Food Research International 177 (2024) 113867

Contents lists available at ScienceDirect

Food Research International

journal homepage: www.elsevier.com/locate/foodres

Quantitative and qualitative composition of proanthocyanidins and other polyphenols in commercial red wines and their contribution to sensorially evaluated tannicity

Juuso Erik Laitila [a,*], Juha-Pekka Salminen [a]

[a] Natural Chemistry Research Group, Department of Chemistry, University of Turku, FI-20500, Turku, Finland

ARTICLE INFO

Keywords:
Adduct
Astringency
Mass spectrometry
Chromatographic fingerprint
Tannin
Proanthocyanidin

ABSTRACT

The analysis of proanthocyanidins (PA) in red wine has typically been conducted using few key methods, such as phloroglucinolysis or precipitation assays. Here, the content of PAs and other common polyphenol groups in commercial red wines were analyzed with a group-specific liquid chromatography–tandem mass spectrometry method. Besides concentrations, the method provides qualitative information about the detected compound groups in the form of two-dimensional (2D) chromatographic fingerprints. The 2D fingerprints of PAs have not been utilized in analysis of red wine before. For instance, 2D chromatographic fingerprints revealed that the complex PA compositions were qualitatively notably similar between many wine types, even when there were considerable differences in concentrations. Finally, 201 commercial red wines had been categorized as either tannic or medium tannic based on their sensorial evaluations. The content of PAs and three different groups of oligomeric adducts of malvidin glycosides and PAs were measured from these wines. The compositional features of the PAs and PA–malvidin glycoside adducts were more important than concentrations in explaining the perceived tannicity.

1. Introduction

Proanthocyanidins (PAs) are oligomers and polymers consisting of flavan-3-ol monomers (FL; Fig. 1). The most common monomeric building units of PAs in red wine are catechins and gallocatechins (Hanlin et al., 2011), which are subsequently called procyanidin (PC) and prodelphinidin (PD) units when they are part of an oligomer or polymer. There is some additional variation in the sub-unit composition of PAs in the plant kingdom due to the variation in, e.g., the hydroxylation patterns of the phenolic rings (Dixon et al., 2005). However, the complexity of PA mixtures comes from the numerous ways that the building units can be linked to one another and, overall, plants can produce hundreds of different types of PAs (Leppä et al., 2018; Salminen, 2018).

PA moieties are found in red wine in different types of oligomeric or polymeric adducts that are formed from reactions between anthocyanins and PAs during wine making (Cheynier et al., 2006). This increases the complexity of the sub-unit composition of PA containing compounds in red wine in comparison to naturally occurring PAs. The anthocyanin moieties can be located in the terminal end of the oligomeric chain (Salas et al., 2003), in the extension units (Remy-Tanneau et al., 2003) or the anthocyanin and PA moieties can be linked via an aryl or alkyl linkage, typically methylmethine linkage (Fig. 1; Lee et al., 2004). The notation PA–Mv adducts will be used in this article to refer to the oligomeric adducts of PAs and malvidin glycosides (Mv) in general without referencing any specific sub-group, while abbreviations for the above-

---

Abbreviations: ANOVA, analysis of variance; FL, flavan-3-ol monomer; FL/PA ratio, ratio of the concentrations of FLs and proanthocyanidins; HSD, Honest Significant Difference; LC, liquid chromatography; LOA, large oligomeric adduct; MOA, medium-sized oligomeric adduct; MRM, multiple reaction monitoring; MS, mass spectrometry; MS/MS, tandem mass spectrometry; Mv, malvidin glycoside; PA, proanthocyanidin; PA–methylmethine–Mv$^+$, methylmethine-linked proanthocyanidin–malvidin glycoside adducts; PA–Mv$^+$, directly-linked proanthocyanidin–malvidin glycoside adduct where the malvidin unit is the terminal unit; PA–Mv–FL, proanthocyanin–malvidin glycoside–flavan-3-ol adducts where the bicyclic Mv–FL unit is in the terminal end; PC, procyanidin; PD, prodelphinidin; PLS-DA, partial least squares discriminant analysis; SOAs, small oligomeric adduct; UPLC, ultra-performance liquid chromatography; VIP, variable importance on projection.

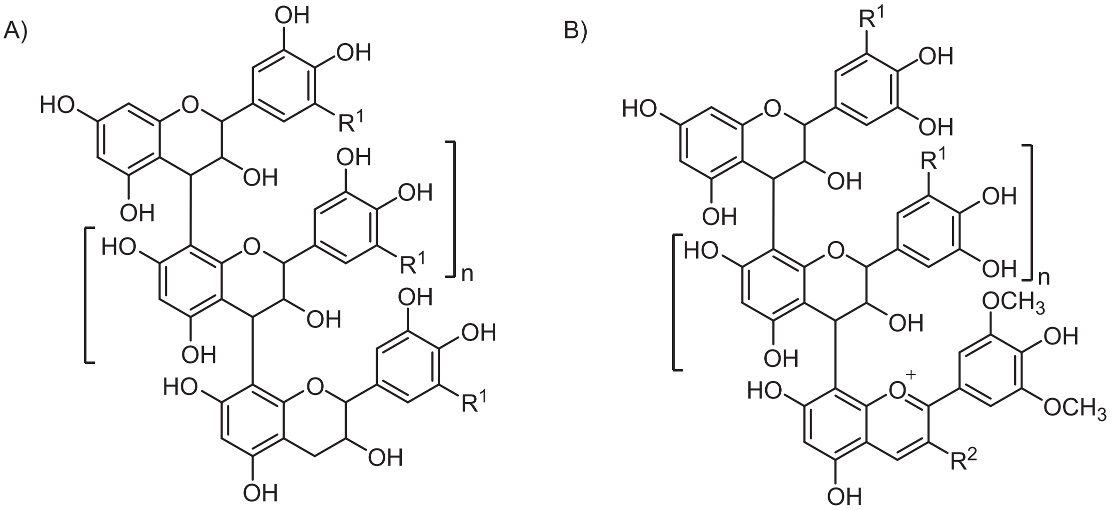
[*] Corresponding author.

E-mail addresses: juerlai@utu.fi (J.E. Laitila), j-p.salminen@utu.fi (J.-P. Salminen).

https://doi.org/10.1016/j.foodres.2023.113867
Received 13 September 2023; Received in revised form 4 December 2023; Accepted 14 December 2023

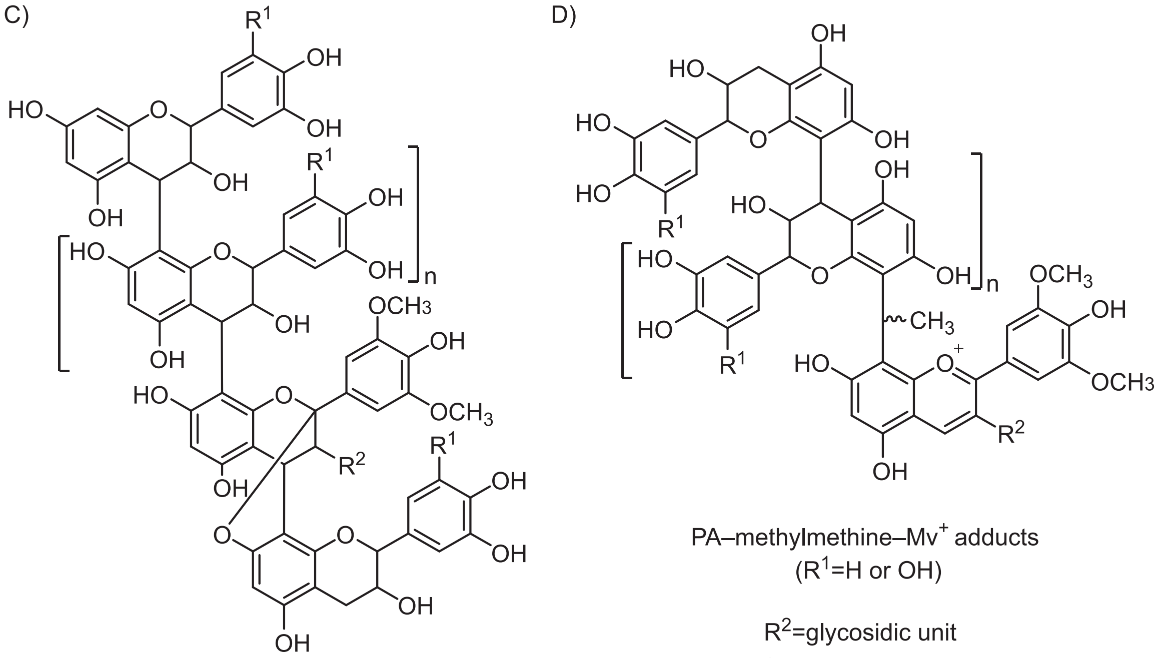
mentioned specific sub-groups are given in Fig. 1. PAs themselves can also react with one another via acetaldehyde yielding methylmethine-linked PAs, red wine has been shown to contain unusual crown PAs, and vinylcatechins released from methylmethine-linked PAs by acid-catalyzed depolymerization can form pyranoanthocyanin–PA adducts with anthocyanins (Es-Safi et al., 1999; Longo et al., 2019; Mateus et al., 2004). The given examples do not make up a complete list of all types of PA containing compound groups in red wine, but instead they are examples that demonstrate how PAs are found in various and often unique compound groups in red wine.

Engström et al. (2014) published a group-specific UPLC–MS/MS method in 2014 that detects PC and PD containing PAs. The Engström method utilizes in-source collision induced dissociation to first fragment PAs to their subunits by quinone methide fragmentation (Friedrich et al., 2000), and then uses selected reaction monitoring to detect the released PC and PD units with a separate detection channel for their terminal and extension units. This methodology produces two-dimensional (2D) fingerprints of the targeted compound groups that preserve information about the initial compound composition, even though PAs are depolymerized in the ion source. This is possible, because the depolymerization



Proanthocyanidins (PAs):
Procyanidin unit: R^1=H
Prodelphinidin unit: R^1=OH

PA–Mv^+ adducts
(R^1=H or OH)



PA–Mv–FL adducts
(R^1=H or OH)

R^2=glycosidic unit
(commonly glucoside,
acetylglucoside or
coumaroylglucoside)

Fig. 1. Structures of (A) proanthocyanidins (PA) and three groups of PA–malvidin glycoside adducts (B–D). Abbreviations: PA, proanthocyanidin; FL, flavan-3-ol monomer; Mv, malvidin glycoside.

is preceded by chromatographic separation of the intact PAs in the UPLC system. The Engström method can be used to quantify concentrations of PAs and to determine mean degree of polymerization of PAs as well, and the method has been used in red wine analytics for this purpose before (Engström et al., 2014; Laitila, 2021). However, the 2D fingerprints provide an entirely new way to measure and evaluate qualitative composition of PA containing compounds in red wine that is not achieved by other methods based on depolymerization, such as phloroglucinolysis or thiolysis. This is especially valuable in analysis of red wine, because the PA containing compounds dot not consist only of natural PAs as discussed. Laitila et al. (2019) utilized similar concept as in the Engström method to develop a group-specific UPLC–MS/MS method for the detection of three different groups oligomeric adducts of PA–Mv adducts (Fig. 1). In the context of red wine, the Engström method is a more universal method for the detection of many types of compounds with PA moieties (as proposed in Fig. S1), whereas the method of Laitila et al. (2019) detects a specific portion of the compounds that the Engström method detects, i.e., certain types of PA–Mv adducts.

The first objective of the present study was to investigate the polyphenolic composition of commercial red wines (n = 234) by screening the red wines with the Engström method (2014, 2015). Besides PAs, the Engström measures 2D fingerprints and concentrations of several other common polyphenolic compound groups found in red wine, i.e., quercetin, myricetin, hexahydroxydiphenoyl (HHDP), and galloyl derivatives, and catechins and gallocatechins (Engström et al., 2015). However, the emphasis was on the PAs in this study. The specific aims were to utilize the 2D chromatographic fingerprints to identify any qualitative compositional differences between red wines, and to examine whether the quantitative or qualitative composition varies more between red wines. Additionally, the 2D chromatographic fingerprints of the detected compound groups were used to study their qualitative compositional evolution, together with the quantitative evolution. Finally, the 2D fingerprints of PAs produced by the Engström method were to be visually compared to the published 2D fingerprints of the PA–Mv adducts in red wine (Laitila, 2021) to examine whether features of the 2D fingerprints of the PA–Mv adducts were visible in the 2D fingerprints of PAs produced by the Engström method.

The second major objective was to study the linkage between sensorially evaluated tannin content, i.e., tannicity, and the composition of PAs and PA–Mv adducts. In total, 201 commercial red wines that were provided by Alko Inc. were evaluated in discriminant sensorial analysis at Alko Inc. as either “medium tannic” or “tannic”, with the latter group was perceived as the one with higher tannin content. Alko Inc. is a Finnish state-owned company that has the sole right in Finland to engage in the retail sale of beverages with more than 5.5 % alcohol by volume. Polymeric pigments have been linked to astringency in general, and to specific astringency-related sensations before (Casassa et al., 2013; Gawel et al., 2007; Ristic et al., 2010). However, the utilized analytical methods have been universal for polymeric pigments in general, and they have been unable to detect specific subgroups separately. The UPLC–MS/MS method of Laitila et al. (2019) overcomes many of the previous analytical limitations as the method can detect three specific sub-groups of PA–Mv adducts regardless of the size and structure of the PA moieties in them. The method also provides information about the relative sizes of the measured adducts between wine samples (Laitila et al., 2019). Therefore, the wines that were sensorially evaluated were analyzed using the UPLC–MS/MS methods of Laitila et al. (2019) and Engström et al. (2014) to gain information about the composition and concentrations of PAs and PA–Mv adducts that could explain the sensorially evaluated tannicity.

2. Materials and methods

2.1. Chemicals and red wines

LC–MS grade acetonitrile was purchased from Sigma-Aldrich (Steinheim, Germany) and LC–MS grade formic acid was purchased from VWR (Helsinki, Finland). Ultra-pure type 1 water was produced with Millipore Synergy UV water purification system (Merck KGaA, Darmstadt Germany). Catechin, epicatechin, gallogatechin, epigallocatechin, quercetin-3-*O*-galactoside and myricetin-3-*O*-rhamnoside were purchased from Extrasynthese (Lyon, France) and gallic acid from Sigma. This study utilized a set of commercial red wines that were provided mostly by Alko Inc. (n = 272), while some were purchased by the Natural Chemistry Research Group of University of Turku (n = 45). Specific sub-sets of the whole wine set were utilized in this study instead of the whole wine set, and the characteristics of the sub-sets are given in Supplementary Tables S1 and S2. Rationale for selection of the wines is provided in results and discussion. All wines were commercial red wines and aliquots of six milliliters were stored at –80 °C prior to the chemical analyzes. Formation of precipitate was not observed during the storage in any of the wines, and few wines were monitored periodically using the UPLC system described in chapter 2.3. to ensure that the chemical composition was not altered during the storage.

2.2. Well-plate reader assay for total polyphenolic content

Total polyphenolic content of the red wines was measured with a 96-well plate reader using Folin–Ciocalteu method as modified by (Salminen & Karonen, 2011).

2.3. UPLC–DAD-TQ–MS/MS analyzes

The Acquity UPLC system consisted of a binary solvent manager, sample manager, column oven and diode array detector. The column was an Acquity BEH Phenyl column (100 mm × 2.1 mm i.d.; 1.7 µm particle size; Waters Corp, Wexford, Ireland). The eluents (acetonitrile and 0.1 % HCOOH) and the gradient were the same that were used in the UPLC–MS/MS method in Laitila et al. (2019). The LC system was coupled to a Xevo triple quadrupole mass spectrometer (Waters Corporation, Milford, MA). The PC and PD containing PAs, catechins, gallocatechins and quercetin, myricetin, HHDP and galloyl derivatives were analyzed using the UPLC–MS/MS method of Engström et al., (2014, 2015) as described in Leppä et al (2020). The Engström method contains additionally an optimized selected reaction monitoring channel for the detection of monomeric FLs (Engström et al., 2014) and this channel was used to detect catechin, epicatechin, gallocatechin and epigallocatechin. Quercetin derivatives were quantified using quercetin-3-*O*-galactoside and myricetin derivatives by using myricetin-3-*O*-rhamnoside. HHDP and galloyl derivatives (excluding gallic acid) were not converted into mg/L concentrations due to lack of suitable standards, but instead they were semi-quantified as areas of the 2D chromatographic fingerprints (chromatograms). Only ellagic acid peak was integrated from HHDP fingerprints because it was the only major compound in this group. Gallic acid was quantified from the UPLC–DAD trace (280 nm) against an authentic standard. Total FL concentration, catechin-% of the FLs and FL/PA-ratio were calculated from the individually quantified catechins and gallocatechins. The complete operating parameters of the MS instrument for the Engström method are described in Engström et al. (2014). Briefly, the capillary voltage of the ion source was 2.4 kV, desolvation temperature was 650 °C and the source temperature was 150 °C, nitrogen was used as desolvation and cone gas and their flow rates were 1000 and 100 L/h, respectively. The oligomeric PA–Mv^+, PA–methylmethine–Mv^+ and PA–Mv–FL adducts were detected with the same instrumentation and analytical conditions (except ionization mode) by using the UPLC–MS/MS method of Laitila et al. (2019) as described in Laitila and Salminen (2020). All red wines were analyzed as such after filtration through 0.20 µm PTFE filters.

2.4. UPLC–DAD–Orbitrap–MS analyzes

The LC system was identical to the one used in the

UPLC–DAD–TQ–MS/MS analyzes (section 2.3.). The UPLC system was coupled to a Q Exactive Orbitrap MS system (Thermo Fisher Scientific GmbH, Bremen, Germany). The heated electrospray ion source was operated in negative ion mode. Spray voltage was set to –3.0 kV and capillary temperature to 380 °C. Nitrogen gas was used as sheath gas and auxiliary gas and their flow rates were 60 and 20 units, respectively. S-lens radio frequency level was set to 60. In the full scan analyzes, the scanned *m/z* range was 150–2250 and the resolution was set to 70 000, automatic gain control to 3×10^6 and data was collected between 0 and 8 min. External calibration was done using Pierce ESI Negative Ion Calibration Solution (Thermo Fischer Scientific Inc., Waltham, MA). Data was processed with Thermo Xcalibur Qual Browser (version 4.1.31.9).

2.5. Sensorial analysis of tannicity

201 red wines (Table S2) were categorized in discriminant sensorial analyzes as either “tannic” or “medium tannic”, with the latter being the group perceived as less tannic. Tannicity was defined as the drying sensation in the mucous membrane in mouth, which could be felt with a delay after expectoration. This definition coincides with the definition of astringency by the American Society of Testing Materials, who defines astringency as “the complex of sensations due to shrinking, drawing or puckering of the epithelium as a result of exposure to substances such as alums or tannins” (García-Estévez et al., 2018). The sensorial analyzes were performed at Alko Inc. as part of their routine analyzes of red wines in their selection. A tasting group consisting of approximately 30 employees of Alko Inc. did the sensory evaluations of the wines. The suitability and expertise of each member for sensorial analysis was evaluated before they got to join the tasting group. The sensorial analysis team receives regular training about sensorial analysis in general, but they were not specifically trained for these particular sensorial analyzes. First, each wine was individually and separately evaluated by three or four persons. The evaluations were combined and if the panelists were not unanimous on the tannicity, the wine could be re-evaluated before deciding the final classification. Alko Inc. only provided the final classification of each wine, and no information about the evaluations of the individual panelists, or their identity, or any other information concerning the panelists was provided, ensuring complete anonymity of the panelists. Alko Inc. routinely performs similar sensorial evaluations, and the tannicity categories are provided at the web of Alko Inc. at https://www.alko.fi for most red wines in their selection. Therefore, the sensorial data could be considered as public information.

2.6. Statistical analyzes

All statistical analyzes were conducted using R (version 3.6.3) in Rstudio integrated development environment (version 1.2.5033; Rstudio Team, 2019; R Core Team, 2020). Principal component analysis was done with the prcomp function and data was log-transformed and standardized to zero means and unit variances for the analysis. Subsequent k-means clustering was done using the kmeans function and the principal component scores of the first four components. Samples were divided to five clusters starting from 100 random centers. Data visualization related to the principal component and clustering analyzes was done using the “*ggplot2*” (version 3.3.0) and “*factoextra*” (version 1.0.6.) packages (Kassambara & Mundt, 2019; Wickham, 2016). Quantitative differences in the concentrations of the determined compound groups and other continuous variables were tested with one-way analysis of variance (ANOVA, aov function in R) and Tukey’s Honest Significant Difference tests (HSD, TukeyHSD function in R). The tests were done at risk level of 5 %. The connection between the perceived tannin content of the wines (tannicity; response) and their PA content and content of PA–Mv adducts (predictors) was analyzed with partial least squares discriminant analysis (PLS-DA) using “*mixOmics*” package (version 6.11.11.; Rohart et al., 2017). The data was log-transformed and standardized to zero means and unit variances to ensure that all variables are in the same dimensionless scale and that they have similar weight in the analysis. The PLS–DA results were validated using 15-fold cross-validation, which was repeated 50 times. In this process, the red wines were randomly divided to 15 parts, 14 of which are used to train a PLS-DA model, and one was used as validation set to test the model. The whole process was repeated 50 times. Optimal number of latent variables for the final model was chosen based on the average error rates in the cross-validation. Additionally, average area under curve (AUC) of receiver operating characteristics plots (ROC) was calculated from the cross-validation results and from the whole wine set to supplement the error rates in evaluating the performance of the PLS-DA model.

The qualitative comparisons of the detected compound groups were done by visually inspecting 2D quantile fingerprints, which were formed using the protocol presented by Laitila (Laitila, 2021). The protocol enables visualization and summarization of chromatographic information from multiple samples in only one quantile fingerprint. A quantile fingerprint consisted of a median fingerprint and 10th and 90th percentiles around the median, which were used to visualize the variation in the chromatographic data in the selected sample. Briefly, the key steps in calculating the quantile fingerprints were selecting the sample, normalizing the individual fingerprints to highest intensity and aligning them to account for variation in retention times, calculating the median and the 10th and 90th percentiles and plotting the data. The protocol is described in detail in Laitila (2021) where the steps of the protocol are demonstrated visually as well. To compare the qualitative polyphenolic composition of the k-mean clusters, 2D quantile fingerprints of the detected compound groups were calculated from 15 randomly sampled wines from each of the five clusters. The wines in cluster **V** were primarily used to study evolution of the compound groups. The 2D quantile fingerprints were formed from ten randomly sampled 1-year-old wines and all ten 3-year-old and 5–7 years old wines.

3. Results and discussion

3.1. Clustering of the red wines by their quantitative chemical composition

The same wine set that was now used has been previously used in two prior publications (Laitila, 2021; Laitila & Salminen, 2020). The complete wine set (n = 317) was heterogeneous and it contained wines from many different countries and regions, and numerous different grape varieties were used in their making. To reduce the complexity of the wine set, only such wines were selected for the first analyzes of this study that had at least eight replicates based on the principal grape variety (n = 234, refer to Table S1 for a summary of wines). These wine varieties were Pinot Noir, Shiraz, Merlot, Cabernet Sauvignon, Touriga Nacional, Touriga Franca, Blaufrankisch and Tempranillo.

An unsupervised principal component analysis followed by k-means clustering was done to group the wines, i.e., cluster them objectively based on their quantitative chemical composition. This approach was adopted as there was no reason to make any assumptions on how to group the wines beforehand. For instance, it could not be assumed that wines made from the same grape varieties but in different regions would be chemically identical, or that wines made from different grapes would have chemical differences between them. The first four principal components had eigenvalues higher than one (i.e., they contained most of the information about the original variables) and they were used in the k-means clustering. The four principal components explained 87 % of the variance in the original variables. K-means clustering algorithm was run several times to produce multiple clustering solutions (k = 3, 4, 5, 6), and eventually the wines were divided into five clusters (Fig. 2). Five clusters provided a satisfactory compromise between the number of clusters and the complexity of individual clusters. Table 1 summarizes which types of wines were appointed to which cluster.

The Pinot Noir wines in the present wine set originated mostly from two separate regions, which were the Beaune region in France and the

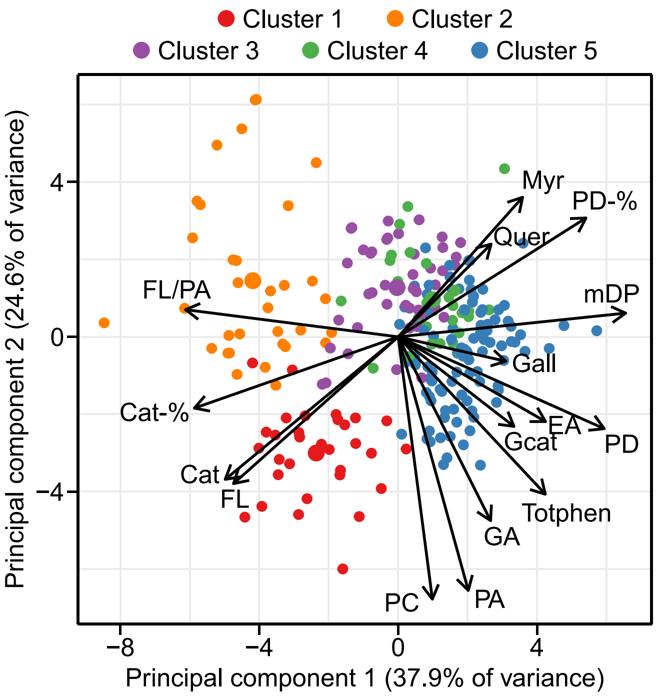


Fig. 2. Biplot of the scores of the samples and the loadings of the variables on the first two components of the principal component analysis of the red wines (n = 234). The wines were assigned to five different clusters with k-means clustering algorithm using the principal component scores of the first four principal components. Abbreviations: Cat, catechins; Cat-%, proportion of catechins of the total concentration of flavan-3-ol monomers; EA, ellagic acid; FL, flavan-3-ol monomer; FL/PA, ratio of concentrations of flavan-3-ol monomers and proanthocyanidins; GA, gallic acid; Gall, galloyl derivatives; Gcat, gallocatechins; mDP, mean degree of polymerization; Myr, myricetin derivatives; PA, proanthocyanidins; PC, procyanidins; PD prodelphinidins; PD-%; proportion of prodelphinidins out of proanthocyanidins; Quer, quercetin derivatives; Totphen, total polyphenols.

Pfalz region in Germany. Nearly all wines from the Beaune region (96 %) were in cluster **I**, whereas nearly all wines from the Pfalz region (89 %) were in cluster **II** (Table 1). Correspondingly, the wines from these regions were the predominant wines in their own clusters. The wines in cluster **I** were nearly exclusively Pinot Noir wines while in cluster **II** a fifth of the wines were blends made of Blaufrankisch, Zweigelt and St. Laurent grapes. Excluding the Blaufrankisch blends, most wines in clusters **I** and **II** were single-varietal wines as the Pinot Noir wines did not contain any additional grape varieties. A third relatively homogenous cluster was the cluster **IV**, which was mostly characterized by the large abundance of wines from the Douro region (Table 1). The whole wine set contained two different types of wines from the Douro region. First, there were the wines that were blends of Touriga Nacional, Touriga Franca and Tinta Roriz grapes, and then there were the blends made from Touriga Franca and Tinta Amarela grapes. All wines made of the Touriga Franca and Tinta Amarela grapes were in cluster **IV**, whereas the wines made from Touriga Nacional, Touriga Franca and Tinta Roriz grapes were divided between cluster **IV** and **V** with a small majority being in the cluster **IV** (55 %).

The clusters **III** and **V** were more heterogeneous than the other three clusters, but there were still observable trends about the distribution of the wines in them (Table 1). Most wines in cluster **III** were made of Shiraz and Merlot grapes, and roughly 75 % of the wines were single-varietal wines. Third substantial group of wines in cluster **III** were the Blaufrankisch blends, majority of which were classified to cluster **II**. Interestingly, all the Merlot wines in cluster **III** were made in the USA (n = 12) in the regions of Washington or California (Table 1). These wines were single-varietal wines. The Shiraz wines in cluster **III** originated mainly from two different regions in Australia (n = 19), i.e., from Barossa Valley and from South Eastern Australia. Majority of the wines from Barossa Valley were in cluster **III** (67 %) and the minority were in cluster **IV** (20 %), whereas roughly half of the wines from South Eastern Australia were cluster **III** (42 %), while the other half were in cluster **V** (53 %). That is to say, the clustering results showed how the Shiraz wines of the two different regions had chemical differences, since the wines were partially divided into different clusters. The cluster **V** was the biggest of the five clusters as it contained almost 40 % of all wine samples (Table 1). A major characteristic of cluster **V** was that it contained most of the Cabernet Sauvignon wines (87 %), and most of the wines that contained Cabernet Sauvignon as secondary grape as well (86 %, Table 1). Another major characteristic of cluster **V** was that it contained 64 % (n = 25) of the Merlot wines, and of these wines a big proportion (n = 22) was made in France. The French Merlot wines were mainly made in Listrac-Medoc region (n = 17) and they contained Cabernet Sauvignon as secondary grape variety. Moreover, all Tempranillo wines were cluster **V**, 45 % of Touriga Nacional blends were in cluster **V** and 40 % of Shiraz wines were in cluster **V** (Table 1).

3.2. *Quantitative chemical composition of the wines and the clusters*

The whole wine set that was used in the clustering was first analyzed by calculating interquartile ranges of the measured features, i.e., the ranges which contain the middle 50 % of observations. The total polyphenolic concentration was typically 673.0–975.0 mg/L and the PAs were the main polyphenols in terms of concentration (332.6–553.8 mg/L, Table S3). The PA concentrations were similar to the levels reported by Harbertson et al. in their large screening study of over 1000 red wines, as they reported a mean concentration of 544 mg/L and standard deviation of 293 mg/L for the whole wine set (Harbertson et al., 2008). The mDP of the PAs was typically 4.4–6.6 and the PD-share of PAs was 12.4 %–20.6 %. Again, the present results agreed with previous studies about quantitative PA composition of red wines (Gómez-Plaza et al., 2016; Hanlin et al., 2011), which demonstrated that PAs in red wine can be quantified at compound group level using the UPLC–MS/MS method of Engström et al. (2014). Most of the FLs were catechins (85.8 %–93.3 %) and the FL concentration was 9.4–19.8 mg/L (Table S3). The concentration of flavonols was typically low compared to the other quantified compound groups, while the gallic acid concentration was high (33.2–56.2 mg/L, Table S3). Typically, the quantitative composition of polyphenols differed markedly between the clusters, which is next discussed in detail.

The clusters **I** and **II** that were mainly characterized by the high proportions of Pinot Noir wines from the Beaune and Pfalz regions, respectively, had interesting quantitative differences between them (Fig. 3). Both clusters were rich in FLs as their concentrations were over two-fold higher compared to the other clusters (Fig. 3I), and clusters **I** and **II** had the highest catechin-% of FLs of all clusters (Fig. 3J). The concentration of PCs, and PAs altogether, were much higher in cluster **I** than in cluster **II,** or in any other cluster (Fig. 3L, N). However, the PD-% and mDP of PAs were still relatively similar between the cluster **I** and **II**, and they were the lowest out of all clusters. Finally, the cluster **I** had the highest gallic acid concentration of all clusters, whereas the cluster **II** had the lowest, and the total phenolic concentration in cluster **II** was the lowest of all clusters as well. Considering that the only known separating factor between the Pinot Noir wines in these clusters were their origins (region and country), the chemical differences between clusters **I** and **II** were considerable, especially regarding the PA concentrations.

Cluster **III** was best characterized by having the highest concentrations of both the quercetin and myricetin derivatives (Fig. 3E, F) and galloyl derivatives (Fig. 3B). The cluster **III** also had the second lowest concentration of total polyphenols (Fig. 3A). The Douro region wines (cluster **IV**) did not stand out with the highest concentration of any compound group, but they did have the second highest mDP of PAs (Fig. 3O). The Douro region wines had a notably low concentration of

Table 1

Summary of the wines in the five k-means clusters (Fig. 1, n = 234). The number n in parantheses tells the absolute number of wines while the percentage tells how large proportion of all wines made from the given grape varieties or wines originating from the given countries or regions were in each cluster. For instance, there were 31 wines in cluster **I**, 27 of which were made from Pinot Noir grapes, which was 52 % of all wines made from Pinot Noir grapes. 161 wines did not contain tertiary or further grape varieties and the most frequent tertiary grape varieties were Tinta Roriz (n = 27, used in Touriga Ciol wines from Douro region) and Petit Verdot (n = 20, used in Merlot wines from Listrac-Medoc region).

	Principal grape variety		Secondary grape variety		Country		Region	
Cluster **I** *n* = 31	Pinot Noir	(n = 27, 52 %)	None	(n = 27, 21 %)	France	(n = 26, 40 %)	Beaune	(n = 23, 96 %)
	Blaufrankisch	(n = 2, 15 %)	Zweigelt	(n = 2, 15 %)	Austria	(n = 2, 15 %)	Burgenland	(n = 2, 15 %)
	Cabernet Sauvignon	(n = 1, 3 %)	Cabernet Sauvignon	(n = 1, 5 %)	Argentina	(n = 1, 13 %)	Languedoc-Roussillon	(n = 2, 18 %)
	Merlot	(n = 1, 3 %)	Shiraz	(n = 1, 17 %)	Australia	(n = 1, 3 %)	Listrac-Medoc	(n = 1, 6 %)
					Spain	(n = 1, 8 %)	Others	(n = 3)
Cluster **II** *n* = 33	Pinot Noir	(n = 23, 44 %)	None	(n = 24, 19 %)	Germany	(n = 16, 84 %)	Pfalz	(n = 16, 89 %)
	Blaufrankisch	(n = 6, 46 %)	Zweigelt	(n = 6, 46 %)	Austria	(n = 6, 46 %)	Burgenland	(n = 6, 46 %)
	Shiraz	(n = 2, 4 %)	Greche	(n = 1, 50 %)	France	(n = 5, 8 %)	Barossa Valley	(n = 1, 7 %)
	Cabernet Sauvignon	(n = 1, 3 %)	Merlot	(n = 1, 9 %)	Argentina	(n = 2, 25 %)	Beaune	(n = 1, 4 %)
	Merlot	(n = 1, 3 %)	Others	(n = 1)	New Zealand	(n = 2, 100 %)	Other	(n = 8)
					Other	(n = 2)		
Cluster **III** *n* = 43	Shiraz	(n = 23, 48 %)	None	(n = 32, 25 %)	Australia	(n = 19, 48 %)	Barossa Valley	(n = 10, 67 %)
	Merlot	(n = 12, 31 %)	Zweigelt	(n = 5, 38 %)	USA	(n = 12, 100 %)	Washington	(n = 10, 100 %)
	Blaufrankisch	(n = 5,38 %)	Cabernet Sauvignon	(n = 2, 9 %)	Austria	(n = 5, 38 %)	South Eastern Australia	(n = 8, 42 %)
	Pinot Noir	(n = 2, 4 %)	Zinfandel	(n = 2, 100 %)	South Africa	(n = 2, 29 %)	Burgenland	(n = 5, 38 %)
	Cabernet Sauvignon	(n = 1, 3 %)	Other	(n = 2)	Germany	(n = 2, 11 %)	California	(n = 2, 100 %)
					Other	(n = 3)	Pfalz	(n = 2, 11 %)
							Others	(n = 6)
Cluster **IV** *n* = 32	Touriga Ciol	(n = 16, 55 %)	Touriga Franca	(n = 16, 55 %)	Portugal	(n = 27, 64 %)	Douro	(n = 27, 69 %)
	Touriga Franca	(n = 11, 100 %)	Tinta Amarela	(n = 11, 100 %)	Australia	(n = 4, 10 %)	Barossa Valley	(n = 4, 20 %)
	Shiraz	(n = 4, 8 %)	None	(n = 4, 3 %)	South Africa	(n = 1, 14 %)	Other	(n = 2)
	Cabernet Sauvignon	(n = 1, 3 %)	Merlot	(n = 1, 9 %)				
Cluster **V** *n* = 95	Cabernet Sauvignon	(n = 27, 87 %)	None	(n = 39, 31 %)	France	(n = 34, 52 %)	Listra-Medoc	(n = 17, 94 %)
	Merlot	(n = 25, 64 %)	Cabernet Sauvignon	(n = 19, 86 %)	Australia	(n = 15, 38 %)	Douro	(n = 12, 31 %)
	Shiraz	(n = 19, 40 %)	Touriga Franca	(n = 13, 45 %)	Portugal	(n = 14, 33 %)	South Eastern Australia	(n = 10, 53 %)
	Touriga Ciol	(n = 13, 45 %)	Merlot	(n = 9, 82 %)	Spain	(n = 10, 83 %)	Languedoc-Roussillon	(n = 8, 73 %)
	Tempranillo	(n = 11, 100 %)	Shiraz	(n = 4, 67 %)	Chile	(n = 8, 89 %)	Mendoza	(n = 5, 83 %)
			Other	(n = 11)	Argentina	(n = 5, 63 %)	Rapel Valley	(n = 4, 100 %)
					South Africa	(n = 4, 57 %)	Valdepenas	(n = 4, 100 %)
					Italy	(n = 4, 100 %)	Other	(n = 35)
					Germany	(n = 1, 5 %)		

catechins and total FLs (Fig. 3G, I), which subsequently meant that the ratio of FLs to PAs was the lowest of the five clusters (Fig. 3K). The cluster **V** contained nearly twice the amount of PDs compared to the other clusters, and the cluster had the second highest amount of PAs (Fig. 3L, M). The PD-% and mDP of PAs were the highest in cluster **V** (Fig. 3O, P). The gallocatechin concentrations did not have meaningful differences between clusters **I**–**IV** but in cluster **V** their concentration was significantly higher.

3.3. *Qualitative chemical composition of the clusters*

The utilized UPLC–MS/MS method of Engström et al. (2014) provides information about the concentrations at compound group level, but that is only one aspect about the composition of the detected polyphenols. The other aspect, and the novelty of the method in analysis of red wine, is the qualitative composition imprinted in the 2D chromatographic fingerprints. There are usually a few different approaches in detecting and quantifying PAs in wines. One is to utilize targeted or untargeted MS-based approaches to detect individual oligomers (Teixeira et al., 2016). Second is to use methods which provide the overall concentration of PAs, such as precipitation of PAs by bovine serum albumin (Harbertson et al., 2008). Third, is to use thiolysis or phloroglucinolysis, which provide information about the sub-unit composition of PAs along with the total concentration (Hanlin et al., 2011; Kennedy & Jones, 2001).These methods provide useful information about PAs, but information about the initial compound composition is either lost during the analysis, e.g., during depolymerization or it is not measured completely in the first place (detection of individual oligomers). Another advantage of the group-specific UPLC–MS/MS methodology is that it is not necessary to know beforehand what compounds are present in each detected compound group or red wine. Individual compounds can be characterized after the group-specific analysis to identify the main compounds in the 2D chromatographic fingerprints. In this study, the compound characterization was done directly from the red wines with high-resolution UPLC–Orbitrap–MS (Table S20). Finally, the 2D chromatographic fingerprints could be presented using the data-visualization and summarizing protocol reported by Laitila (2021). This method forms 2D quantile chromatographic fingerprints from dozens of individual samples, which enabled comparison of the qualitative composition of the clusters in an intuitive and robust way.

The variation of the PC composition within the clusters (within-group variation) was relatively minor, which could be seen from the narrow ranges between the 10th and 90th percentiles around the median fingerprints (the grey areas; Fig. 4). Additionally, the within-group variation seemed to be relatively symmetrical around the median throughout the fingerprints, i.e., there were no strong outlier wines in the selected samples. Only in cluster **II** there was more within-cluster variation than in other clusters (Fig. 4B). The quantile PC fingerprints, and subsequent characterization by Orbitrap-MS, showed that there

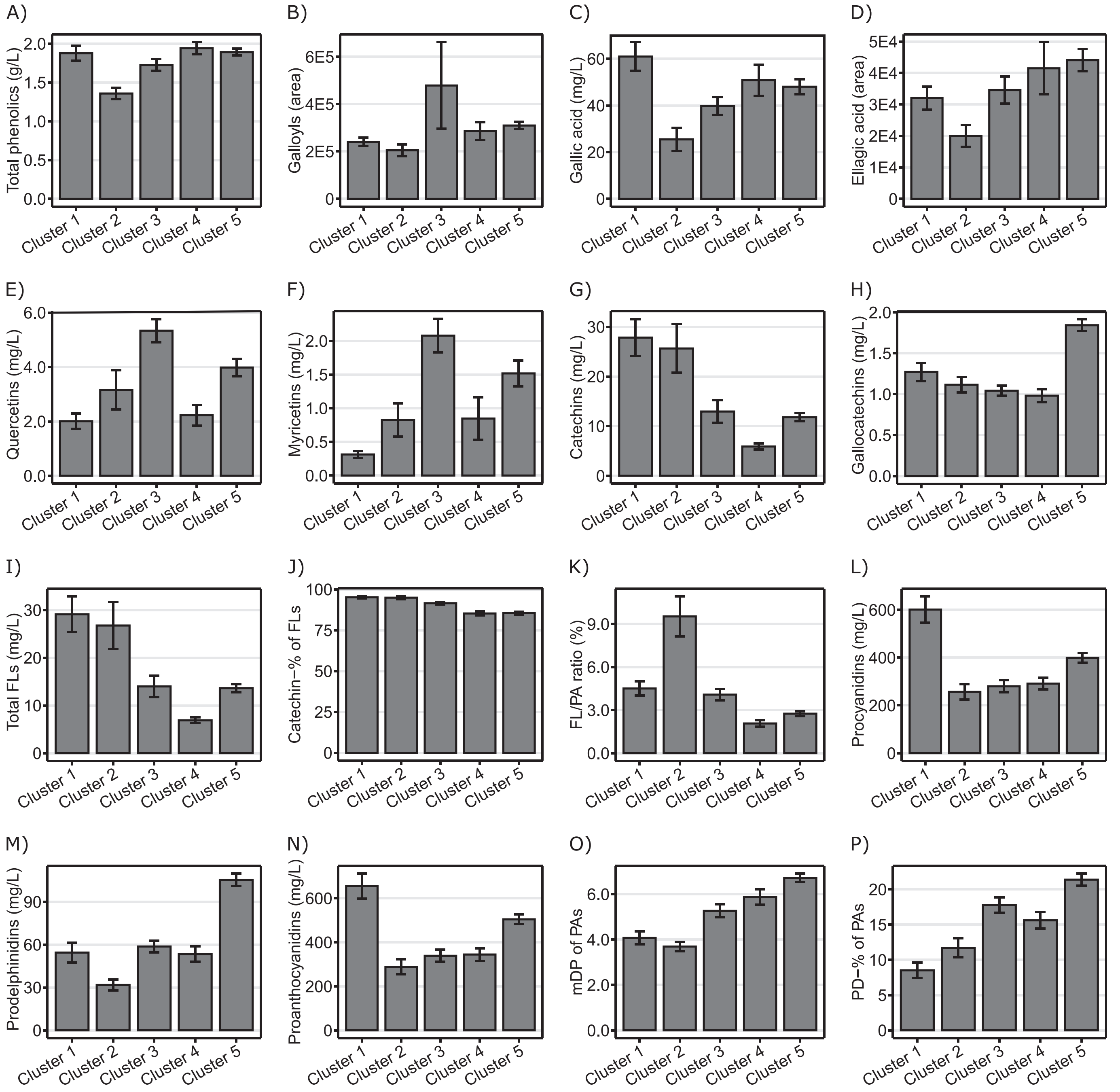


Fig. 3. Barplots of the quantitative variables measured from the red wines (n = 234). The samples were clustered into five clusters with k-means clustering algorithm using the principal component scores of the first four principal components (Fig. 2). Error bars represent 95 % confidence intervals. Tukey's HSD tests were used to test for statistically significant differences between the clusters, and the p-values are reported in Supplementary Tables S4–S19. Abbreviations: FL; flavan-3-ol monomer; PA, proanthocyanidins; PD, prodelphinidins; PD-%, proportions of PDs out of total PA concentration; mDP, mean degree of polymerization.

were two PC–PC dimers, with roughly similar abundances, that were the most abundant individual PC oligomers in every cluster (Fig. 4A–E, Table S20). The quantile PC fingerprint of cluster **II** stood out the most from the five clusters as the PC hump was not as dominant feature as in the other clusters. On the other hand, the PC fingerprints of clusters **I** and **III**–**V** were remarkably similar between one another, even though very subtle differences could be observed. For instance, the tailing parts of the PC humps in clusters **IV** and **V** were more prevalent than in cluster **I** and **III**, which could be seen from the intensities at five- and six-minute regions, for instance. These differences, however, were modest compared to the quantitative differences between the clusters (chapter 3.2., Fig. 3). For instance, the PA concentration of cluster **I** was approximately two-fold compared to cluster **III** and **IV** despite similar qualitative compositions. Therefore, the same conclusion could be made about the PC composition of the red wines in this study as was previously made of the compositions of the oligomeric PA–Mv adducts in young red wines (Laitila, 2021). That is, the differences in concentrations between wine types were typically larger than the qualitative differences.

Laitila (2021) reported previously that the fingerprints of PA–Mv$^+$ and PA–Mv–FL adducts had clear main compounds in their 2D fingerprints (dimers of catechin and malvidin glycosides), but these peaks could not be distinguished by their retention times from the 2D fingerprints of PCs measured in this study. The general shape of the 2D fingerprints of the PA–Mv adducts (Laitila, 2021) could not be

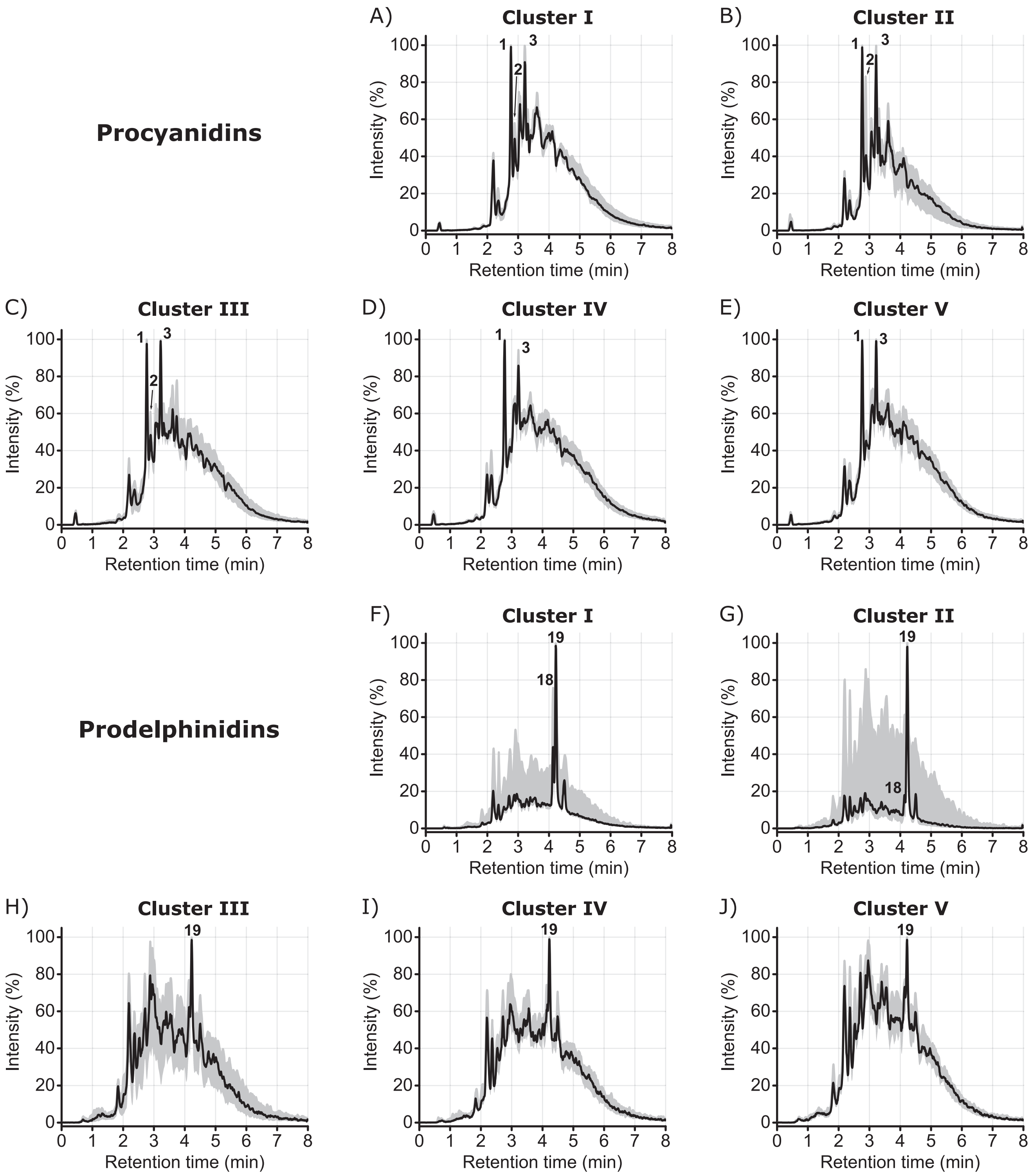


Fig. 4. Two-dimensional (2D) quantile fingerprints of procyanidin (PC) and prodelphinidin (PD) units. The black lines represent the median fingerprints, and the grey areas are the ranges between the 10th and 90th percentiles. In other words, the grey areas represent within-group variation. 15 wines were randomly sampled from each cluster to form the 2D quantile fingerprints. Altogether, both PC (A–E) and PD (F–J) quantile fingerprints summarize chromatographic information from 450 individual chromatograms because the extension and terminal units of PCs and PDs were detected separately as different fragments and by using three different cone voltages. Refer to Tables S20 for the characterization of the compounds.

distinguished from the 2D fingerprints of the PCs either. This result did not necessarily mean that the PA–Mv adducts would not be detected by the Engström method. Most likely, the PA–Mv adducts were concealed by the complexity of the PC fingerprints, which emphasized other and more abundant PC containing oligomers and polymers, such as the naturally occurring PCs themselves.

The comparison of PD composition between the clusters was affected by taxifolin (dihydroquercetin) derivatives (Table S20), which were detected by the PD-specific method as false positives (Fig. 4F–J). This was most likely caused by the singly-charged molecular ion of taxifolin aglycone having the same *m/z* value as the PDs that are released as extension units (*m/z* 303). Moreover, taxifolin aglycone and the extension units of PDs are structurally alike as well. For the quantification of PDs, the areas of the false positive signals were subtracted from the PD humps during integration. The quantification results were not affected by the false positive signals, and the concentrations of PAs, PDs and the PD-% of PAs agreed with the literature data, as was discussed earlier (Gómez-Plaza et al., 2016; Hanlin et al., 2011). Because the most abundant individual compounds in the PD fingerprints were not actual PDs, the comparison of the within-group variation was challenging, as the variation in the quantile fingerprints was affected by both variation in the PD and taxifolin compositions. For this reason, the following observations were based on the fingerprints, which were scaled to emphasize the median fingerprints (Fig. S2). Again, the PD hump of the cluster **II** was not as dominant feature as in the other cluster, and especially in the tailing end of the PD fingerprint (Fig. 4). However, the median PD fingerprints were overall dominated by the chromatographic humps and there were not as clear main compounds as the PC–PC dimers in the PC fingerprints (Fig. S2). Just like the PC fingerprints, the PD fingerprints were highly alike between the different clusters, implying of similar qualitative PD composition between the clusters.

There were no compositional differences in the 2D quantile fingerprints of catechins and gallocatechins, and the ratios between the epimers were similar in both fingerprints and in each cluster (Figs. S3, S4). The HHDP fingerprint consisted practically only of ellagic acid, and the HHDP fingerprints were similar in every cluster (Fig. S7). On the other hand, the 2D quantile fingerprints of galloyl derivatives showed interesting features (Fig. S8). The main galloyl derivatives were gallic acid and ethyl gallate (Fig. S8, Table S20), but the baseline was slightly elevated as well. This was most likely caused by galloylated PAs, which are known to comprise approximately few percent of the PAs in red wine (Gómez-Plaza et al., 2016; Hanlin et al., 2011). The biggest differences in the monomeric polyphenol composition between the clusters were observed with quercetin and myricetin derivatives (Figs. S5, S6). The composition of myricetin derivatives was the most complex in the cluster **I** (Fig. S5), while the concentration of myricetin derivatives was simultaneously the lowest out of all clusters (Fig. 3). Myricetin glucoside was the main individual myricetin derivative in each cluster, while myricetin aglycone and laricitrin aglycone (methylated myricetin aglycone) were the present in lower abundance (Table S2). The main compounds in the quercetin fingerprints were quercetin aglycone, quercetin glucuronide and isorhamnetin aglycone (methylated quercetin aglycone; Table S20). These compounds were present in clusters **III**–**V** in roughly similar proportions and the composition of quercetin derivatives was similar in the three clusters. The cluster **II** contained substantially less quercetin aglycone than the other clusters, while the abundance of quercetin aglycone was the highest in the cluster **I**. Overall, the quantile 2D chromatographic fingerprints of the monomeric polyphenol groups were similar between most clusters, and the differences in concentrations were, again, larger than the differences in qualitative composition. Mainly, the composition of quercetin and myricetin derivatives in clusters **I** and **II** differed from the other clusters, but even in this case the differences in qualitative composition were less substantial than the difference in the concentrations.

3.4. Evolution of the polyphenol composition

The evolution of the polyphenolic groups was studied by analyzing the wines in cluster **V**. With 95 wines, the cluster **V** was the biggest cluster and the ages of 73 wines were known. There were 26 one-year-old wines, 15 two-year-old wines, 10 three-year-old wines, 12 four-year-old and 10 wines with ages between five and seven years. All these wines were used in the comparison of the quantitative evolutionary patterns. Ten wines were randomly sampled from the 1-year-old wines and all three and 5–7 years old wines were used to display the qualitative evolution of the quantile 2D chromatographic fingerprints. The quantitative evolutionary patterns were additionally examined using the wines in clusters **III**–**V** (data not shown), but the results were similar to the ones obtained using only the wines from cluster **V**. Therefore, the following results are presented using the wines from cluster **V**.

The PD-% of the PAs did not change as the wines aged (ANOVA: $p = 0.350$) but the changes in the mDP of PAs (ANOVA: $p < 0.001$) and in the PA concentration (ANOVA: $p = 0.024$) were statistically significant (Fig. 5). However, the subsequent multiple comparison showed that the only significant differences were between the 1-year-old and 5–7 years old wines (Tables S30 and S31), and even these differences were minor (Fig. 5). In other words, despite the statistical differences, the quantitative metrics did not show any strong evolutionary trends in the concentrations or in the mDP of PAs.. In other studies, the protein precipitable tannin content has also been shown to remain stable after maceration during bottled storage for the first few years (Casassa et al., 2013) or a modest increasing trend has been reported in the PA concentration in the first two years of bottled storage (Monagas et al., 2006). Regarding the mDP of PAs, the mDP has been shown to decline in a comprehensive and homogeneous time series containing Cabernet Sauvignon wines over three decades (Chira et al., 2012). In shorter time scale of few years, the mDP has been reported to remain stable after maceration or it increases slightly (Casassa et al., 2013).

The evolutionary trends of the PA–Mv$^+$ and PA–methylmethine–Mv$^+$ adducts in red wine reported by Laitila (2021) were more noticeable than the evolutionary trends of the PAs. The concentrations of these adduct groups decreased towards older commercial wines, while the average sizes of the oligomeric adducts increased (Laitila, 2021). These trends were especially strong with the PA–methylmethine–Mv$^+$ adducts. The combined results demonstrate well how the evolution of the PAs, as measured by the Engström method, did not reflect the evolution of the more specific groups of PA containing adducts in red wine.

The evolution of the PA composition was additionally estimated by visually inspecting the 2D quantile fingerprints of the PCs and PDs (Fig. 5D–I). Subtle evolutionary patterns were observed in both PCs and PDs. The intensity of the chromatographic hump of the PCs increased in the tailing end of the chromatogram forming a small shoulder in the older wines (Fig. 5D–F). Additionally, the main compounds in the PC fingerprints, i.e., the PC–PC dimers, decreased in relative abundance, even though they remained as the main individual PC oligomers (Fig. 5D–F). Similar small changes occurred in the 2D quantile fingerprints of PDs as well (Fig. 5G–I). The changes in the 2D quantile fingerprints were subtle but they were still clear enough to be considered reliable, especially because the visual comparison was simultaneously based on the combined information from 30 commercial red wines. The evolutionary trends would have been more difficult to recognize if the comparison of the fingerprints would have been done between individual wines one at time. Overall, the previously observed evolutionary trends in composition of the PA–Mv$^+$, PA–methylmethine–Mv$^+$ and PA–Mv–FL adducts were not similar to the trends observed in the PC and PD fingerprints (Laitila, 2021). In other words, the changes in the PC and PD composition observed in this study were most likely not caused by the three above mentioned sub-groups of PA–Mv adducts, but rather by evolution of some other PA-derived compound groups, such as methylmethine-linked PAs, or by evolution of the initial PAs originating from the grapes.

The quantitative and qualitative evolution of the monomeric compound groups were analyzed in similar way (Figs. S9 and S10). No significant qualitative evolutionary patterns could be established (Fig. S10), and only the FL/PA ratio showed a meaningful and systemic trend as the ratio decreased towards older wines (Fig. S9K). This could be explained by a mechanism where the monomeric FLs are consumed in the constantly ongoing depolymerization and repolymerization reactions, which have been suggested to occur in red wines (Herderich & Smith, 2005). This would lead to increase in the PA concentration

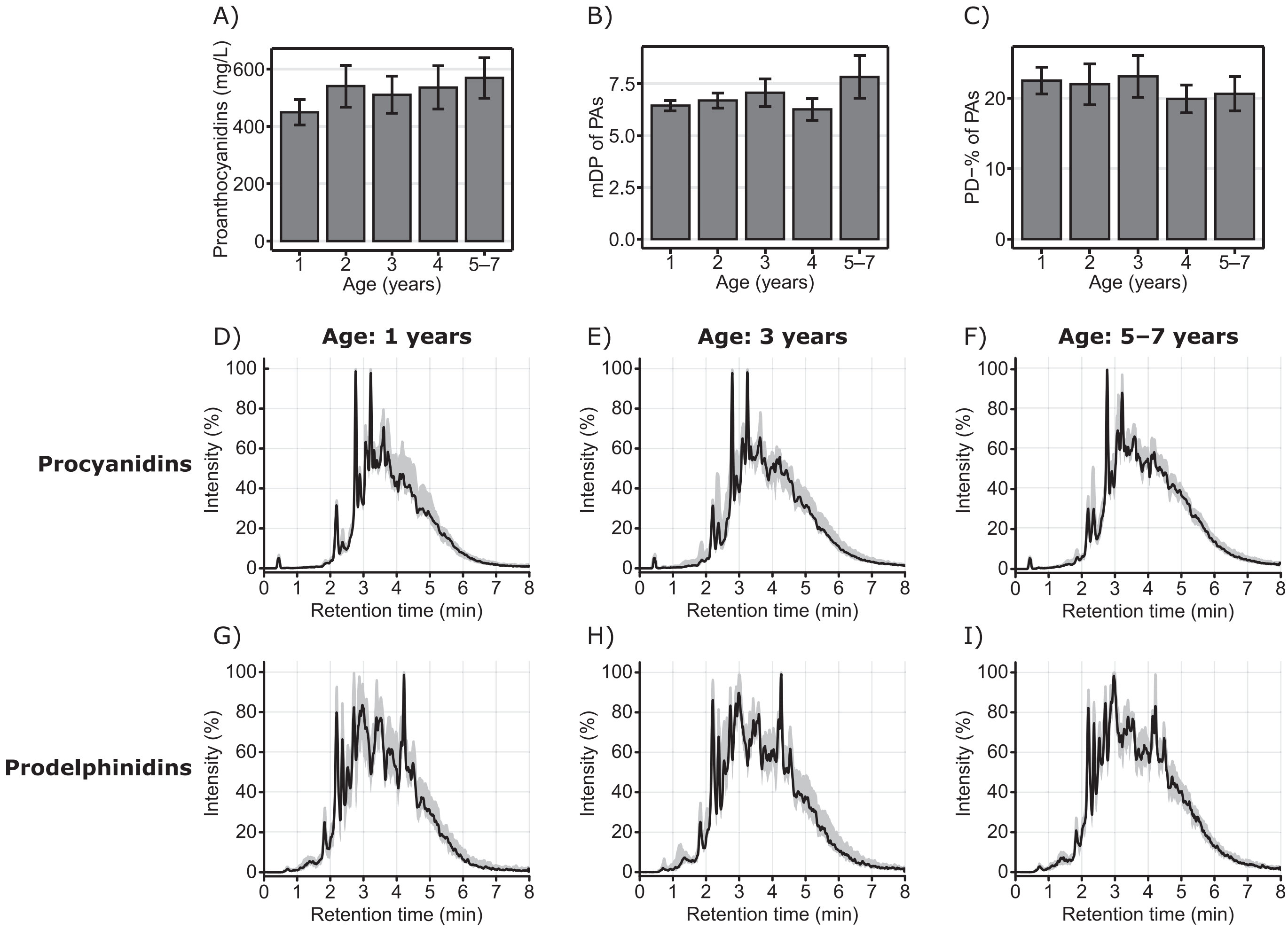


Fig. 5. Evolution of the proanthocyanidin composition in the red wines in cluster **V** (Fig. 2, Table 1). The age of 73 red wines was known out of the 95 wines in cluster **V**. There were 26 one-year-old wines, 15 two-year-old wines, 10 three-year-old wines, 12 four-year-old wines and 10 wines with ages between five and seven years. All wines with known ages were used to study the quantitative evolution (A–C, $n = 73$). There was no statistically significant differences in the PD-% of PAs based on ANOVA ($p > 0.05$), and the p-values of Tukey's HSD tests about the PA concentrations and the mDPs of the PAs are reported in Tables S31 and S32 in the supplementary data. Ten red wines were randomly sampled from the 1-year-old wines while all 3-year-old ($n = 10$) and 5–7 years old wines ($n = 10$) were used in the 2D quantile fingerprints D–I). The black lines represent the median fingerprints, and the grey areas are the ranges between 10th and 90th percentiles. In other words, the grey areas represent within-group variation. Altogether, both PC (D–F) and PD (G–I) quantile fingerprints summarize chromatographic information from 180 individual chromatograms as the extension and terminal units of PCs and PDs were detected separately as different fragments and by using three different cone voltages. Refer to Figs S9 and S10 in the supplementary data for the quantitative and qualitative evolutionary patterns of the monomeric polyphenol groups and the total phenolic content. Abbreviations: PA, proanthocyanidin; PD-%, proportion of prodelphinidins out of proanthocyanidins; mDP, mean degree of polymerization. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

relative to the monomeric FL concentration. Interestingly, the FL/PA ratio showed a stronger evolutionary trend than the individual FL and PA concentrations (Fig. S9I, K, M).

3.5. *Tannicities of the wines*

Sensorial evaluations are routinely done at Alko Inc. for red wines in their selection. The primary purpose of the sensorial evaluations is to provide information to customers and to support the salespeople in selecting right wines for their customers. In other words, the sensorial analyzes were not specifically done for this study, but the provided sensorial results could be used for research purposes. The panelists were not specifically trained for these analyzes, but specific training is not a mandatory for discriminant type evaluations (Cheynier & Sarni-Manchado, 2010). The experimental design might not have entirely followed generally preferred procedures, and this had both upsides and downsides. A downside was that the classification of any single individual wine was prone to more uncertainty compared to more tightly controlled evaluations with specific training and more panelists evaluating each wine, for instance. A benefit, on the other hand, was the size of the wine set, which enabled robust statistical analyzes with extensive cross-validation. Moreover, any discovered trends would be generalizable because they would have been discovered despite the expected uncertainty in the evaluation of the individual wines. The cross-validation ensured that any combination of the data set was equally well modelled, and that the complete model was not overfitted so that it would have poor predictive capabilities.

For the statistical modelling, the PA concentration, mDP and PD-% of PAs, and the concentrations and proportions of the large oligomeric adducts (LOA-%) of the PA–Mv$^+$, PA–methylmethine–Mv$^+$ and PA–Mv–FL adducts were used as predictors for the tannicities of the wines. All predictors have either been shown to contribute to the mouthfeel of red wines (PAs; Vidal et al., 2004) or there is a chemical reasoning to believe that they could contribute to it (PA–Mv adducts). A separate model containing the other polyphenol and phenolic acid groups as additional variables, e.g., quercetin and myricetin derivatives was also fitted (data not shown). However, this more comprehensive model had identical explanatory power compared to the more condensed model with only PA-related variables, and the same individual variables were the most important in explaining the categorization of the wines in both models. Therefore, the results will be presented using the model with only PA-related variables, as these oligomers and polymers were the main focus of this study. The LOA–% of the PA–Mv adducts has previously been used as a metric for the average size of the oligomeric PA–Mv adducts as it provides information about the proportion of the largest detectable oligomeric adducts (Laitila & Salminen, 2020). The metric does not reveal absolute information about the sizes of the adducts but it provides relative information about the differences in average sizes between wine samples. A PLS-DA model was chosen for the statistical analysis because it is well suited to cope with multicollinearity in the predictor matrix, which could affect the results and interpretations of typical generalized linear models or discriminant analyzes.

The best classification outcome in the cross-validation, and the lowest classification error rate was achieved by using maximum distance as prediction distance (Figs. 6, S11, Table S32). Three latent variables were selected for the model as they achieved the lowest classification error rate (Fig. S11, Table S32). The average error rate of the three-component model was 18.8 % (standard deviation 0.7 %) in the cross-validation, and the average balanced error rate was correspondingly 19.7 % (standard deviation 0.7 %). The average area under ROC curve was 0.861 (standard deviation 0.002). Subsequently, the average error rate for the model of the whole wine set was 17.4 %, average balanced error rate was 18.4 % and the area under ROC curve was 0.875. Perfect separation between the medium tannic and tannic wines could not be achieved but majority of the wines could be accurately classified based on the composition of the PAs and the PA–Mv adducts (Fig. 6A). The most contributing predictors were all related to the compositional features of the predictors rather than to the concentrations (Fig. 6B, C). The tannic wines got high negative scores on the first latent variable and the

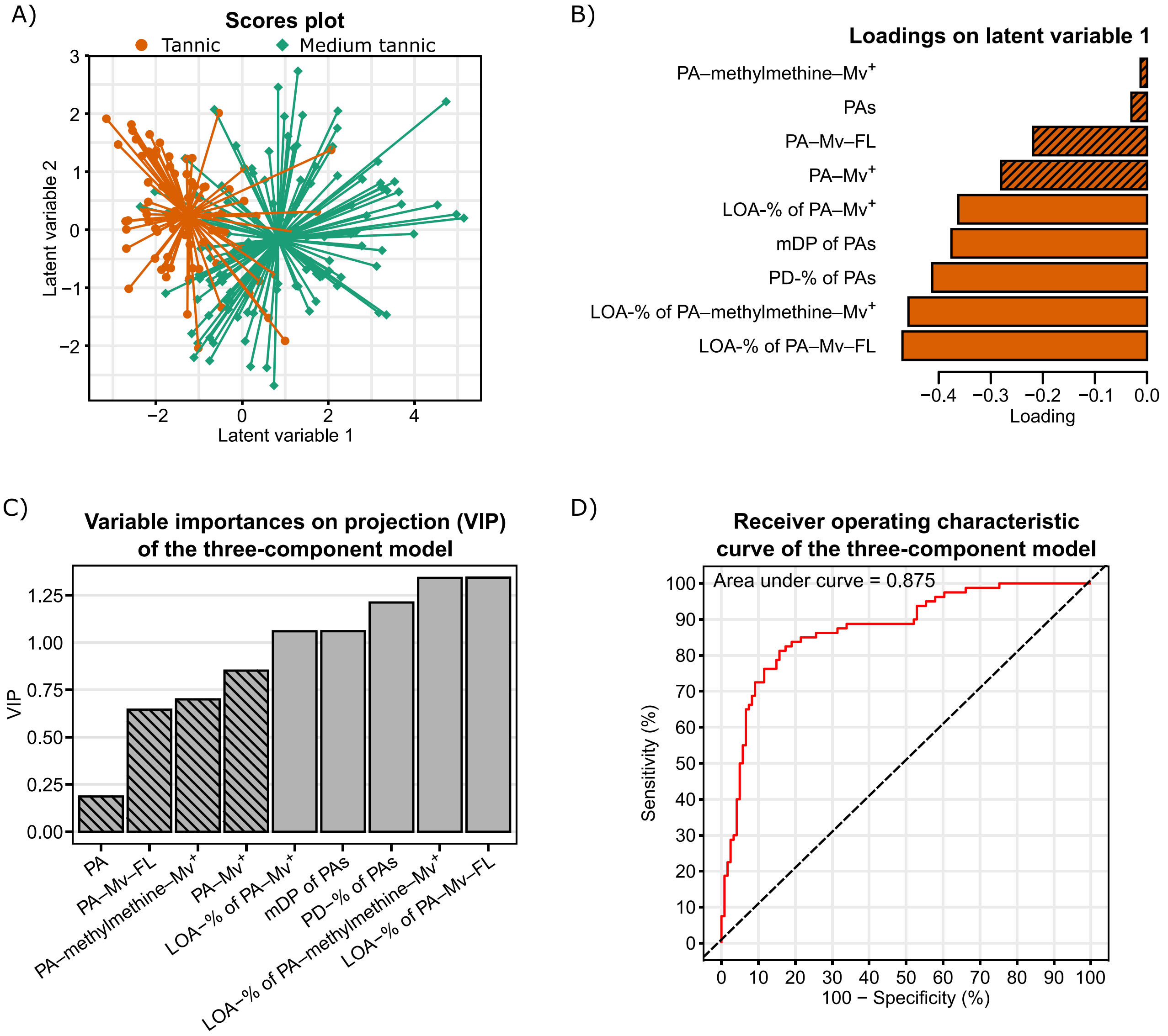


Fig. 6. Results of the partial least squares discriminant analysis (PLS-DA) modelling the connection between the proanthocyanidins (PA) and oligomeric proanthocyanidin–malvidin glycoside adducts (PA–Mv; predictors), and the perceived tannin content (tannicity; response) in 201 commercial red wines. Panel A shows the scores plot of the wine samples on the first two latent variables, panel B shows the loadings of variables on the first latent variable, and panel C shows the variable importances on projection (VIP) of the three-component model. Panel D presents the receiver operating characteristic plot indicating the performance of the model. In panels B and C, the rasterized bars represent variables related to concentrations, and non-rasterized bars represent variables related to the composition of the compound groups. Abbreviations: FL, flavan-3-ol monomer; PA, proanthocyanidin; PC, procyanidin; PD, prodelphinidin; Mv, malvidin glycoside; mDP, mean degree of polymerization; LOA-%, proportion of large oligomeric adducts; VIP, variable importance on projection.

five variables with the strongest negative loadings on the first latent variable were the proportions of the LOAs of the PA–Mv adducts and the mDP and PD-% of PAs (Fig. 6A, B). These same variables had the highest VIP (variable importance on projection) scores in the three-component model (Fig. 6C). Correspondingly, the four least loaded variables on the first latent variable and the variables with the lowest VIP scores were the concentrations of the PA–Mv adducts and PAs (Fig. 6B and C). In other words, the PLS-DA model implied that the compositional factors were more important in discriminating tannic wines from medium tannic wines, and that wines with higher average size of PAs and PA–Mv adducts were more likely to be perceived as tannic than wines with lower average size of PAs and PA–Mv adducts. The imperfect classification of the wines could be related to uncertainty in the evaluations of the individual wines due to low number of panelists, but equally probable reason is that the PLS-DA model was lacking information from other substances that contribute to astringency, such as organic acids and ethanol content (Cheynier & Sarni-Manchado, 2010).

The polyphenolic composition of red wine and its linkage to astringency has been widely studied (Casassa et al., 2013; Chira et al., 2012; Ferrer-Gallego et al., 2015; García-Estévez et al., 2017; Vidal et al., 2004). A commonly accepted idea is that the mDP of PAs is a key feature causing astringency in red wines, and this result was recreated in the present study, which gave some validation for the chosen methodology. The importance of the mDP of PAs has been established both in sensorial evaluations with trained panelists and in *in vitro* measurements, which have been done using human salivary proline rich proteins (PRP) and isolated compounds or purified PA fractions (Chira et al., 2012; De Freitas & Mateus, 2002; García-Estévez et al., 2018; Vidal et al., 2004). The effect of hydroxylation of the B-ring, i.e., the effect of the PD-% of PAs, has not been as clearly established. It has been shown using an oligomer series from monomers to trimers that the PDs induce different type of oral sensations than PCs (Ferrer-Gallego et al., 2015). In this study, the contribution of the PD-% on the perceived tannicity could not be distinguished from the contribution of the mDP because the two variables were strongly correlated ($r = 0.78$). The PA concentration itself was not an important factor in separating the two tannicity groups and similar result has been obtained in other studies previously as well (Quijada-Morín et al., 2012; Watrelot et al., 2016). It could be that the concentration was not an important factor because of the relatively high concentration range used in this study, just like Quijada-Morín (2012) concluded in their study.

Total concentrations of oligomeric pigments, as measured by the HPLC–DAD method of Peng et al. (2002), has been previously linked to sensorial properties of Shiraz wines (Gawel et al., 2007; Ristic et al., 2010) and Merlot wines (Casassa et al., 2013). In the study by Gawel et al. (2007), the concentration of pigmented polymers was strongly correlated with total polyphenolic concentration and total PA concentration, which were in turn related to puckery sensation in the wines post expectoration. Ristic et al. (2010) found a strong linear correlation between polymeric pigments and astringency related attributes, such as grainy, coarseness and dryness (Ristic et al., 2010). However, the utilized HPLC–DAD method was relatively universal in the detection of the polymeric pigments, i.e., pigmented sub-groups could not be detected separately, meaning that contributions of specific sub-groups of oligomeric pigments could not be separately identified. Casassa et al. (2013) reported a correlation between astringency and content of large polymeric pigments but, in this case as well, the utilized analytical method was such that it cannot separate specific sub-groups of oligomeric pigments form one another. The present study is the first time that compositions of specific groups of PA–Mv adducts were linked to the perceived tannicity and mouthfeel of red wine at compound group level.

4. Conclusions

This study was the first time that the PA content of red wine was studied using group-specific UPLC–MS/MS methodology and the 2D chromatographic fingerprints that the method produced. It was reasoned that the PAs and PA-containing compounds could not be so clearly defined in red wine because of the increased variation in sub-unit composition due to the presence of PA–anthocyanin adducts and other PA containing adducts. However, the fingerprints of PCs and PDs had notable similarity between different clusters of wines, and the main compounds in the 2D fingerprints of PCs were PC–PC dimers known to exist in wines. No visible trace of the fingerprints of PA–Mv adducts or their individual compounds could be seen in the PA fingerprints. Moreover, it was shown how the quantitative and qualitative evolutionary trends of the PAs differed from the previously reported trends of the PA–Mv adducts (Laitila, 2021). This demonstrated that in red wine analytics it is preferable to use as structurally specific methods as possible to gain truly accurate information about the composition of the oligomeric and polymeric polyphenols in red wine. Finally, the sensorially evaluated tannin content was strongly linked to the content of PAs and the three groups of PA–Mv adducts. Importantly, the concentrations of these oligomers and polymers were less important in explaining the tannicity than features related to their average size (mDP of PAs and LOA-% of PA-Mv adducts) and sub-unit composition (PD-% of PAs). The results concerning the PA–Mv adducts were enabled by the group-specific methodology, and this was the first time that specific groups of PA–Mv adducts were shown to contribute to mouthfeel of red wine at compound group level.

CRediT authorship contribution statement

Juuso Erik Laitila: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration, Funding acquisition. **Juha-Pekka Salminen:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This study was supported by the University of Turku Graduate School (UTUGS) and by the Turku University Foundation (personal grant to J.E. L). Anne Koivuniemi, Milla Leppä and Jorma Kim are acknowledged for maintenance of the UPLC–MS/MS instrument. Jussi Suvanto is acknowledged for preparing calibration curves for the monomeric flavan-3-ols. Maarit Karonen kindly helped to improve an early version of this manuscript. Alko Inc. provided a red wine set consisting of 272 commercial red wines and part of this wine set was utilized in this study. Additionally, Alko Inc. did the sensorial analyzes of the red wines and Alko Inc. is gratefully acknowledged for their contribution, which greatly improved the quality and the scope of the present paper. The whole of Natural Chemistry Research Group is thanked for general help and discussion.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodres.2023.113867.

References

Casassa, L. F., Beaver, C. W., Mireles, M. S., & Harbertson, J. F. (2013). Effect of extended maceration and ethanol concentration on the extraction and evolution of phenolics, colour components and sensory attributes of Merlot wines. *Australian Journal of Grape and Wine Research, 19*(1), 25–39. https://doi.org/10.1111/ajgw.12009

Cheynier, V., Dueñas-Paton, M., Salas, E., Maury, C., Souquet, J. M., Sarni-Manchado, P., & Fulcrand, H. (2006). Structure and properties of wine pigments and tannins. *American Journal of Enology and Viticulture, 57*(3), 298–305.

Cheynier, V., & Sarni-Manchado, P. (2010). Wine taste and mouthfeel. In *Managing Wine Quality: Viticulture and Wine Quality* (pp. 29–72). Elsevier Inc. https://doi.org/10.1533/9781845699284.1.30.

Chira, K., Jourdes, M., & Teissédre, P.-L. (2012). Cabernet sauvignon red wine astringency quality control by tannin characterization and polymerization during storage. *European Food Research and Technology, 234*(2), 253–261. https://doi.org/10.1007/s00217-011-1627-1

De Freitas, V., & Mateus, N. (2002). Nephelometric study of salivary protein-tannin aggregates. *Journal of the Science of Food and Agriculture, 82*(1), 113–119. https://doi.org/10.1002/jsfa.1016

Dixon, R. A., Xie, D. Y., & Sharma, S. B. (2005). Proanthocyanidins - A final frontier in flavonoid research? *In New Phytologist (Vol., 165*(1), 9–28. https://doi.org/10.1111/j.1469-8137.2004.01217.x

Engström, M. T., Pälijärvi, M., Fryganas, C., Grabber, J. H., Mueller-Harvey, I., & Salminen, J.-P. (2014). Rapid qualitative and quantitative analyses of proanthocyanidin oligomers and polymers by UPLC-MS/MS. *Journal of Agricultural and Food Chemistry, 62*(15), 3390–3399. https://doi.org/10.1021/jf500745y

Engström, M. T., Pälijärvi, M., & Salminen, J.-P. (2015). Rapid fingerprint analysis of plant extracts for ellagitannins, gallic acid, and quinic acid derivatives and quercetin-, kaempferol- and myricetin-based flavonol glycosides by UPLC-QqQ-MS/MS. *Journal of Agricultural and Food Chemistry, 63*(16), 4068–4079. https://doi.org/10.1021/acs.jafc.5b00595

Es-Safi, N.-E., Fulcrand, H., Cheynier, V., & Moutounet, M. (1999). Studies on the acetaldehyde-induced condensation of (-)-epicatechin and malvidin 3-*O*-glucoside in a model solution system. *Journal of Agricultural and Food Chemistry, 47*(5), 2096–2102. https://doi.org/10.1021/jf9806309

Ferrer-Gallego, R., Quijada-Morín, N., Brás, N. F., Gomes, P., de Freitas, V., Rivas-Gonzalo, J. C., & Escribano-Bailón, M. T. (2015). Characterization of sensory properties of flavanols–A molecular dynamic approach. *Chemical Senses, 40*(6), 381–390. https://doi.org/10.1093/chemse/bjv018

Friedrich, W., Eberhardt, A., & Galensa, R. (2000). Investigation of proanthocyanidins by HPLC with electrospray ionization mass spectrometry. *European Food Research and Technology, 211*(1), 56–64. https://doi.org/10.1007/s002170050589

García-Estévez, I., Cruz, L., Oliveira, J., Mateus, N., de Freitas, V., & Soares, S. (2017). First evidences of interaction between pyranoanthocyanins and salivary proline-rich proteins. *Food Chemistry, 228*, 574–581. https://doi.org/10.1016/j.foodchem.2017.02.030

García-Estévez, I., Ramos-Pineda, A. M., & Escribano-Bailón, M. T. (2018). Interactions between wine phenolic compounds and human saliva in astringency perception. *Food and Function, 9*(3), 1294–1309. https://doi.org/10.1039/c7fo02030a

Gawel, R., Francis, L., & Waters, E. J. (2007). Statistical correlations between the in-mouth textural characteristics and the chemical composition of Shiraz wines. *Journal of Agricultural and Food Chemistry, 55*(7), 2683–2687. https://doi.org/10.1021/jf0633950

Gómez-Plaza, E., Olmos, O., & Bautista-Ortín, A. B. (2016). Tannin profile of different Monastrell wines and its relation to projected market prices. *Food Chemistry, 204*, 506–512. https://doi.org/10.1016/j.foodchem.2016.02.124

Hanlin, R. L., Kelm, M. A., Wilkinson, K. L., & Downey, M. O. (2011). Detailed characterization of proanthocyanidins in skin, seeds, and wine of shiraz and cabernet sauvignon wine grapes (*Vitis vinifera*). *Journal of Agricultural and Food Chemistry, 59* (24), 13265–13276. https://doi.org/10.1021/jf203466u

Harbertson, J. F., Hodgins, R. E., Thurston, L. N., Schaffer, L. J., Reid, M. S., Landon, J. L., … Adams, D. O. (2008). Variability of tannin concentration in red wines. *American Journal of Enology and Viticulture, 59*(2), 210–214. https://doi.org/10.5344/ajev.2008.59.2.210

Herderich, M. J., & Smith, P. A. (2005). Analysis of grape and wine tannins: Methods, applications and challenges. *Australian Journal Of Grape And Wine Research, 11*(2), 205–214. https://doi.org/10.1111/j.1755-0238.2005.tb00288.x

Kassambara, A., & Mundt, F. (2019). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses* (1.0.6.).

Kennedy, J. A., & Jones, G. P. (2001). Analysis of proanthocyanidin cleavage products following acid-catalysis in the presence of excess phloroglucinol. *Journal of Agricultural and Food Chemistry, 49*(4), 1740–1746. https://doi.org/10.1021/jf001030o

Laitila, J. E. (2021). Composition and evolution of oligomeric proanthocyanidin–malvidin glycoside adducts in commercial red wines. *Food Chemistry, 340*, Article 127905. https://doi.org/10.1016/j.foodchem.2020.127905

Laitila, J. E., & Salminen, J.-P. (2020). Relevance of the concentrations and sizes of oligomeric red wine pigments to the color intensity of commercial red wines. *Journal of Agricultural and Food Chemistry, 68*(11), 3576–3584. https://doi.org/10.1021/acs.jafc.9b07941

Laitila, J. E., Suvanto, J., & Salminen, J.-P. (2019). Liquid chromatography–tandem mass spectrometry reveals detailed chromatographic fingerprints of anthocyanins and anthocyanin adducts in red wine. *Food Chemistry, 294*, 138–151. https://doi.org/10.1016/j.foodchem.2019.02.136

Lee, D. F., Swinny, E. E., & Jones, G. P. (2004). NMR identification of ethyl-linked anthocyanin-flavanol pigments formed in model wine ferments. *Tetrahedron Letters, 45*(8), 1671–1674. https://doi.org/10.1016/j.tetlet.2003.12.110

Leppä, M. M., Karonen, M., Tähtinen, P., Engström, M. T., & Salminen, J.-P. (2018). Isolation of chemically well-defined semipreparative liquid chromatography fractions from complex mixtures of proanthocyanidin oligomers and polymers. *Journal of Chromatography A, 1576*, 67–79. https://doi.org/10.1016/j.chroma.2018.09.034

Longo, E., Rossetti, F., Jouin, A., Teissedre, P. L., Jourdes, M., & Boselli, E. (2019). Distribution of crown hexameric procyanidin and its tetrameric and pentameric congeners in red and white wines. *Food Chemistry, 299*(July), Article 125125. https://doi.org/10.1016/j.foodchem.2019.125125

Mateus, N., Oliveira, J., Santos Buelga, C., Silva, A. M. S., & Freitas, V. (2004). NMR structure characterization of a new vinylpyranoanthocyanin-catechin pigment (a portisin). *Tetrahedron Letters, 45*(17), 3455–3457. https://doi.org/10.1016/j.tetlet.2004.03.007

Monagas, M., Gómez-Cordovés, C., & Bartolomé, B. (2006). Evolution of the phenolic content of red wines from Vitis vinifera L. during ageing in bottle. *Food Chemistry, 95* (3), 405–412. https://doi.org/10.1016/j.foodchem.2005.01.004

Peng, Z., Iland, P. G., Oberholster, A., Sefton, M. A., & Waters, E. J. (2002). Analysis of pigmented polymers in red wine by reverse phase HPLC. *Australian Journal of Grape and Wine Research, 8*(1), 70–75. https://doi.org/10.1111/j.1755-0238.2002.tb00213.x

Quijada-Morín, N., Regueiro, J., Simal-Gándara, J., Tomás, E., Rivas-Gonzalo, J. C., & Escribano-Bailón, M. T. (2012). Relationship between the sensory-determined astringency and the flavanolic composition of red wines. *Journal of Agricultural and Food Chemistry, 60*(50), 12355–12361. https://doi.org/10.1021/jf3044346

Remy-Tanneau, S., Le Guernevé, C., Meudec, E., & Cheynier, V. (2003). Characterization of a colorless anthocyanin-flavan-3-ol dimer containing both carbon-carbon and ether interflavanoid linkages by NMR and mass spectrometry. *Journal of Agricultural and Food Chemistry, 51*(12), 3592–3597. https://doi.org/10.1021/jf021227b

Ristic, R., Bindon, K., Francis, L. I., Herderich, M. J., & Iland, P. G. (2010). Flavonoids and C13-norisoprenoids in *Vitis vinifera* L. cv. Shiraz: Relationships between grape and wine composition, wine colour and wine sensory properties. *Australian Journal of Grape and Wine Research, 16*(3), 369–388. https://doi.org/10.1111/j.1755-0238.2010.00099.x

Rohart, F., Gautier, B., Singh, A., & Lê Cao, K.-A. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLOS Computational Biology, 13*(11), e1005752.

Salas, E., Fulcrand, H., Meudec, E., & Cheynier, V. (2003). Reactions of anthocyanins and tannins in model solutions. *Journal of Agricultural and Food Chemistry, 51*(27), 7951–7961. https://doi.org/10.1021/jf0345402

Salminen, J.-P. (2018). Two-dimensional tannin fingerprints by liquid chromatography tandem mass spectrometry offer a new dimension to plant tannin analyses and help to visualize the tannin diversity in plants. *Journal of Agricultural and Food Chemistry, 66*(35), 9162–9171. https://doi.org/10.1021/acs.jafc.8b02115

Salminen, J.-P., & Karonen, M. (2011). Chemical ecology of tannins and other phenolics: We need a change in approach. *Functional Ecology, 25*(2), 325–338. https://doi.org/10.1111/j.1365-2435.2010.01826.x

Teixeira, N., Azevedo, J., Mateus, N., & Freitas, V. D. (2016). Proanthocyanidin screening by LC-ESI-MS of Portuguese red wines made with teinturier grapes. *Food Chemistry, 190*, 300–307. https://doi.org/10.1016/j.foodchem.2015.05.065

Vidal, S., Francis, L., Noble, A., Kwiatkowski, M., Cheynier, V., & Waters, E. (2004). Taste and mouth-feel properties of different types of tannin-like polyphenolic compounds and anthocyanins in wine. *Analytica Chimica Acta, 513*(1), 57–65. https://doi.org/10.1016/j.aca.2003.10.017

Watrelot, A. A., Byrnes, N. K., Heymann, H., & Kennedy, J. A. (2016). Understanding the relationship between red wine matrix, tannin activity, and sensory properties. *Journal of Agricultural and Food Chemistry, 64*(47), 9116–9123. https://doi.org/10.1021/acs.jafc.6b03767

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. https://ggplot2.tidyverse.org.

Rstudio Team (2019). RStudio: Integrated Development Environment for R (1.2.5033). Rstudio, PBC, Boston, MA. URL http://www.rstudio.com/.

R Core Team (2020). R: A language and environment for statistical computing (3.6.3). R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.