustered again as nearest neighbours in the dendrogram (Figure 4).

222

223   The annotation process of the pBOWs showed that several of the metabolites belongs to the

224   chemical classes of polyphenols, amino acids, dipeptides, tripeptides, bounded terpenoids, sugars

225   and organic acids (Supplementary Tables S2-3). Therefore it was decided to take advantage of the

226   annotation achieved previously using the same protocol in oenological projects and to study more

227   in depth these groups of known metabolites.[11,13,14,21,25,28] With this aim, we turned back to the raw

228   files and integrated a big number of metabolites. This integration process was independent to

229   Progenesis QI workflow, therefore this was also a way to manually check the possible presence of

230   false positive and false negatives markers. Then the integrated areas peak table was uploaded to

231   the Metaboanalyst platform for further statistical analysis and data visualization.

232   Figures 5-7 show the (bio)synthetic pathway of several metabolites of oenological interest,

233   annotated and detected as markers in this study. For each metabolite, data from the heatmap of

234   Supplementary Figures S1 is also shown, in order to compare the relative concentration of each

235   metabolite in the different wine groups. Concerning the amino acids included in Figure 5, Primitivo

236   was the group with the highest amount of leucine, arginine, tyrosine, valine and phenylalanine. On

237   the opposite end, the wine groups with the smallest amounts of the same amino acids were

238   Nebbiolo and Sangiovese. We should take in consideration that yeasts could consume the majority

239   of the amino acids during the alcoholic fermentation as a nitrogen source.[31] Thus common

240   oenological practices, such as addition of inorganic and/or organic nitrogen to support yeast's

241   growth would strongly affect the concentration of amino acids in wine.[31] Since the wines from each

242   group originated from different wineries that followed different winemaking practices, we should not

243   exclude that amino acids could be markers to discriminate wines originated from different cultivar.

244   In the past, the amino acids profile has been proposed as a tool to wine discrimination.[31–33] Proline

245   is the only amino acid not consumed by the yeast in anaerobic condition[31] and because of this

246   characteristic it has been used in food frauds analysis.[34] According to our results, Primitivo wines

247   showed relatively low concentration for this amino acid, with Teroldego showing the highest and

248   Nerello the lowest.

249   Moreover, several di- and tri-peptides were tentatively annotated (3rd level annotation) as markers.

250   According to the nitrogen rule/principle in mass spectrometry, odd m/z values indicate organic

10

251   compound with odd number of nitrogen (thus at least one) and even *m/z* values indicate organic

252   compound with zero or even number of nitrogen. Of course, this rule is valid for organic

253   compounds containing exclusively H, C, N, O, Si, P, S, and halogen, and for high resolution mass

254   spectrometers it is more accurate for *m/z* values below 500. Primitivo wines pBOWs included

255   several ions with odd *m/z* values (Supplementary Tables S2-3), had the highest concentrations in

256   several amino acids (Figure 5), and the tentatively annotated compounds included di- and tri-

257   peptide. If this issue is characteristic for Primitivo, further experiments are necessary to validate

258   this hypothesis and better understand the composition of Primitivo wines, and the contribution of

259   the cultivar and its terroir in determining this unusually richness in nitrogen compounds. Lately,

260   Sherman et al.[35] discovered that wine sensorial quality was positively correlated with markers

261   annotated as di- and tri-peptides. To validate the hypothesis that amino acid profile could be used

262   to distinguish the cultivar in wines, bit analysis on wines produced in more than one harvest are

263   necessary as well as the used of wines produced under the same winemaking conditions and

264   under well controlled agronomical conditions. Indeed, it is well known that in addition to the cultivar,

265   also the terroir (fertilization with nitrogen, grape maturity, climate and the sanitary status) can

266   greatly influence the concentration in nitrogen containing compounds.[36]

267   Primitivo and Sagrantino were the richest wines in tryptophan, while Sangiovese and Raboso and

268   Nebbiolo were the poorest. Conversely, Sangiovese wines were the richest in tryptophol, thus the

269   Ehrlich reaction tryptophan product during the alcoholic formation, and Primitivo the poorest

270   (Figure 5). This was an indication that tryptophan was used by the yeast during the alcoholic

271   fermentation of Sangiovese wines.[31] The lower presence of tryptophol in Primitivo wines was

272   expected, since the Ehrlich pathway is not a preferred way of nitrogen assimilation in presence of

273   an abundant content in amino acids in the juice. Moreover, we found that Sangiovese wines were

274   also the richest in sulfonated tryptophol (Supplementary Figure S1), which is a product of the

275   sulfonation of tryptophol and it formation is favoured by oxygen and lower pH.[14,37,38] Primitivo wines

276   were also the richest in two other N-containing metabolites, tryptophan products during the

277   alcoholic fermentation, N-acetyl-tryptophan ethyl ester and tryptophan ethyl ester.[37] Apparently,

278   tryptophan during Primitivo winemaking process turned to these two ethyl esters and not to the

11

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

305 cultivar or to possible plant stress as a fungal infection,[41] Montepulciano showed the highest

306 concentrations for the glucosidic forms.

307 Figure 6 summarizes another important branch of the general pathway for the synthesis of

308 polyphenols, where the metabolites are divided based on the number of their B-ring substitutes.

309 This figure includes the families of flavanonols (dihydroquercetin, dihyrokaempferol and

310 dihydromirycetin), flavonols (quercetin, isorhamnetin, kaempferol, syringetin, myricetin and

311 laricitrin), anthocyanins (cyanidin, peonidin, delphinidin, malvidin and petunidin), and flavanols

312 (catechin, *epi*catechin, gallocatechin, etc). The kaempferol pathway have only one substitute,

313 quercetin two and myricetin three. It is known that the ratio between these three chemical groups

314 are genetically controlled and often used to distinguish cultivars.[4,5] Teroldego was characterized by

315 the highest concentration in the tri-substitute families, thus to the derivatives of myricetin,

316 delphinidin, petunidin and malvidin. Moreover, Teroldego wines appeared to those with the highest

317 amount of all anthocyanins. Sangiovese wines were the richest in quercetin, followed by Nebbiolo

318 and Nerello. These data are in agreement with a previous study on grapes, where all grapes vines

319 were cultivated in the same vineyard and under the same condition.[5] According to Mattivi et al.[5],

320 myricetin had the highest % between all the flavonols for Teroldego (74%) and Sagrantino (82%),

321 while quercetin the highest for Sangiovese (67%) and Nebbiolo (70%). The same study, that

322 included all the cultivars of the present project except Nerello Mascalese, is in agreement with our

323 outcome about the richness of Teroldego in anthocyanins. In recent years, Sangiovese wines

324 suffer from a problem of instability involving quercetin (and other flavonols), generating flakes

325 floating in the wine that appears in bottled wine.[42] The chemical analysis demonstrated that the

326 major component of these flakes is quercetin aglycon, and so it is believed that is occurring under

327 high amount of quercetin in the wines.[42] As far as our knowledge is concerning this problem was

328 not reported so far in Nebbiolo or Nerello wines, which according to our results had the highest

329 concentration of quercetin after Sangiovese.

330 Nebbiolo was also the group of wines with the highest amount of isorhamnetin, which is the

331 methylation product of quercetin and di-substituted in the B-ring. Also this result was in agreement

332 with Mattivi et al.[5], where isorhamnetin represented the 15% of all flavonols for Nebbiolo. After

13

333   Teroldego, Raboso was the second group of wines in terms of the cyanidin and peonidin amount.

334   For the tri-substitute anthocyanins, after Teroldego, Montepulciano and Sagrantino were the

335   richest cultivars, followed by Aglianico and Cannonau.

336   For the monomeric flavanols, Aglianico was the richest group for catechin and *epi*catechin,

337   followed by Sagrantino and Teroldego for epicatechin, and Sagrantino, Nerello, Nebbiolo and

338   Corvina for catechin. Teroldego was the richest group for epicatechin gallate, followed by

339   Sagrantino and Sangiovese from Romagna. Nerello was the richest in gallocatechin and Teroldego

340   the richest for epigallocatechin. Finally, Sagrantino was also the richest for epigallocatechin gallate

341   (Figure 6). Flavanols is an important family of polyphenols in wine since, between others, it

342   influences wine astringency and bitterness. According to Cheynier et al.[43] *epi*catechin is more bitter

343   that catechin, and the galloylation increases the astringency. Several other monomeric polyphenols

344   inserted having a large variability have also been described to be sensory active, affecting the

345   quality of bitterness and astringency of the red wines.[44]

346   Wine is not just a grape product, but includes a complex technological process (alcoholic

347   fermentation, malolactic fermentation, etc.) and each step enriches and modifies the wine

348   metabolomic fingerprint. Additionally wine metabolites continuously evolve during aging.

349   Anthocyanins, which are the metabolites responsible for the red colour of the wines (and many

350   other food and flowers), participate to a number of reaction during wine aging leading to the

351   production of several classes of wine pigments. As Figure 7 depicted, Teroldego was the richest

352   group in grape anthocyanins, but Aglianico was richest in direct linked and ethyl-bridged linked

353   flavanols-anthocyanins, probably because of its higher content in epicatechin. Particularly rich in

354   ethyl-bridged flavanols-anthocyanins were also Sagrantino, Cannonau and Primitivo. After

355   Aglianico, the richest group in directed linked flavanols-anthocyanins were Sagrantino, Teroldego,

356   Cannonau and Sangiovese. Cannonau, which was the richest in caftaric acid (Figure 5) was also

357   the richest group for some pinotins which are condensation products of hydroxycinnamates with

358   the anthocyanins (Figure 7). Finally, the product of the reaction between malvidin 3-glucoside and

359   acetaldehyde, B-type vitisin was to found to characterize more Cannonau, Raboso and Aglianico;

360   while the product between malvidin 3-glucoside and pyruvic acid characterised the groups of

361   Montepulciano, Aglianico, Sagrantino and Teroldego (Figure 7).

362   One central objective of this project was to study tannins of Italian red wines originated from the

363   grapes, so all the wines were prepared without any tannin addition or contact with wooden barrels.

364   Figure 8 demonstrates a comparison of the wine groups for different monomeric, dimeric, trimeric

365   and tetrameric flavanols, and also included some monomeric sulfonated flavanols. Moreover, the

366   metabolites were divided in 4 families based on the B-ring substitution: a) procyanidins, only

367   constituted by the di-substitute catechin and epicatechin; b) proanthocyanidins, which have at least

368   one tri-substituted gallocatechin or epigallocatechin, and one di-substituted catechin or

369   epicatechin; c) prodelphinidins, only constituted by the tri-substituted gallocatechin and

370   epigallocatechin; and d) gallates, that include at least one galloyl moiety. According to previous

371   researches, the polymerization of tannins decreases the bitterness, and dimers, trimers and

372   tetramers areperceived as more bitter than astringent. As the polymerization increases, initially

373   astringency increases (oligomeric tannins), but as the polymerization further increases astringency

374   decreases (polymeric tannins).[43]

375   Aglianico group was the richest in procyanidins type tannins, followed by Sagrantino and Nebbiolo.

376   These three cultivars are known to produce wines with astringent character. Conversely,

377   Cannonau, Corvina, Montepulciano, Raboso and Nerello showed the smallest amounts of

378   procyanidins. Sagrantino wines were also the richest in mixed proanthocyanidins, followed by

379   Nerello and Nebbiolo; while Primitivo, Corvina and Teroldego were the poorest. As far as concerns

380   the prodelphinidins Sagrantino, Sangiovese, Nerello, Nebbiolo and Teroldego were the richest; and

381   Primitivo, Corvina and Cannonau were the poorest. Sagrantino, Aglianico, Teroldego and Nebbiolo

382   were the richest in galloylated flavanols; while Primitivo, Corvina, Cannonau and Nerello contained

383   the lowest amounts. Raboso, Nerello, Sangiovese from Tuscany and Montepulciano were the

384   wines with the highest concentration on sulfonated tannins (Figure 8).

385

386   Generally, this analytical survey on the untargeted metabolomic fingerprint of 11 Italian mono-

387   cultivar red wines, all together for the first time, highlighted the huge diversity in the composition of

388   these Italian Origin Wines, and generated hypothesis that will need to be validated in the future

389   with targeted approaches. Primitivo was the wine group with the most distinctive metabolome,

390   characterized by the highest amount in several amino acids (tyrosine, phenylalanine, arginine,

391   valine, leucine and isoleucine), and the lowest levels of proline. In agreement with these findings,

392   Primitivo wines were also characterized by a large number of N-containing metabolites. One

393   additional characteristic of Primitivo was the increased level of methylation of both

394   hydroxycinnamates and flavonols. Finally, Primitivo wines were poor in anthocyanins and

395   oligomeric flavanols.

396   Teroldego was also a group of wine with a distinctive metabolomic fingerprint, characterised by the

397   highest amount of anthocyanins, in particular three-substituted anthocyanins at the B-ring.

398   Increased B-ring substitution in Teroldego was also observed for flavonols.

399   Nebbiolo wines were poor in amino acids, hydroxycinnamates, anthocyanins and their derivatives;

400   but rich in kaempferol, isorhametin and quercetin (the 2nd richest group in quercetin after

401   Sangiovese). Condensed tannins were detected in high concentration in Nebbiolo wines, as well

402   as procyanidin gallates and gallic acid. This high galloylation could perhaps explain the astringent

403   character of renowned Nebbiolo wines (Barolo, Barbaresco, etc).[45,46]

404   Aglianico wines were the richest in catechin, epicatechin, procyanidins, type A vitisin, type B vitisin,

405   and the products of reaction between anthocyanins and flavanols (both ethyl-linked and direct-

406   linked). Aglianico samples did not exhibit particularly high levels of anthocyanins, possibly due to

407   the high rate of reaction with flavanols, resulting in the synthesis of stable anthocyanins adducts

408   and therefore more stable color. The high content of monomeric and oligomeric procyanidins could

409   be also responsible for the high astringent character of Aglianico wines.[47,48]

410   Sangiovese, which is the most widespread Italian cultivar, was close to Nebbiolo and Nerello in

411   ESI- analysis, whereas for ESI+ it showed a metabolite profile close to Nebbiolo and Raboso. If we

412   take into consideration all the wine groups, Sangiovese wines were characterized by the B-ring di-

413   substituted flavonols (quercetin derivatives) and anthocyanins (cyanidin 3-glucoside), and the di-

414   substituted hydroxycinnamates (coutaric acid). The tannins of Sangiovese were rich in

415   proanthocyanidins/prodelphinidins with tri-substituted flavanols (gallocatechin and/or

416   epigallocatechin units), while the Sangiovese wines from Tuscany were rich in sulfonated

417   oligomeric flavanols. Finally, Sangiovese wines were poor in amino acids and N-containing

418   metabolites. Overall, Sangiovese wines from Tuscany and Romagna were close and had a very

419   similar metabolome.

420   Cannonau wines were characterised by various caffeic acid metabolites (caftaric acid, caffeoyl

421   derivatives, sulfonated caftaric acid, and pinotins). They were also rich in B-type vitisin, arginine

422   and B-ring methylated flavonoids (syringetin, laricitrin and malvidin derivatives), while relatively

423   poor in tannins.

424   Sagrantino wines showed the highest content of tryptophan, and had intermediate amounts for the

425   other amino acids. Oligomeric tannins were generally high in Sagrantino, both direct-linked and

426   ethyl-linked flavanol-anthocyanins, and the highest levels in proanthocyanidins and

427   epigallocatechin gallate were also detected. Sagrantino wines were also characterized by the

428   highest amounts in flavanonols (dihydroxykaempferol and dihydroxyquercetin), and for relatively

429   high levels of coutaric acid than caftaric and fertaric.

430   Corvina wines were the less homogenous group, with generally low levels in polyphenols (except

431   flavanonols), and highest content of sulfonated glutathione and sulfonated indole lactic acid

432   glucoside. Raboso were characterised by the di-substituted anthocyanins, cyanidin 3-glucoside

433   and peonidin 3- glucoside, and the sulfonated tannins. Montepulciano group was characterised by

434   the acetylated anthocyanins, indole lactic acid and its glucoside, and ellagic acid.

435   In conclusion, the use of a robust untargeted LC-MS based analytical protocol together with a

436   targeted sampling protocol covering a large portion of Italian enological biodiversity produced an

437   interesting publicly available database. For the 11 mono-cultivar red wines investigated, Primitivo,

438   Teroldego and Nebbiolo had the highest number of pBOWs; and a second group comprised

439   Sangiovese, Aglianico, Cannonau and Raboso. Primitivo and Teroldego had the most

440 distinguished metabolomic fingerprint, while Sangiovese with Nebbiolo and Montepulciano with

441 Cannonau had very similar metabolomes. Between the pBOWs were annotated several N-

442 containing metabolites (amino acids, di- and tri-peptides, etc), showing that could be promising

443 metabolites to understand and exploit wine diversity. Especially Primitivo wines were very rich in N-

444 containing metabolites tentative markers. The wines with the richest metabolome in condensed

445 tannins were Sagrantino, Nebbiolo and Aglianico. Teroldego was characterised by the highest

446 amount in anthocyanins, followed by Raboso, Montepulciano, Sagrantino and Aglianico.

447 Sangiovese, Nebbiolo, Nerello and Raboso were characterised by di-substituted flavonoids in the

448 B-ring; and Primitivo, Teroldego, Aglianico, Cannonau and Montepulciano by tri-substituted. In

449 parallel, mono-substituted hydroxycinnamates characterised Sangiovese, Nerello and Raboso, and

450 di- and tri-substituted characterised Primitivo and Cannonau. As expected, the pathway of

451 polyphenols offers many tools in order to understand the metabolomic diversity of the wines.

452 Moreover, even if all wines had the same total SO_2, this wine preservative reacted in a different

453 manner with the metabolites of each wine. In Corvina, Montepulciano and Raboso reacted with

454 ILA-glu; in Teroldego, Corvina, Raboso and Primitivo with glutathione; and Nerello, Sangiovese

455 and Raboso with flavanols. Both raw and analysed data are publicly available, in order to help

456 other researchers in their aim to understand better the Italian oenological diversity and quality.

457

458 **Abbreviations Used:**

459 AGL, Aglianico; PRI, Primitivo; TER, Teroldego; NER, Nerello Mascalese; RAB, Raboso; COR,

460 Corvina; CAN, Cannonau; MON, Montepulciano; SAG, Sagrantino; SAT, Sangiovese Tuscany;

461 SAR, Sangiovese Romagna; NEB, Nebbiolo; QC, Quality Control; pBOWs, putative Biomarkers of

462 Origin Wines; LC, Liquid Chromatography; MS, Mass Spectrometry; FTICR, Fourier-transform ion

463 cyclotron resonance; UPLC-QTOF MS, Ultra-high Performance Liquid Chromatography-

464 Quadrupole Time-of-Flight Mass Spectrometry; PCA, Principal component analysis.

465

466 **Acknowledgements:**

467   Authors acknowledge the other members of the D:Wines consortium: A. Curioni[3], A. Gambuti[4], S.

468   Giacosa[5], D. Perenzoni[1], Luigi Moio[4], A. Rinaldi[4], S. Río Segade[5], G.P. Parpinello[6], A. Ricci[6], B.

469   Simonato[2], G. Tornielli[2], D. Slaghenaufi[2], S. Vincenzi[3]

470

473

474   **Conflict of interest**

475   The authors declare that they have no conflict of interest.

476

477   **Author contributions**

478   All authors conceived and designed the experiment, collected the samples, and read and approved

479   the manuscript; PA performed the LC-MS based metabolomics analysis and data analysis,

480   interpreted the data, and prepared the Tables and Figures. PA, FM, MU, MM and PP wrote the

481   manuscript.

482

483   **AUTHOR INFORMATION**

484   **Corresponding Author**

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

486   **ORCID**

487   Panagiotis Arapitsas 0000-0002-9378-9287

488   Maurizio Ugliano 0000-0002-6487-2866

489   Matteo Marangon 0000-0001-5358-2428

490   Paola Piombino 0000-0001-8705-9846

19

491   Luca Rolle 0000-0002-6075-079X

492   Vincenzo Gerbi 0000-0001-7748-120X

493   Andrea Versari 0000-0003-2960-0619

494   Fulvio Mattivi 0000-0003-4935-5876

495

496   **Supporting Information description**

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

498   Supplementary Table S2. Putative markers list for the analysis in ESI-; including information about

499   the annotation, annotation level, statistical data and the group(s) of wines that were markers.

500   Supplementary Table S3. Putative markers list for the analysis in ESI+; including information about

501   the annotation, annotation level, statistical data and the group(s) of wines that were markers.

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

503

504   **References**

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

506   (2)   Lacombe, T.; Audeguin, L.; Boselli, M.; Bucchetti, B.; Cabello, F.; Chatelet, P.; Crespan, M.;

507   D'Onofrio, C.; Eiras Dias, J.; Ercisli, S.; et al. Grapevine European Catalogue: Towards a

508   Comprehensive List. *VITIS* **2011**, *50* (2), 65–68.

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

510   (4)   Flamini, R.; Mattivi, F.; De Rosso, M.; Arapitsas, P.; Bavaresco, L. Advanced Knowledge of

511   Three Important Classes of Grape Phenolics: Anthocyanins, Stilbenes and Flavonols. *Int. J.*

512   *Mol. Sci.* **2013**, *14* (10), 19651–19669. https://doi.org/10.3390/ijms141019651.

513   (5)   Mattivi, F.; Guzzon, R.; Vrhovsek, U.; Stefanini, M.; Velasco, R. Metabolite Profiling of

514   Grape: Flavonols and Anthocyanins. *J. Agric. Food Chem.* **2006**, *54* (20), 7692–7702.

20

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

516   (6)   de Villiers, A.; Venter, P.; Pasch, H. Recent Advances and Trends in the Liquid-

517   Chromatography–Mass Spectrometry Analysis of Flavonoids. *J. Chromatogr. A* **2016**, *1430*,

518   16–78. https://doi.org/10.1016/J.CHROMA.2015.11.077.

519   (7)   Black, C. A.; Parker, M.; Siebert, T. E.; Capone, D. L.; Francis, I. L. Terpenoids and Their

520   Role in Wine Flavour: Recent Advances. *Aust. J. Grape Wine Res.* **2015**, *21*, 582–600.

521   https://doi.org/10.1111/ajgw.12186.

522   (8)   Gika, H.; Theodoridis, G.; Mattivi, F.; Vrhovsek, U.; Pappa-Louisi, A. Hydrophilic Interaction

523   Ultra Performance Liquid Chromatography Retention Prediction under Gradient Elution.

524   *Anal. Bioanal. Chem.* **2012**, *404* (3), 701–709. https://doi.org/10.1007/s00216-012-6015-6.

525   (9)   Cozzolino, D. Metabolomics in Grape and Wine: Definition, Current Status and Future

526   Prospects. *Food Anal. Methods* **2016**, *9* (11), 2986–2997. https://doi.org/10.1007/s12161-

527   016-0502-x.

528   (10) Diamantidou, D.; Zotou, A.; Theodoridis, G. Wine and Grape Marc Spirits Metabolomics.

529   *Metabolomics* **2018**, *14* (12), 159. https://doi.org/10.1007/s11306-018-1458-1.

530   (11) Arapitsas, P.; Mattivi, F. LC-MS Untargeted Protocol for the Analysis of Wine; Humana

531   Press, New York, NY, 2018; pp 225–235. https://doi.org/10.1007/978-1-4939-7643-0_16.

532   (12) Pinu, F.; Pinu; R, F. Grape and Wine Metabolomics to Develop New Insights Using

533   Untargeted and Targeted Approaches. *Fermentation* **2018**, *4* (4), 92.

534   https://doi.org/10.3390/fermentation4040092.

535   (13) Arapitsas, P.; Speri, G.; Angeli, A.; Perenzoni, D.; Mattivi, F. The Influence of Storage on the

536   "Chemical Age" of Red Wines. *Metabolomics* **2014**, *10* (5), 816–832.

537   https://doi.org/10.1007/s11306-014-0638-x.

538   (14) Arapitsas, P.; Ugliano, M.; Perenzoni, D.; Angeli, A.; Pangrazzi, P.; Mattivi, F. Wine

539   Metabolomics Reveals New Sulfonated Products in Bottled White Wines, Promoted by

540   Small Amounts of Oxygen. *J. Chromatogr. A* **2016**, *1429*, 155–165.

21

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

542   (15)   Carlin, S.; Vrhovsek, U.; Franceschi, P.; Lotti, C.; Bontempo, L.; Camin, F.; Toubiana, D.;

543   Zottele, F.; Toller, G.; Fait, A.; et al. Regional Features of Northern Italian Sparkling Wines,

544   Identified Using Solid-Phase Micro Extraction and Comprehensive Two-Dimensional Gas

545   Chromatography Coupled with Time-of-Flight Mass Spectrometry. *Food Chem.* **2016**, *208*,

546   68–80. https://doi.org/10.1016/j.foodchem.2016.03.112.

547   (16)   Roullier-Gall, C.; Hemmler, D.; Gonsior, M.; Li, Y.; Nikolantonaki, M.; Aron, A.; Coelho, C.;

548   Gougeon, R. D.; Schmitt-Kopplin, P. Sulfites and the Wine Metabolome. *Food Chem.* **2017**,

549   *237*, 106–113. https://doi.org/10.1016/J.FOODCHEM.2017.05.039.

550   (17)   De Rosso, M.; Mayr, C. M.; Girardi, G.; Dalla Vedova, A.; Flamini, R. High-Resolution Mass

551   Spectrometry Metabolomics of Grape Chemical Markers to Reveal Use of Not-Allowed

552   Varieties in the Production of Amarone and Recioto Wines. *Metabolomics* **123AD**, *1*, 124.

553   https://doi.org/10.1007/s11306-018-1415-z.

554   (18)   Arapitsas, P.; Corte, A. D. A. Della; Gika, H.; Narduzzi, L.; Mattivi, F.; Theodoridis, G.

555   Studying the Effect of Storage Conditions on the Metabolite Content of Red Wine Using

556   HILIC LC-MS Based Metabolomics. *Food Chem.* **2016**, *197*, 1331–1340.

557   https://doi.org/10.1016/j.foodchem.2015.09.084.

558   (19)   Roullier-Gall, C.; Witting, M.; Moritz, F.; Gil, R. B.; Goffette, D.; Valade, M.; Schmitt-Kopplin,

559   P.; Gougeon, R. D. Natural Oxygenation of Champagne Wine during Ageing on Lees: A

560   Metabolomics Picture of Hormesis. *Food Chem.* **2016**, *203*, 207–215.

561   https://doi.org/10.1016/j.foodchem.2016.02.043.

562   (20)   Jeandet, P.; Heinzmann, S. S.; Roullier-Gall, C.; Cilindre, C.; Aron, A.; Deville, M. A.; Moritz,

563   F.; Karbowiak, T.; Demarville, D.; Brun, C.; et al. Chemical Messages in 170-Year-Old

564   Champagne Bottles from the Baltic Sea: Revealing Tastes from the Past. *Proc. Natl. Acad.*

565   *Sci. U. S. A.* **2015**, *112* (19), 5893–5898. https://doi.org/10.1073/pnas.1500783112.

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

567   G.; Carlin, S.; Vrhovsek, U.; Mattivi, F.; et al. MetaDB a Data Processing Workflow in

568   Untargeted MS-Based Metabolomics Experiments. *Front. Bioeng. Biotechnol.* **2014**, *2*

569   (DEC), 72. https://doi.org/10.3389/fbioe.2014.00072.

570   (22)   Theodoridis, G.; Gika, H.; Franceschi, P.; Caputi, L.; Arapitsas, P.; Scholz, M.; Masuero, D.;

571   Wehrens, R.; Vrhovsek, U.; Mattivi, F. LC-MS Based Global Metabolite Profiling of Grapes:

572   Solvent Extraction Protocol Optimisation. *Metabolomics* **2012**, *8* (2), 175–185.

573   https://doi.org/10.1007/s11306-011-0298-z.

574   (23)   Etièvant, P.; Schlich, P.; Cantagrel, R.; Bertrand, M.; Bouvier, J.-C. Varietal and Geographic

575   Classification of French Red Wines in Terms of Major Acids. *J. Sci. Food Agric.* **1989**, *46*

576   (4), 421–438. https://doi.org/10.1002/jsfa.2740460405.

577   (24)   Parpinello, G. P.; Ricci, A.; Arapitsas, P.; Curioni, A.; Moio, L.; Riosegade, S.; Ugliano, M.;

578   VERSARI, A. Multivariate Characterisation of Italian Monovarietal Red Wines Using MIR

579   Spectroscopy. *OENO One* **2019**, *53* (4). https://doi.org/10.20870/oeno-one.2019.53.4.2558.

580   (25)   Shahaf, N.; Franceschi, P.; Arapitsas, P.; Rogachev, I.; Vrhovsek, U.; Wehrens, R.

581   Constructing a Mass Measurement Error Surface to Improve Automatic Annotations in

582   Liquid Chromatography/Mass Spectrometry Based Metabolomics. *Rapid Commun. Mass*

583   *Spectrom.* **2013**, *27* (21), 2425–2431. https://doi.org/10.1002/rcm.6705.

584   (26)   Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W.-

585   M.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; et al. Proposed Minimum Reporting Standards for

586   Chemical Analysis. *Metabolomics* **2007**, *3* (3), 211–221. https://doi.org/10.1007/s11306-007-

587   0082-2.

588   (27)   Moro, L.; Da Ros, A.; da Mota, R. V.; Purgatto, E.; Mattivi, F.; Arapitsas, P. LC–MS

589   Untargeted Approach Showed That Methyl Jasmonate Application on Vitis Labrusca L.

590   Grapes Increases Phenolics at Subtropical Brazilian Regions. *Metabolomics* **2020**, *16* (2),

591   18. https://doi.org/10.1007/s11306-020-1641-z.

592   (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 *OenoIvas2019*; 2019; p 107.

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

619   (37)   Álvarez-Fernández, M. A.; Fernandez-Cruz, E.; García Parrilla, M. C.; Troncoso, A. M.;

620   Mattivi, F.; Vrhovsek, U.; Arapitsas, P. Saccharomyces Cerevisiae and Torulaspora

621   Delbrueckii Intra- and Extra-Cellular Aromatic Amino Acids Metabolism. *J. Agric. Food*

622   *Chem.* **2019**, *67* (28), 7942–7953. https://doi.org/10.1021/acs.jafc.9b01844.

623   (38)   Arapitsas, P.; Guella, G.; Mattivi, F. The Impact of SO2 on Wine Flavanols and Indoles in

624   Relation to Wine Style and Age. *Sci. Rep.* **2018**, *8* (1), 858. https://doi.org/10.1038/s41598-

625   018-19185-5.

626   (39)   Ong, B. Y.; NAGEL, C. W. Hydroxycinnamic Acid-Tartaric Acid Ester Content in Mature

627   Grapes and During the Maturation of White Riesling Grapes. *Am. J. Enol. Vitic.* **1978**, *29*,

628   277–281.

629   (40)   Singleton, V. L.; Zaya, J.; Trousdale, E. K. Caftaric and Coutaric Acids in Fruit of Vitis.

630   *Phytochemistry* **1986**, *25* (9), 2127–2133. https://doi.org/10.1016/0031-9422(86)80078-4.

631   (41)   Gatto, P.; Vrhovsek, U.; Muth, J.; Segala, C.; Romualdi, C.; Fontana, P.; Pruefer, D.;

632   Stefanini, M.; Moser, C.; Mattivi, F.; et al. Ripening and Genotype Control Stilbene

633   Accumulation in Healthy Grapes. *J. Agric. Food Chem.* **2008**, *56* (24), 11773–11785.

634   https://doi.org/10.1021/jf8017707.

635   (42)   Mattivi, F.; Arapitsas, P.; Perenzoni, D.; Guella, G.; Speri, G.; Angeli, A.; Perenzoni, D.;

636   Mattivi, F. The Influence of Storage on the "Chemical Age" of Red Wines. *ACS Symp. Ser.*

637   **2014**, *1203*, 29–49. https://doi.org/10.1007/s11306-014-0638-x.

638   (43)   Cheynier, V.; Prieur, C.; Guyot, S.; Rigaud, J.; Moutounet, M. The Structures of Tannins in

639   Grapes and Wines and Their Interactions with Proteins; 1997; pp 81–93.

640   https://doi.org/10.1021/bk-1997-0661.ch008.

641   (44)   Hufnagel, J. C.; Hofmann, T. Orosensory-Directed Identification of Astringent Mouthfeel and

642   Bitter-Tasting Compounds in Red Wine. *J. Agric. Food Chem.* **2008**, *56* (4), 1376–1386.

643   https://doi.org/10.1021/jf073031n.

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

645   Hydroxycinnamoyl Tartrates and Acylated Anthocyanins in the Skin of 34 Vitis Vinifera

646   Genotypes. *J. Agric. Food Chem.* **2012**, *60* (19), 4931–4945.

647   https://doi.org/10.1021/jf2045608.

648   (46)   Petrozziello, M.; Torchio, F.; Piano, F.; Giacosa, S.; Ugliano, M.; Bosso, A.; Rolle, L. Impact

649   of Increasing Levels of Oxygen Consumption on the Evolution of Color, Phenolic, and

650   Volatile Compounds of Nebbiolo Wines. *Front. Chem.* **2018**, *6*, 137.

651   https://doi.org/10.3389/fchem.2018.00137.

652   (47)   Gambuti, A.; Rinaldi, A.; Pessina, R.; Moio, L. Evaluation of Aglianico Grape Skin and Seed

653   Polyphenol Astringency by SDS–PAGE Electrophoresis of Salivary Proteins after the

654   Binding Reaction. *Food Chem.* **2006**, *97* (4), 614–620.

655   https://doi.org/10.1016/J.FOODCHEM.2005.05.038.

656   (48)   Rinaldi, A.; Jourdes, M.; Teissedre, P. L.; Moio, L. A Preliminary Characterization of

657   Aglianico (Vitis Vinifera L. Cv.) Grape Proanthocyanidins and Evaluation of Their Reactivity

658   towards Salivary Proteins. *Food Chem.* **2014**, *164*, 142–149.

659   https://doi.org/10.1016/J.FOODCHEM.2014.05.050.

660

661

662 **FIGURE CAPTIONS**

663

664 **Figure 1.** Distribution of the wine sample set according to their cultivar (black) and region (red).

665 The principal denomination of origin of each cultivar/region are also presented (light blue). The

666 cultivation area refers to all Italy for each cultivar for the year 2015.[1]

667 **Figure 2.** PCA plots of all the wines in ESI+ (above) and ESI- (bellow). AGL, Aglianico; PRI,

668 Primitivo; TER, Teroldego; NER, Nerello Mascalese; RAB, Raboso; COR, Corvina; CAN,

669 Cannonau; MON, Montepulciano; SAG, Sagrantino; SAT, Sangiovese Tuscany; SAR, Sangiovese

670 Romagna; NEB, Nebbiolo; QC, Quality Control.

671 **Figure 3.** Number of pBOW features for each cultivar in ESI+ and ESI-. Unique are the pBOWs

672 that helps to discriminate the cultivar for all the others.

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

674 **Figure 5.** Biosynthesis and synthesis of N-containing metabolites, hydroxycinnamates and

675 stilbenoids annotated in this study. Colours refers to the heat-map of Supplementary Figure S1 and

676 represent a comparison of the concentration of each metabolite between the various mono-cultivar

677 wine groups. The heatmap was build using Pareto scaling and Euclidean distance. AGL, Aglianico;

678 PRI, Primitivo; TER, Teroldego; NER, Nerello Mascalese; RAB, Raboso; COR, Corvina; CAN,

679 Cannonau; MON, Montepulciano; SAG, Sagrantino; SAT, Sangiovese Tuscany; SAR, Sangiovese

680 Romagna; NEB, Nebbiolo.

681 **Figure 6.** General pattern for flavonoids biosynthesis, with the metabolites annotated in this study.

682 Colours refers to the heat-map of Supplementary Figure S1 and represent a comparison of the

683 concentration of each metabolite between the various mono-cultivar wine groups. The heatmap

684 was build using Pareto scaling and Euclidean distance. AGL, Aglianico; PRI, Primitivo; TER,

685 Teroldego; NER, Nerello Mascalese; RAB, Raboso; COR, Corvina; CAN, Cannonau; MON,

686 Montepulciano; SAG, Sagrantino; SAT, Sangiovese Tuscany; SAR, Sangiovese Romagna; NEB,
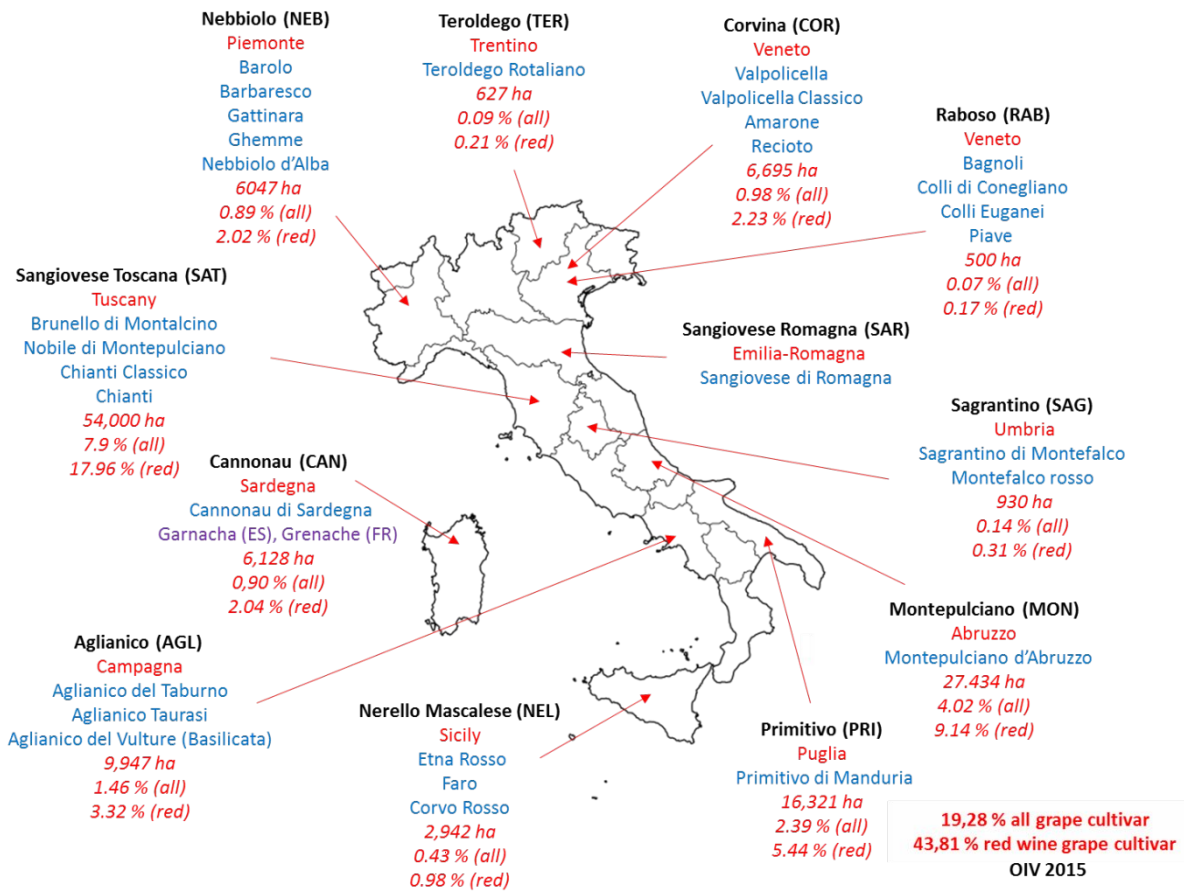
687 Nebbiolo.

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.

704   **FIGURES**

705

Nebbiolo (NEB)
Piemonte
Barolo
Barbaresco
Gattinara
Ghemme
Nebbiolo d'Alba
6047 ha
0.89 % (all)
2.02 % (red)

Teroldego (TER)
Trentino
Teroldego Rotaliano
627 ha
0.09 % (all)
0.21 % (red)

Corvina (COR)
Veneto
Valpolicella
Valpolicella Classico
Amarone
Recioto
6,695 ha
0.98 % (all)
2.23 % (red)

Raboso (RAB)
Veneto
Bagnoli
Colli di Conegliano
Colli Euganei
Piave
500 ha
0.07 % (all)
0.17 % (red)

Sangiovese Toscana (SAT)
Tuscany
Brunello di Montalcino
Nobile di Montepulciano
Chianti Classico
Chianti
54,000 ha
7.9 % (all)
17.96 % (red)

Sangiovese Romagna (SAR)
Emilia-Romagna
Sangiovese di Romagna

Sagrantino (SAG)
Umbria
Sagrantino di Montefalco
Montefalco rosso
930 ha
0.14 % (all)
0.31 % (red)

Cannonau (CAN)
Sardegna
Cannonau di Sardegna
Garnacha (ES), Grenache (FR)
6,128 ha
0,90 % (all)
2.04 % (red)

Montepulciano (MON)
Abruzzo
Montepulciano d'Abruzzo
27.434 ha
4.02 % (all)
9.14 % (red)

Aglianico (AGL)
Campagna
Aglianico del Taburno
Aglianico Taurasi
Aglianico del Vulture (Basilicata)
9,947 ha
1.46 % (all)
3.32 % (red)

Nerello Mascalese (NEL)
Sicily
Etna Rosso
Faro
Corvo Rosso
2,942 ha
0.43 % (all)
0.98 % (red)

Primitivo (PRI)
Puglia
Primitivo di Manduria
16,321 ha
2.39 % (all)
5.44 % (red)

19,28 % all grape cultivar
43,81 % red wine grape cultivar
OIV 2015
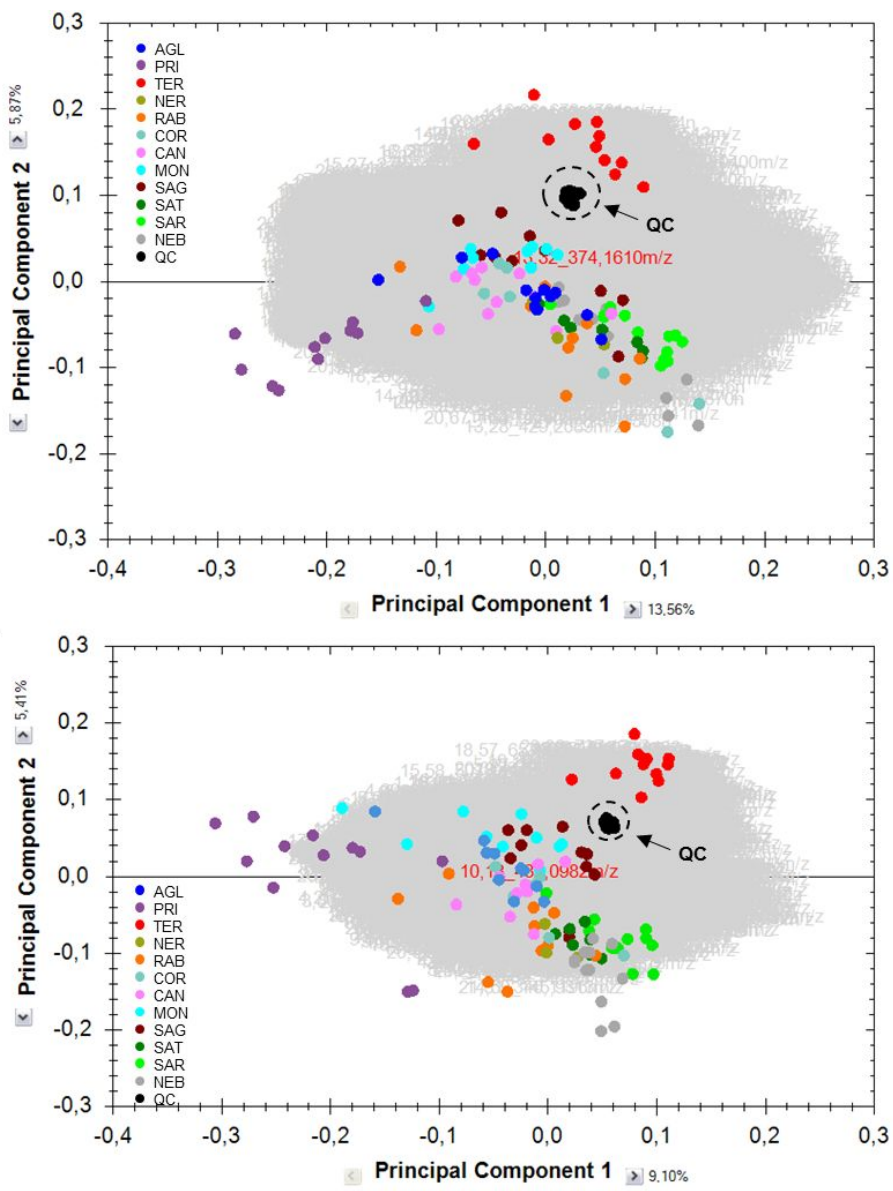
706

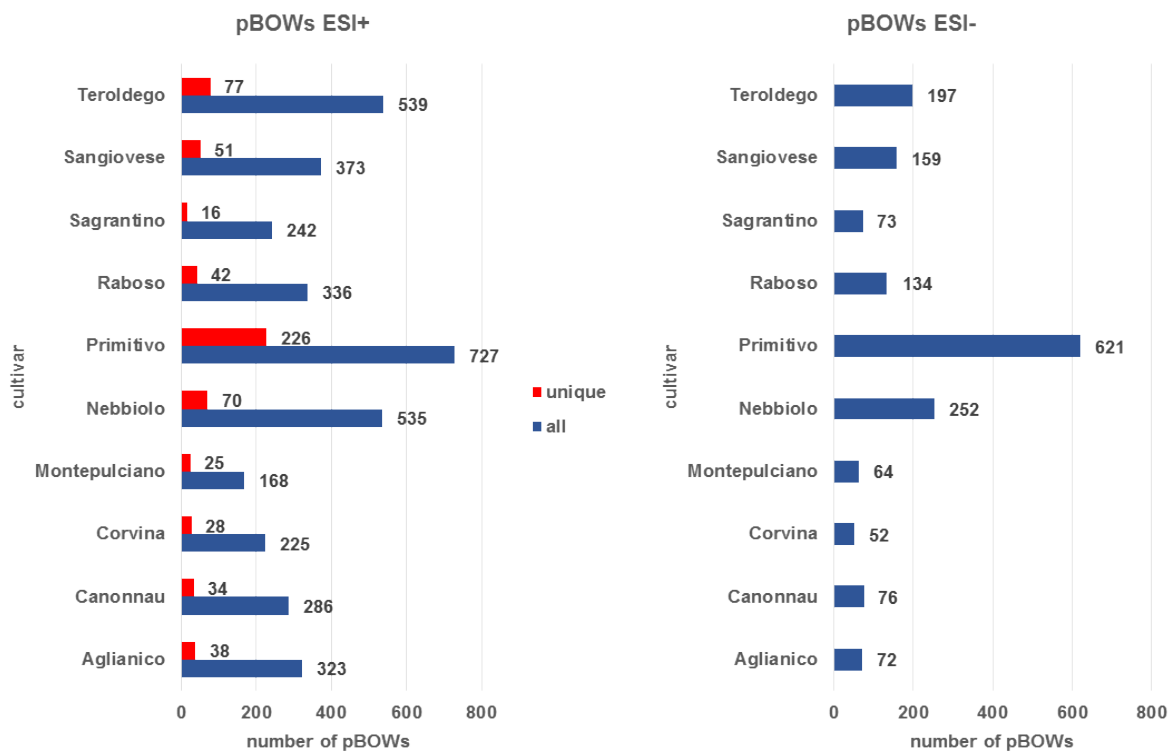707   Figure 1

708

709

Figure 2

710

711   Figure 3

712
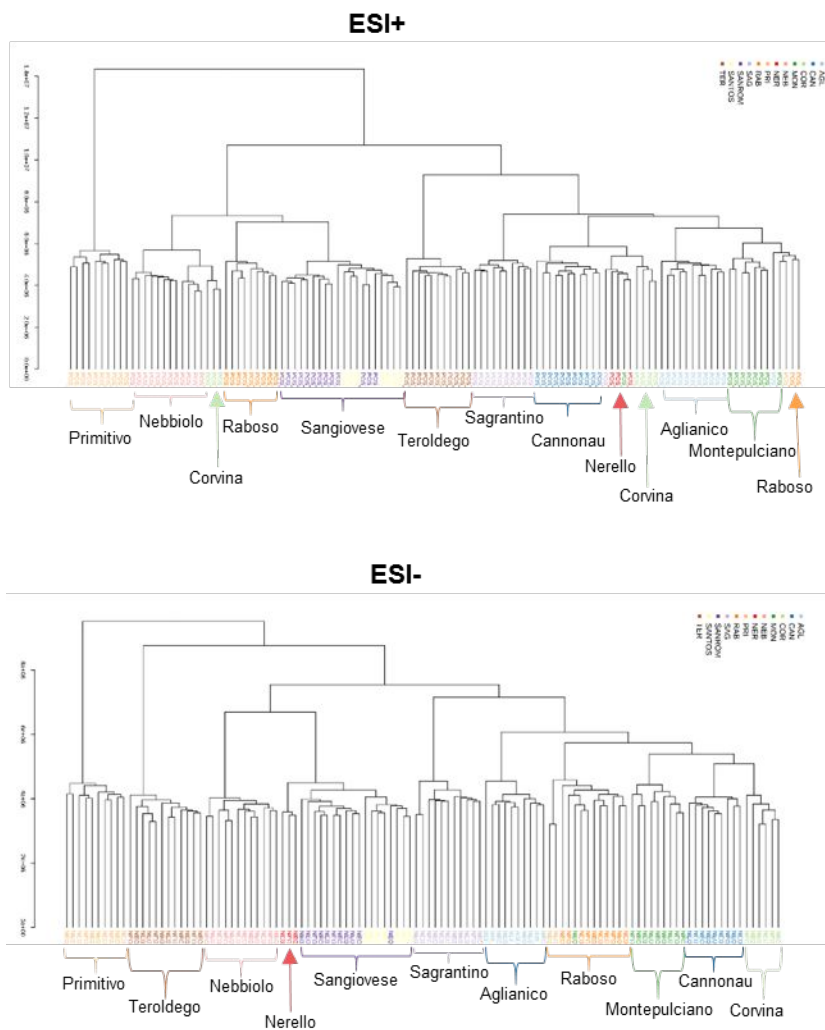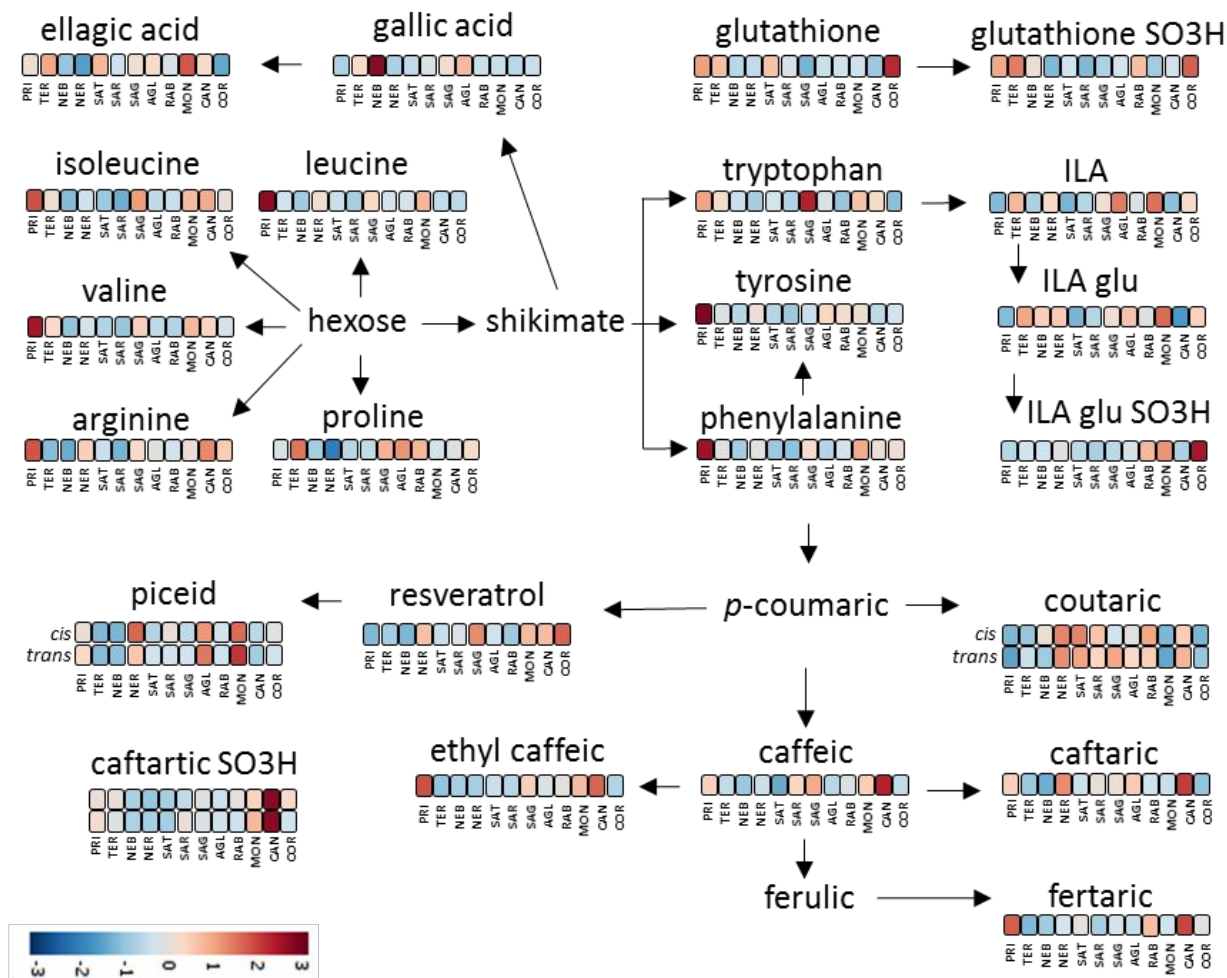
Figure 4

713

714

715

716
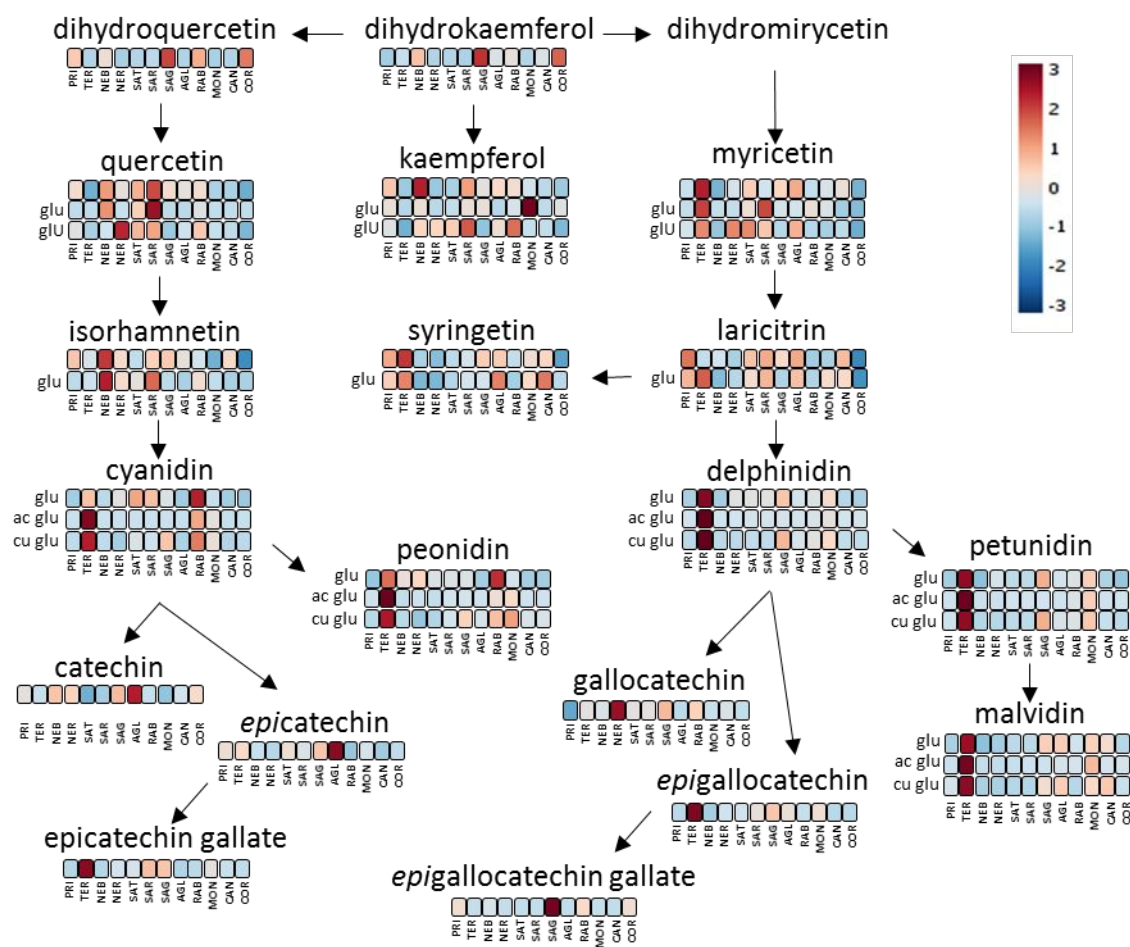
717

718

719   Figure 5
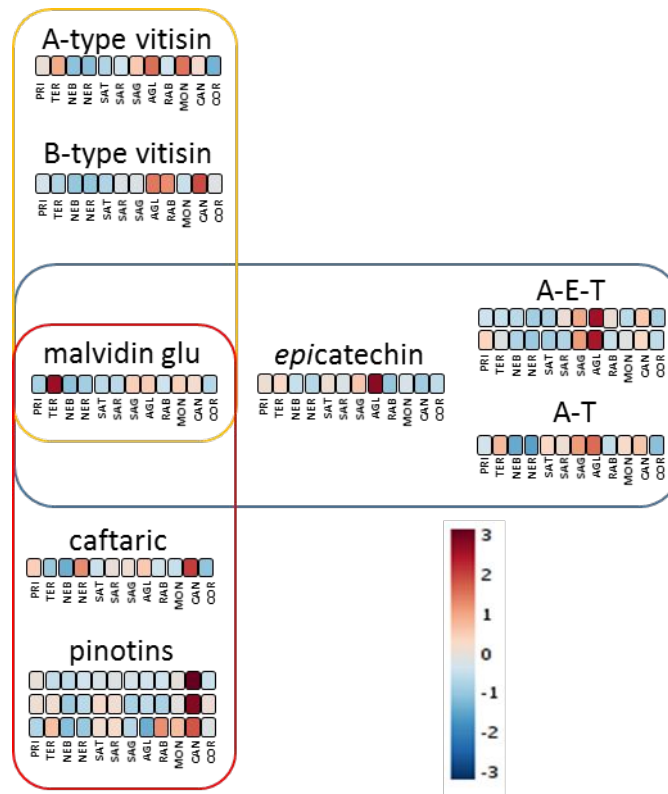
720
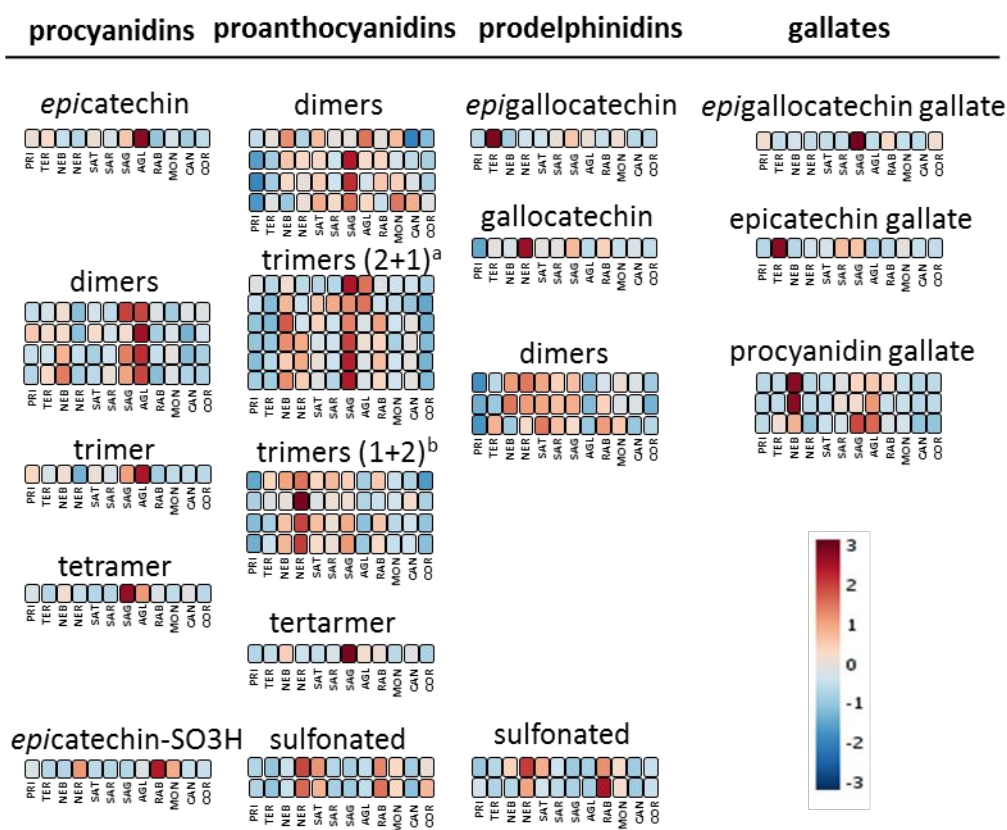
721

722

723   Figure 6

724

725   Figure 7

726

727

728

729

730

Figure 8