https://doi.org/10.1016/j.aca.2023.341966

Rapid Authentication of Red Wine by MALDI-MS Combined With DART-MS

Xuewei Lin [a, b, c], Hao Wu [d], Gefei Huang [a, b, c], Qian Wu [a, b, c], Zhong-Ping Yao [a, b, c, *]

[a] State Key Laboratory of Chemical Biology and Drug Discovery, and Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong Special Administrative Region, China

[b] Research Institute for Future Food, and Research Center for Chinese Medicine Innovation, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong Special Administrative Region, China

[c] State Key Laboratory of Chinese Medicine and Molecular Pharmacology (Incubation), and Shenzhen Key Laboratory of Food Biological Safety Control, Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen 518057, China

[d] Key Laboratory of the Ministry of Education for Coastal and Wetland Ecosystems, College of the Environment and Ecology, Xiamen University, Fujian 361102, China

*To whom correspondence should be addressed:

Zhong-Ping Yao
Department of Applied Biology and Chemical Technology
The Hong Kong Polytechnic University
Hung Hom, Kowloon
Hong Kong
Tel: (852) 34008792. Fax: (852)23649932.
E-mail: zhongping.yao@polyu.edu.hk

Abstract

A simple, rapid and high-throughput approach was developed for authentication of red wine for the first time, by combining spectral results from matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) and direct analysis in real time mass spectrometry (DART-MS). By coupling with orthogonal partial least squares discrimination analysis (OPLS-DA), this approach enabled successful classification of 535 wines from 8 countries, with the correct classification rates of 100% on the calibration set and over 90% on the validation set for almost all countries, and 26 potential characteristic markers selected. Compared to one single technique, this approach allowed detection of more compound ions, and with better fitting and predictive performances. The satisfactory differentiation results of vintages and grape varieties further verified the robustness of the approach. This study demonstrated the feasibility of combining multiple mass spectrometric techniques for wine analysis, which can be extended to other fields or to combinations of other analytical techniques.

Keywords: Red wine; Classification; Non-target analysis; Direct MS techniques; Combination; Chemometrics.

1 Introduction

Red wine is one of the most popular alcoholic beverages and occupies a large market worldwide. For example, 679 million liters of wine were imported to China in 2018 (Liu & Song, 2021), and 9.05 billion Euros of bottled wine were exported from France in 2017 (Ugaglia, Cardebat, & Corsi, 2019). Due to its commercial profits, counterfeiting of wine, e.g., dilution, adulteration and mislabeling of brands or years, has long time been a severe problem (Herrero-Latorre, Barciela-Garcia, Garcia-Martin, & Pena-Crecente, 2019), though considerable efforts have been made to protect both consumers and producers (Meloni, Anderson, Deconinck, & Swinnen, 2019). An annual income loss of nearly 1.3 billion Euros because of the counterfeit alcoholic beverages and wines was reported for the EU market in 2016 (Wajsman, Arias Burgos, & Davies, 2016). Authenticity control of wine regarding its origins, vintage years, grape varieties, etc., has thus become an important issue, and establishing methods for rapid differentiation of wines from key wine-producing countries is urgently needed for globalization of the wine trade.

The chemical profile of red wine can be different as a result of the producing and storage process, and it is associated with geography, grape type and vintage age. Currently, a wide variety of techniques have been employed in the field of wine fingerprints, including nuclear magnetic resonance (NMR), ultraviolet-visible (UV-Vis) spectroscopy and Fourier transform infrared (FT-IR) spectroscopy for studying grape varieties, vintage years and regions (Geana, Ciucure, Apetrei, & Artem, 2019; Riovanto, Cynkar, Berzaghi, & Cozzolino, 2011; Son et al., 2008). Mass

spectrometry (MS) is a very sensitive and widely used technique, and has been applied for wine characterization, mainly by coupling with isotopic ratio (IR-MS) (Wu et al., 2019), gas chromatography (GC-MS) (Ziolkowska, Wasowicz, & Jelen, 2016) and liquid chromatography (LC-MS) (Zhang et al., 2020). IR-MS cannot provide the molecular profiling of wine components, and GC-MS and LC-MS typically require sample pretreatment and chromatographic separation.

Direct MS techniques, requiring neither sophisticated extraction nor separation steps, can provide rich information from numerous samples in a relatively short time. Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) is such a technique with high tolerance to impurities and has the high-throughput capacity to automatically analyze hundreds of sample spots on a MALDI plate. MALDI-MS has been extensively employed in various applications, such as compound profiling and differentiation of blended oils and herbal medicines (Lai et al., 2018; Ng et al., 2018). Direct analysis in real time mass spectrometry (DART-MS) is another direct MS technique that has been demonstrated to be very effective in analysis of drugs, environmental pollutants, residual pesticides, etc. (Wang, Zhao, Zhang, Li, & Pan, 2012; Wu, Yuan, Li, Huang, & Hu, 2020). MALDI-MS is typically operated under vacuum with the use of a matrix, while DART-MS analyzes samples in the open air by exposing the samples to a stream of excited gas (Cody, 2009; Gross, 2014). Up to now, literatures on red wine analysis using MALDI-MS or DART-MS are still very limited, primarily focusing on some specific compounds or simple samples at a small scale (Alecu, Albu, Litescu, Eremia, & Radu, 2016; Rubert, Lacina, Fauhl-

Hassek, & Hajslova, 2014). Moreover, while both targeted and untargeted methods have been used for wine characterization, untargeted methods have superiority in discovering discriminative markers, including metabolites and secondary metabolites produced during the fermentation and aging (Liu et al., 2020).

In this study, we developed untargeted and high-throughput methods for rapid authentication of wine using MALDI-MS and DART-MS, combined with a powerful multivariate statistical modeling tool, orthogonal partial least squares discrimination analysis (OPLS-DA) (Boccard & Rutledge, 2013). To the best of our knowledge, this work presents the first approach on combining MALDI-MS and DART-MS for sample analysis. 535 wine samples from 8 key wine-producing countries have been investigated to demonstrate the feasibility of the approach for controlling wine quality in the market.

2 Material and Methods

2.1 Chemicals and wine samples

α-Cyano-4-hydroxycinnamic acid (CHCA) and trifluoroacetic acid (TFA) were from Sigma-Aldrich (St. Louis, MO, USA). HPLC-grade acetonitrile (ACN) and HPLC-grade acetone were purchased from Anaqua Chemical Supply (Houston, TX, USA) and Acros Organic (Waltham, MA, USA), respectively. Ammonium acetate was obtained from Panreac (Barcelona, Spain). The standard Tuning Mix was purchased from Agilent Technologies (Santa Clara, CA, USA).

A total of 535 certified wine samples were kindly provided by Shenzhen Customs, China (Wu et al., 2019). They were collected from 8 major wine-producing countries, with vintage years ranging from 1999 to 2014, more than 10 types of grape varieties, and a considerable number of blended wines (see Table S1 in the Supporting Information for their detailed information). The samples were divided into two groups: the calibration set (nearly 2/3 of total samples) and the validation set (nearly 1/3 of total samples). Before the analysis, the samples were stored at 4 °C.

2.2 MALDI-MS Analysis

Following our optimization results, CHCA matrix was prepared with a concentration of 20 mg mL^{-1} in 75:25 acetone: water (v/v) containing 0.2% TFA, and aliquots of 0.5 μL CHCA solution were loaded onto the MTP 384 target plate ground steel BC (Bruker, Billerica, USA), forming evenly distributed matrix layers after fast air-dried. Aliquots of 10 μL wine samples were vacuum dried, to remove water and ethanol that were found to negatively influence the uniform crystallization and the signals of analytes. The dried samples were then dissolved with 10 μL of 50:50 ACN: H_2O, and aliquots of 0.5 μL dissolved sample solution were dropped onto the matrix layer and air-dried before being introduced into the mass spectrometer. Each wine sample was prepared with 8 spots. The spectra (*m/z* 380-1000 Da) were acquired by a UltrafleXtreme MALDI-TOF-TOF mass spectrometer (Bruker, Billerica, USA) under positive and reflectron mode, with a 355 nm smartbeam-II laser. The voltages for the ion source voltage 1, ion source voltage 2, lens voltage,

reflector voltage 1 and reflector voltage 2 were set to 20.00 kV, 17.75 kV, 7.00 kV, 21.10 kV and 10.85 kV, respectively. The automatic acquisition was applied with a completely random walk to analyze hundreds of sample spots rapidly, and both the external and internal calibration was performed using the CHCA matrix. For each spectrum, eight shots were accumulated with a total of 8000 laser pulses, and a high-signal mass range of *m/z* 400-600 was used for evaluation with resolution higher than 3000. MS/MS spectra of interest peaks were acquired by the same equipment using the LIFT function. The obtained spectra were processed by flexAnalysis (Bruker, Billerica, USA), applying centroid peak detection algorithm, signal to noise over 4, peak width over 0.2 Da and TopHat baseline subtraction.

2.3 DART-MS Analysis

Aliquots of 2 μL wine with 0.2% ammonium acetate were directly spotted onto a clean mesh sampling trip without any prior purification or extraction. Each sample was loaded with 5 spots, separated by a blank spot (Fig. S1) to diminish the interferences from the nearby spots during the DART-MS analysis. The calibration was performed with the standard Tuning Mix. The mass spectra were acquired at a mass range of 100-1000 Da under positive ion mode, with a DART-SVP ion source (IonSense Inc. Saugus, MA, USA) coupling to quadrupole-time of flight mass spectrometer (Agilent Technologies, Santa Clara, USA). A VAPUR® interface was employed to hyphenate the ion source and mass spectrometer, connecting with a membrane pump to maintain low vacuum. Semi-automatic analysis of samples was carried out using a 1-D transmission module

(P/N JCL-2100-A) fixed on a holder. To obtain information-rich and intensity-high mass spectra, the settings of the system were optimized. The settings for DART ionization included the distance between the DART source outlet and the ceramic tube inlet: 9.5 mm, the gap between the ceramic tube outlet and the capillary inlet of MS: 2 mm, module moving speed: 2 mm s^{-1}, helium flow rate: 2 L min^{-1}, gas temperature: 150 °C; and the settings for mass spectrometric detection included gas flow: 3 L min^{-1}, capillary voltage of ESI+: 1000 V, fragmentor voltage: 200 V. The acquired spectra were processed by MassHunter Qualitative Analysis (Agilent Technologies, Santa Clara, USA) with a threshold of relative intensity over 2%. Each sample-loaded spot was measured several times, and the spectral results of the second measurements, which were found to offer higher abundances as well as better reproducibility in the optimization study, were used for the data analysis.

2.4 Data analysis

The procedures for establishment and optimization of the OPLS-DA models are shown in Fig. S2. A total of 611 and 192 variables were obtained from the MALDI-MS and DART-MS results, respectively. Automatic peak picking from samples was achieved using RStudio Desktop (RStudio, Inc., Boston, USA) to form datasheets with variables in columns and samples in rows. Variables from the matrix or backgrounds were removed, and the remaining data were normalized by total intensity or abundance sums. Replicates were averaged and peaks with more than 50% missing values in all groups were excluded. Finally, 180 and 100 variables were left from MALDI-MS and

DART-MS, respectively, for further data analysis.

Data pretreatment was performed to maximize between-group while minimize within-group variance for multivariate data (van den Berg, Hoefsloot, Westerhuis, Smilde, & van der Werf, 2006). In this study, all data were subjected to the third root power transformation as it led to the best results among the transformations tested, and mean-centered before performing supervised OPLS-DA in the software (Simca 16.0; Umetrics, Andover, MA, US). Reliable OPLS-DA models based on the calibration set were established, and several criteria for data screening were considered: samples with standard deviations (SDs) between +4 and −4, and variables with variable importance on projection (VIP) values higher than 0.5. Going back to step 2 was then repeated until both SDs and VIP values met the standards. The goodness of fitting ability (R^2Y) and the cross-validated predictive ability (Q^2) suggest good models when both over 0.5 and excellent models when close to 1.0 (Eriksson, Johansson, Kettaneh-Wold, & Wold, 2001). Permutation plot with 200 numbers was tested to avoid over-fitting of a model. Afterward, the calibration set was qualified to review models, and the external validation set was qualified to verify the classification ability. In addition, the prediction results were inspected by measurements of accuracy, precision, receiver-operating characteristic curve (ROC) and area under curve (AUC) (Saito & Rehmsmeier, 2015). In the final step, potential characteristic markers were picked out based on VIP value > 1.2 and p-value < 0.05, which indicated the highest discriminative potentials in the OPLS-DA score plot. The identification of characteristic markers was performed based on

the MS and MS/MS results by referring to the literatures or databases (Flamini, 2013; Rothwell et al., 2013).

3 Results and discussion

3.1 Spectral results from MALDI-MS and classification of wine origins by MALDI-MS coupled with OPLS-DA

Fig. 1a shows a representative spectrum obtained from MALDI-MS. It is worth noting that the mass range below 380 Da was excluded because the strong signal from the CHCA matrix, especially the peak at *m/z* 379, would result in a significant reduction in the relative intensity of the sample ions. The most intensive peaks such as *m/z* 447, 493, 609, 625, 720 and 755 could be observed in almost all samples, which were commonly reported anthocyanins and their derived pigments that are associated with the wine color (Carpentieri, Marino, & Amoresano, 2007; Oliveira, Alhinho da Silva, Teixeira, De Freitas, & Salas, 2015). Anthocyanin composition has been proven to be an important factor for classification of wines in terms of their regions, grape varieties and even wine-making methods (Gonzalez-San Jose, Santa-Maria, & Diez, 1990). 20 compound ions tentatively identified by MALDI-MS/MS were summarized in Table S2. Taking peak at *m/z* 609 as an example, it was assigned as malvidin-3-O-glucoside-4-vinyl phenol as the neutral loss of glucoside (162 Da) and the remaining fragment ion 447 Da that corresponded to the malvidin residue.

It is widely recognized that proof of origin is a critical issue for food authentication, not to say for valuable red wine. However, most studies were confined to priced wines from the Old World, such

as Italy, France and Spain. With the development of wines in the New World (America, Chile, China, Australia and South Africa), origin authentication of global wine including more key wine-producing regions has become necessary. Therefore, the main objective of this study is to develop an approach for rapid differentiation of wines from 8 key wine-producing countries. Initially, an OPLS-DA model was performed to interpret two-thirds of the samples as the calibration set based on the MALDI-MS results. OPLS-DA can identify the variables with the greatest discriminatory power and is suitable for recognizing differences between groups. In the OPLS-DA plots, the horizontal direction will catch variations between groups, and the differences between the groups will locate them to the left and right sides, respectively, while the vertical direction will catch variations within groups. As shown in Fig. S3a, samples were grouped into eight clusters according to countries, with some overlap observed in the center. It should be noted that wines were collected in a way that covered a wide variety of regions, vintages, grape varieties and manufacturers, all of which are known to influence the wine compositions. The complexity of the samples would narrow the differences between groups, making the spectral results similar and difficult to distinguish. However, an interesting trend in the scattering plot was that northern hemisphere countries (green circle) cluster to the left side and southern hemisphere countries (blue circle) cluster to the right side, suggesting that the difference between the two hemispheres was more pronounced than that between the eight countries. The opposite climate change might account for this, as the ripeness period of grapes is April to October and October to April in the northern hemisphere and southern hemisphere, respectively. Grape growth is largely dependent on climatic conditions, such as

temperature, rainfall and solar radiation, which can critically impact the grape ripeness and consequently the quality of wine (Jones, White, Cooper, & Storchmann, 2005).

When OPLS-DA was applied to characterize countries within the two hemispheres, those from the southern hemisphere were well separated. As for the northern hemisphere, three European countries (Spain, Italy and France, red circle) were distinguished from America and China (Fig. S3b). Since interactions between physical and biological environments lead to unique characteristics of a region, the climate has a remarkable impact on grape types and overall wine styles (Jones et al., 2005; van Leeuwen, 2022). These three European countries are located closed to each other, which are more likely to share similar climate that contributes to similar wine styles. In this case, wine classification could be accomplished step by step. As discussed above, the differences between two hemispheres were found to be more significant than those among 8 countries, and thus model M1 was firstly established to separate the two hemispheres. Two other models (M2 and M3) were carried out to discriminate countries in the southern hemisphere and northern hemisphere, respectively, with Spain, Italy and France as a whole (labeled as Europe). Model M4 was ultimately conducted to distinguish three European countries. In this case, optimal four OPLS-DA models were generated (Fig. 2a-d), and the separations became significantly improved. Notably, the R^2Y and Q^2 values of the four models were all greater than 0.66 (Table S3, M1-M4), suggesting that they were good models. The correct classification rates of the calibration set were 100% for 7 countries and 98.4% for Spain, while for the validation set, the correct

classification rates varied from 68.4% to 90.3% (Table 1).

3.2 Spectral results from DART-MS and classification of wine origins by DART-MS coupled with OPLS-DA

DART-MS has been employed in qualitatively detecting specific compounds or residual pesticides in wine (Guo et al., 2016), but its applications in authentication of wine based on non-targeted analysis have received little attention, particularly at a large scale. A typical spectrum acquired by DART-MS is shown in Fig. 1b. Although a mass range of 100-1000 Da was scanned, nearly no signals were observed over 420 Da, which was in concordance to a previous study (Guo et al., 2021). The commonly observed peaks may be related to volatile compounds that contribute to the wine sensory, which are also parameters for grouping wines (Godelmann et al., 2013). The procedures of building 4 optimal OPLS-DA models were applied to the DART-MS results as well, and very similar clustering trends could be observed (Fig. 2e-h). Slight separation along the vertical direction in Fig. 2e was observed, which might be caused by the differences in sample collection, storage and analysis. However, this intra-group difference was much smaller than the inter-group difference and did not affect the discrimination results. Compared with MALDI-MS, worse R^2Y and Q^2 values were obtained by D1, D2 and D3, but better ones were provided by D4 (Table S3). As displayed in Table 1, when the validation set was classified by the established models, correct classification rates of 64.5%-100% were obtained by DART-MS, with higher ones for the America, France and Italy samples, and lower ones for the China, Spain, Australia, Chile

and South Africa samples, as compared to the MALDI-MS results.

The above results showed that both MALDI-MS and DART-MS were applicable in distinguishing original countries of wine individually, but neither of them was robust enough for unambiguous differentiation of all the countries studied. As mentioned above, MALDI-MS and DART-MS have different ionization mechanisms and favor different types of molecules. MALDI-MS underlined a mass range from 380 Da to 1000 Da in this study, mainly for the detection of anthocyanins and their metabolites, showing limitations in analyzing smaller compounds because of the interferences from the matrix ions. While the major mass range of DART-MS is 100-420 Da in this study, and most of the detected compounds were volatiles. These indicated the complementary roles of these two analytical techniques in the mass range and detectable compounds for the analysis, suggesting that combining their spectral results might be a way to improve the classification results.

3.3 Combination of the spectral results of MALDI-MS and DART-MS and classification of the wine samples

3.3.1 Data analysis

Use of different techniques for determining different wine features (Geana et al., 2019; Wu et al., 2019) and combination of NMR and differential sensing array dataset for differentiation of genetically identical grapevines (Crook et al., 2021) have been reported. However, little attention

has been paid to combining different MS techniques to distinguish complex wine samples in a large scale. An approach of combining spectral results from MALDI-MS and DART-MS for wine analysis was for the first time proposed in this study. As described in Fig. 3, in this approach, the calibration dataset from MALDI-MS was integrated with the calibration dataset from DART-MS with equal contribution to form a new combined dataset, and then followed the procedures from step 2 to step 7 in Fig. S2 for the data analysis. As the intensities of the MALDI-MS spectra did not match with the abundances of the DART-MS spectra, in step 2, the data treatment was conducted within each dataset rather than normalizing them as a whole.

3.3.2 The improved performance of the OPLS-DA models after the combination

As shown in Fig. 2i-l, the performance of the OPLS-DA models after the combination was superior to those of the individual ones. The trend of separation was just the same as described above, but each group had a clearer cluster and less dispersion. Compared with those with the single techniques, significant increases in both R^2Y and Q^2 values were obtained by the current four models (Table S3, MD1-MD4), implying the reliability and better differentiating ability of the models, which were further verified by the results of the permutation plot with 200 numbers. The calibration set was then employed to review the new models. Notably, the correct classification rates in all countries reached 100.0% (Table 1), which was higher than those obtained by MALDI-MS or DART-MS alone. The prediction abilities of the established models on external samples were confirmed by the validation set, and the results showed that the validated samples were well

matched with the calibration ones as the correct classification rates were typically higher than 90%, with 100% for China and France. The exception was the samples from Italy, for which the model gave 68.4% of discrimination accuracy merely, and this was consistent with the previous study (Wu et al., 2019). In more details, the mis-assigned samples were primarily grouped to Spain, suggesting that there might be some similarities between the wines from two countries, and studies have shown that they had similar contents of titratable acidity and proportions of volatile compounds (Sikuten et al., 2021; Stój, Czernecki, Domagała, & Targoński, 2017). As listed in Table S4, the precision, accuracy and area under curve (AUC) values were also calculated to evaluate the four OPLS-DA models for distinguishing the wine samples. Precisions ranging from 82.4% to 100%, accuracies of all above 88.5%, and AUC values of all exceeding 0.86 were obtained, proving the practicality of the established OPLS-DA models. Overall, the wine samples from different counties could be discriminated well using the combined approach. These results demonstrated that combining the spectral results of the two complimentary techniques, MALDI-MS and DART-MS, enabled improved results for differentiation of the wine samples from different countries. For an unknown sample, MD1 can be applied first to determine the hemisphere, and MD2 or MD3 can then be applied for the determination of countries. If it is from European countries, MD4 can be further employed.

3.4 Potential characteristic markers of the eight countries

Different compositions in red wine have been proven to be related to their origins (Berna, Trowell,

Clifford, Cynkar, & Cozzolino, 2009; Green, Parr, Breitmeyer, Valentin, & Sherlock, 2011). Therefore, it is important to find out the characteristic markers that contribute most to the country discrimination. As shown in Table S5, a total of 26 marker ions were selected by applying VIP value > 1.2 together with p value < 0.05. The majority of these marker ions were of too low intensities for conducting MS/MS, and the identities of some marker ions were tentatively assigned based on their *m/z* values and relevant literatures (Table S6). These assigned compounds were typically volatiles and phenols, which are most involved in the wine sensory evaluation. In addition to polyphenol flavonoid monomers, procyanidin C1, the trimer of epicatechin, was also detected (*m/z* 867) (González-Manzano, Santos-Buelga, Pérez-Alonso, Rivas-Gonzalo, & Escribano-Bailón, 2006). Some ions might be the fragments, such as *m/z* 446 and 447 were from derivatized resveratrol and anthocyanins, respectively (De Villiers, Vanhoenacker, Majek, & Sandra, 2004; Soleas et al., 1995). The ion at *m/z* 722 could be fumonisins B_1 that widely exists in wine (Mogensen, Larsen, & Nielsen, 2010). The remaining unknown ones might be metabolites and secondary metabolites produced during the fermentation and storage, or other fragment ions. When looking into the relative intensities/abundances, some marker ions were found to be country-characteristic, such as *m/z* 144 was higher in Chinese wines, and *m/z* 116 was significantly lower in Australia wines compared with that in Chile and South Africa. The relative abundances of ions *m/z* 110 and *m/z* 194 for French wines were almost twice as high as for Italian and Spanish wines, allowing identification with improved intensities and providing a potential method for easy and improved wine authentication for future studies.

3.5 OPLS-DA models for every two countries

We also attempted to build another 28 OPLS-DA models for every two countries (Fig. S4), and it was found to allow enhanced differentiation as compared to those containing more than two countries. As shown in Table S7, every two countries were well distinguished with higher R^2 and Q^2 values. Among them, 19 models showed excellent correct classification rates of over 95.0% for the validation sets, with 12 models exhibiting 100%, which were better than that using 4 models. It is obviously more difficult to separate larger groups as different grape varieties, vintages and brands introduced complex within-group variations that complicated the analysis. Furthermore, the classification of Italian samples also gained significant improvement with total correct classification rates varying from 90.9% to 100% when discriminated with other countries (Table S7, Models S3, S9, S14, S20, S21 and S22). Still, it was 84% when distinguished from Spain (Table S7, Model S19), further suggesting that the wine samples from these two countries were similar. However, testing an unknown sample using 28 models one by one is incredibly time consuming. Consequently, as the overall predictive ability is remarkable enough, 4 models are still applied in this study. These 28 models can be used as a complement to figure out confusing samples, or to rapidly determine whether a sample belongs to country A or country B.

3.6 OPLS-DA models for other features

Origin, grape variety and vintage are known to affect the flavor of a wine. Since country is a

dominant feature for the wine samples collected in this study, the distinction among countries is easier to achieve. To demonstrate the further potential of our approach, we also tried to distinguish the vintages. The vintages of the collected wines spanned a wide range, from 1999 to 2014, while sample numbers for most years were insufficient for the collected samples. Since larger vintage differences allowed better differentiation (Gougeon, da Costa, Guyon, & Richard, 2019), in this study, instead of grouping by specific years, we roughly divided samples into three groups according to their aging time: the young group (vintage year after 2010), the aged group (vintage year between 2008-2010) and the old group (vintage year before 2008). Supervised OPLS-DA models were performed to characterize three groups within each country. Though clear separation between the young group and the old group showed the potential of the method for the distinction, their overlapping with the aged group was observed (not provided), indicating that close vintages were difficult to be discriminated. Therefore, the aged group were temporarily excluded, and the OPLS-DA models were constructed based on the young group and the old group. As show in Fig. 4, it could be discovered that the two groups exhibited good separation in most countries, with R^2 and Q^2 values over 0.5 and even close to 1.0 (Table S8), while the Spain samples showed exceptionally low performance, which needs further investigation, and not enough samples of the old group from Chile were obtained for the differentiation. The average correct classification rate for the new established OPLS-DA models was 95.7% when excluding the Spanish ones.

Discriminating the grape variety of the wine samples was also attempted by our combined

approach. The main grape varieties in this study are Cabernet Sauvignon, Shiraz, Tempranillo, Merlot, Pinot Nior, Carménère, Pinotage and Zinfandel, and other varieties were removed as the numbers were insufficient for multivariate statistical analysis. When all data sets were taken into account, no clear discrimination of wines was observed. Subsequently, three OPLS-DA models were constructed according to the trend of each variety clustered in the scattering plot, which showed distinct clustering (Fig. 5). The average correct classification rate of the 8 varieties was as good as 91.8%. These preliminary results suggested the potential of the approach for differentiation of vintages and grape varieties, which, however, still needs further validation with unknown samples. Improved performance can be also expected with larger number of representative samples for each variety or each year, and more investigations can be explored to develop classification models that cover all the features, and even for the identification of diluted, adulterated and mislabeling wines.

4 Conclusions

Authentication of commercial wine is significantly important, due to the economic and cultural values. In this study, 535 red wines with large variations were collected, and investigated with MALDI-MS and DART-MS, two direct mass spectrometry methods, which require no chromatographic separation. With the minimal sample preparation, hundreds of wine samples could be analyzed within hours. The two techniques were found to work complementarily for wine analysis, with MALDI-MS mainly for detection of anthocyanin-related analytes and DART-MS

for volatile compounds. Their combination was utilized for the first time in the classification of 535 red wine samples from 8 countries, in coupling with OPLS-DA. Combining MALDI-MS and DART-MS expanded the chemical profiles and resulted in better fitting and predictive abilities as well as higher correct classification rates. The potential characteristic markers were also tentatively selected, and the approach was attempted to distinguish other features such as vintage years and grape varieties. Considering the overall time and workload demanded, our approach is profound to be a promising alternative to traditional GC-MS or LC-MS methods for wine analysis, due to its simplicity, minimal sample preparation, rapid analysis and high throughput. The thought of combining different types of mass spectrometric techniques and even other analytical techniques can also be extended to other fields for various applications.

Acknowledgments

This research work was supported by the Hong Kong Research Grants Council (Grant Nos. 15304020, R5013-19, R4005-18 and C4002-20WF) and the National Natural Science Foundation of China (Grant No. 31701699). Thanks are given to the University Research Facility in Life Sciences and the University Research Facility in Chemical and Environmental Analysis of The Hong Kong Polytechnic University (PolyU) for the technical supports. We also thank the Food Inspection Center of Shenzhen Entry-Exit Inspection and Quarantine Bureau for providing the wine samples and Dr. Suying Li (PolyU) for the assistance in preparing this manuscript.

References

Alecu, A., Albu, C., Litescu, S. C., Eremia, S. A. V., & Radu, G. L. (2016). Phenolic and anthocyanin profile of Valea Calugareasca red wines by HPLC-PDA-MS and MALDI-TOF analysis. *Food Analytical Methods, 9* (2), 300-310.

Berna, A. Z., Trowell, S., Clifford, D., Cynkar, W., & Cozzolino, D. (2009). Geographical origin of Sauvignon Blanc wines predicted by mass spectrometry and metal oxide based electronic nose. *Analytica Chimica Acta, 648* (2), 146-152.

Boccard, J., & Rutledge, D. N. (2013). A consensus orthogonal partial least squares discriminant analysis (OPLS-DA) strategy for multiblock Omics data fusion. *Analytica Chimica Acta, 769*, 30-39.

Carpentieri, A., Marino, G., & Amoresano, A. (2007). Rapid fingerprinting of red wines by MALDI mass spectrometry. *Analytical and Bioanalytical Chemistry, 389* (3), 969-982.

Cody, R. B. (2009). Observation of molecular ions and analysis of nonpolar compounds with the direct analysis in real time ion source. *Analytical Chemistry, 81* (3), 1101-1107.

Crook, A. A., Zamora-Olivares, D., Bhinderwala, F., Woods, J., Winkler, M., Rivera, S., Shannon, C. E., Wagner, H. R., Zhuang, D. L., Lynch, J. E., Berryhill, N. R., Runnebaum, R. C., Anslyn, E. V., & Powers, R. (2021). Combination of two analytical techniques improves wine classification by Vineyard, Region, and vintage. *Food Chemistry, 354*, 129531.

De Villiers, A., Vanhoenacker, G., Majek, P., & Sandra, P. (2004). Determination of anthocyanins in wine by direct injection liquid chromatography–diode array detection–mass spectrometry and classification of wines using discriminant analysis. *Journal of Chromatography A, 1054* (1-2), 195-204.

Eriksson, L., Johansson, E., Kettaneh-Wold, N., & Wold, S. (2001). Multi-and megavariate data analysis. *Umetrics Academy, Umeå*, 43.

Flamini, R. (2013). Recent applications of mass spectrometry in the study of grape and wine polyphenols. *ISRN Spectroscopy, 2013*, 1-45.

Geana, E. I., Ciucure, C. T., Apetrei, C., & Artem, V. (2019). Application of spectroscopic UV-Vis and FT-IR screening techniques coupled with multivariate statistical analysis for red wine authentication: Varietal and vintage year discrimination. *Molecules, 24* (22), 4166.

Godelmann, R., Fang, F., Humpfer, E., Schutz, B., Bansbach, M., Schafer, H., & Spraul, M. (2013). Targeted and nontargeted wine analysis by 1H NMR spectroscopy combined with

multivariate statistical analysis. Differentiation of important parameters: grape variety, geographical origin, year of vintage. *Journal of Agricultural and Food Chemistry, 61* (23), 5610-5619.

González-Manzano, S., Santos-Buelga, C., Pérez-Alonso, J. J., Rivas-Gonzalo, J. C., & Escribano-Bailón, M. T. (2006). Characterization of the mean degree of polymerization of proanthocyanidins in red wines using liquid chromatography−mass spectrometry (LC-MS). *Journal of Agricultural and Food Chemistry, 54* (12), 4326-4332.

Gonzalez-San Jose, M. L., Santa-Maria, G., & Diez, C. (1990). Anthocyanins as parameters for differentiating wines by grape variety, wine-growing region, and wine-making methods. *Journal of Food Composition and Analysis, 3* (1), 54-66.

Gougeon, L., da Costa, G., Guyon, F., & Richard, T. (2019). 1H NMR metabolomics applied to Bordeaux red wines. *Food Chemistry, 301*, 125257.

Green, J. A., Parr, W. V., Breitmeyer, J., Valentin, D., & Sherlock, R. (2011). Sensory and chemical characterisation of Sauvignon blanc wine: Influence of source of origin. *Food Research International, 44* (9), 2788-2797.

Gross, J. H. (2014). Direct analysis in real time—a critical review on DART-MS. *Analytical and Bioanalytical Chemistry, 406* (1), 63-80.

Guo, T., Fang, P., Jiang, J., Zhang, F., Yong, W., Liu, J., & Dong, Y. (2016). Rapid screening and quantification of residual pesticides and illegal adulterants in red wine by direct analysis in real time mass spectrometry. *Journal of Chromatography A, 1471*, 27-33.

Guo, T., Li, Y., Yong, W., Fang, P., Qin, Z., Yan, A., & Dong, Y. (2021). Non-target geographic region discrimination of Cabernet Sauvignon wine by direct analysis in real time mass spectrometry with chemometrics methods. *International Journal of Mass Spectrometry, 464*, 116577.

Herrero-Latorre, C., Barciela-Garcia, J., Garcia-Martin, S., & Pena-Crecente, R. M. (2019). Detection and quantification of adulterations in aged wine using RGB digital images combined with multivariate chemometric techniques. *Food Chemistry: X, 3*, 100046.

Jones, G. V., White, M. A., Cooper, O. R., & Storchmann, K. (2005). Climate change and global wine quality. *Climatic Change, 73* (3), 319-343.

Lai, Y. H., Wu, Q., So, P. K., Mok, D. K. W., Che, C. T., & Yao, Z. P. (2018). Rapid differentiation of Schisandra sphenanthera and Schisandra chinensis by matrix-assisted laser desorption/ionization mass spectrometry. *International Journal of Mass Spectrometry, 434*,

258-263.

Liu, A., & Song, H. (2021). Analysis and forecasts of the demand for imported wine in China. *Cornell Hospitality Quarterly, 62* (3), 371-385.

Liu, F. J., Jiang, Y., Li, P., Liu, Y. D., Yao, Z. P., Xin, G. Z., & Li, H. J. (2020). Untargeted metabolomics coupled with chemometric analysis reveals species-specific steroidal alkaloids for the authentication of medicinal Fritillariae Bulbus and relevant products. *Journal of Chromatography A, 1612*, 460630.

Meloni, G., Anderson, K., Deconinck, K., & Swinnen, J. (2019). Wine Regulations. *Applied Economic Perspectives and Policy, 41* (4), 620-649.

Mogensen, J. M., Larsen, T. O., & Nielsen, K. F. (2010). Widespread occurrence of the mycotoxin fumonisin B2 in wine. *Journal of Agricultural and Food Chemistry, 58* (8), 4853-4857.

Ng, T. T., Li, S., Ng, C. C. A., So, P. K., Wong, T. F., Li, Z. Y., Chan, S. T., & Yao, Z. P. (2018). Establishment of a spectral database for classification of edible oils using matrix-assisted laser desorption/ionization mass spectrometry. *Food Chemistry, 252*, 335-342.

Oliveira, J., Alhinho da Silva, M., Teixeira, N., De Freitas, V., & Salas, E. (2015). Screening of anthocyanins and anthocyanin-derived pigments in red wine grape pomace using LC-DAD/MS and MALDI-TOF techniques. *Journal of Agricultural and Food Chemistry, 63* (35), 7636-7644.

Riovanto, R., Cynkar, W. U., Berzaghi, P., & Cozzolino, D. (2011). Discrimination between Shiraz wines from different Australian regions: the role of spectroscopy and chemometrics. *Journal of Agricultural and Food Chemistry, 59* (18), 10356-10360.

Rothwell, J. A., Perez-Jimenez, J., Neveu, V., Medina-Remon, A., M'Hiri, N., Garcia-Lobato, P., Manach, C., Knox, C., Eisner, R., Wishart, D. S., & Scalbert, A. (2013). Phenol-Explorer 3.0: a major update of the Phenol-Explorer database to incorporate data on the effects of food processing on polyphenol content. *Database (Oxford), 2013*, bat070.

Rubert, J., Lacina, O., Fauhl-Hassek, C., & Hajslova, J. (2014). Metabolic fingerprinting based on high-resolution tandem mass spectrometry: a reliable tool for wine authentication? *Analytical and Bioanalytical Chemistry, 406* (27), 6791-6803.

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One, 10* (3), e0118432.

Sikuten, I., Stambuk, P., Tomaz, I., Marchal, C., Kontic, J. K., Lacombe, T., Maletic, E., & Preiner, D. (2021). Discrimination of genetic and geographical groups of grape varieties (Vitis vinifera L.) based on their polyphenolic profiles. *Journal of Food Composition and Analysis, 102*, 104062.

Soleas, G., Goldberg, D., Diamandis, E., Karumanchiri, A., Yan, J., & Ng, E. (1995). A derivatized gas chromatographic-mass spectrometric method for the analysis of both isomers of resveratrol in juice and wine. *American journal of enology and viticulture, 46* (3), 346-352.

Son, H. S., Kim, K. M., van den Berg, F., Hwang, G. S., Park, W. M., Lee, C. H., & Hong, Y. S. (2008). 1H nuclear magnetic resonance-based metabolomic characterization of wines by grape varieties and production areas. *Journal of Agricultural and Food Chemistry, 56* (17), 8007-8016.

Stój, A., Czernecki, T., Domagała, D., & Targoński, Z. (2017). Comparative characterization of volatile profiles of French, Italian, Spanish, and Polish red wines using headspace solid-phase microextraction/gas chromatography-mass spectrometry. *International Journal of Food Properties, 20* (sup1), S830-S845.

Ugaglia, A. A., Cardebat, J.-M., & Corsi, A. (2019). *The Palgrave handbook of wine industry economics* (Vol. 25): Palgrave Macmillan.

van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K., & van der Werf, M. J. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics, 7*, 142.

van Leeuwen, C. (2022). Terroir: The effect of the physical environment on vine growth, grape ripening, and wine sensory attributes. In *Managing Wine Quality* (pp. 341-393): Woodhead Publishing.

Wajsman, N., Arias Burgos, C., & Davies, C. (2016). The economic cost of IPR infringement in spirits and wine. *European Union Intellectual Property Office: Alicante, Spain*.

Wang, L., Zhao, P., Zhang, F., Li, Y., & Pan, C. (2012). Direct analysis in real time mass spectrometry for the rapid identification of four highly hazardous pesticides in agrochemicals. *Rapid Communications in Mass Spectrometry, 26* (16), 1859-1867.

Wu, H., Tian, L., Chen, B., Jin, B., Tian, B., Xie, L., Rogers, K. M., & Lin, G. (2019). Verification of imported red wine origin into China using multi isotope and elemental analyses. *Food Chemistry, 301*, 125137.

Wu, L., Yuan, Z. C., Li, Z. M., Huang, Z., & Hu, B. (2020). In vivo solid-phase microextraction

swab sampling of environmental pollutants and drugs in human body for nano-electrospray ionization mass spectrometry analysis. *Analytica Chimica Acta, 1124*, 71-77.

Zhang, X. K., Li, S. Y., Zhao, X., Pan, Q. H., Shi, Y., & Duan, C. Q. (2020). HPLC-MS/MS-based targeted metabolomic method for profiling of malvidin derivatives in dry red wines. *Food Research International, 134*, 109226.

Ziolkowska, A., Wasowicz, E., & Jelen, H. H. (2016). Differentiation of wines according to grape variety and geographical origin based on volatiles profiling using SPME-MS and SPME-GC/MS methods. *Food Chemistry, 213*, 714-720.

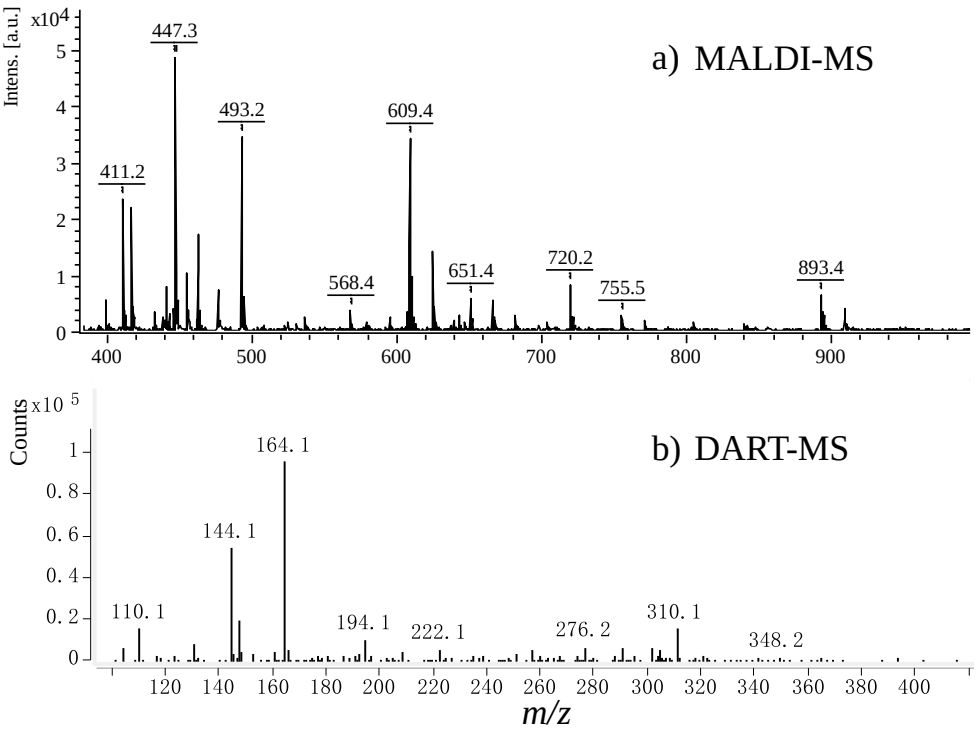

Fig. 1. Representative spectra of a wine sample from France obtained by a) MALDI-MS and b) DART-MS.

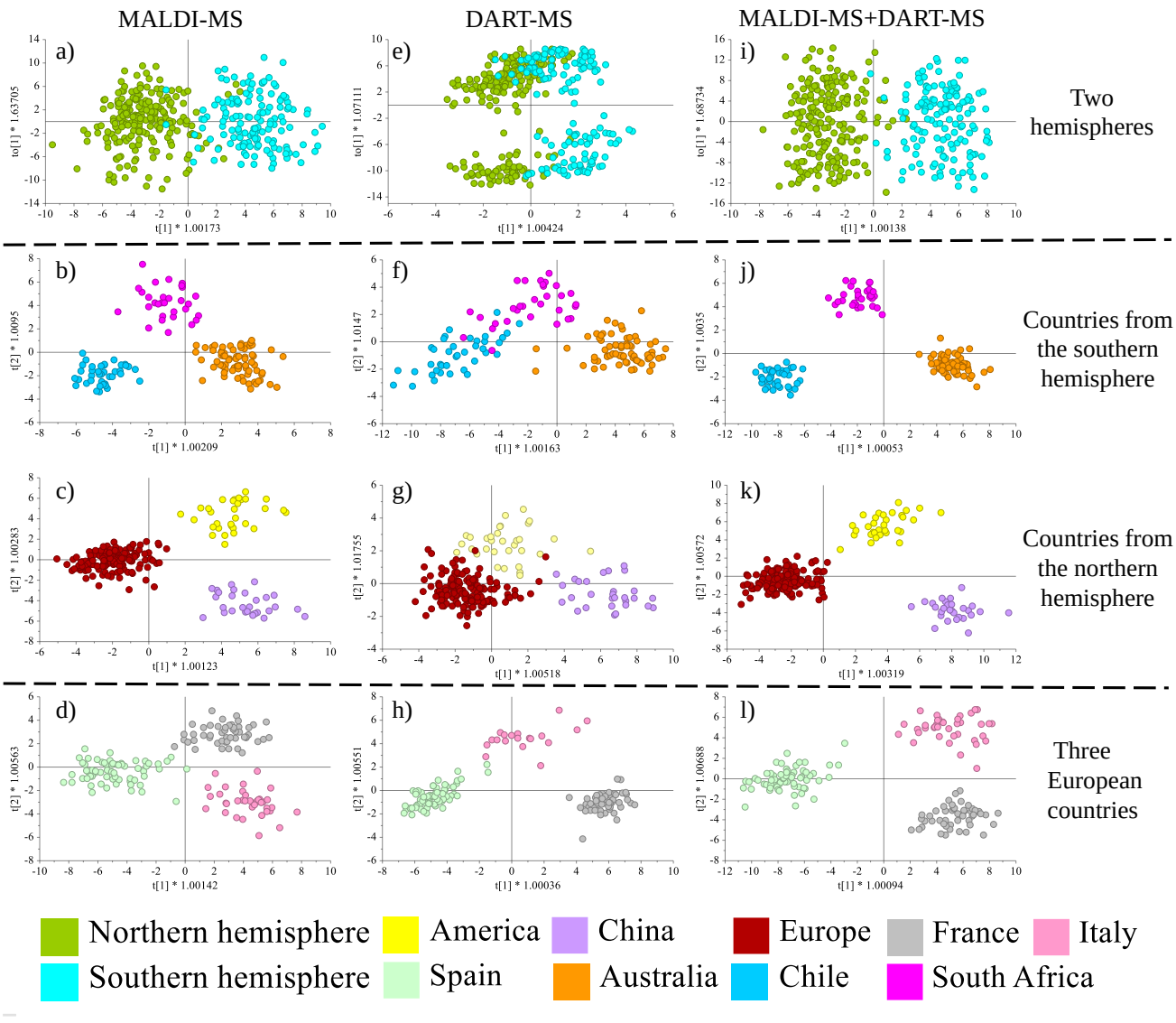

Fig. 2. OPLS-DA classification of the wine samples for differentiating the two hemispheres, countries from the southern hemisphere, countries from the northern hemisphere with Spain, Italy and France as a whole (Europe), and three European countries based on the MALDI-MS results (a-d), the DART-MS results (e-h) and the combined MALDI-MS+DART-MS results (i-l).

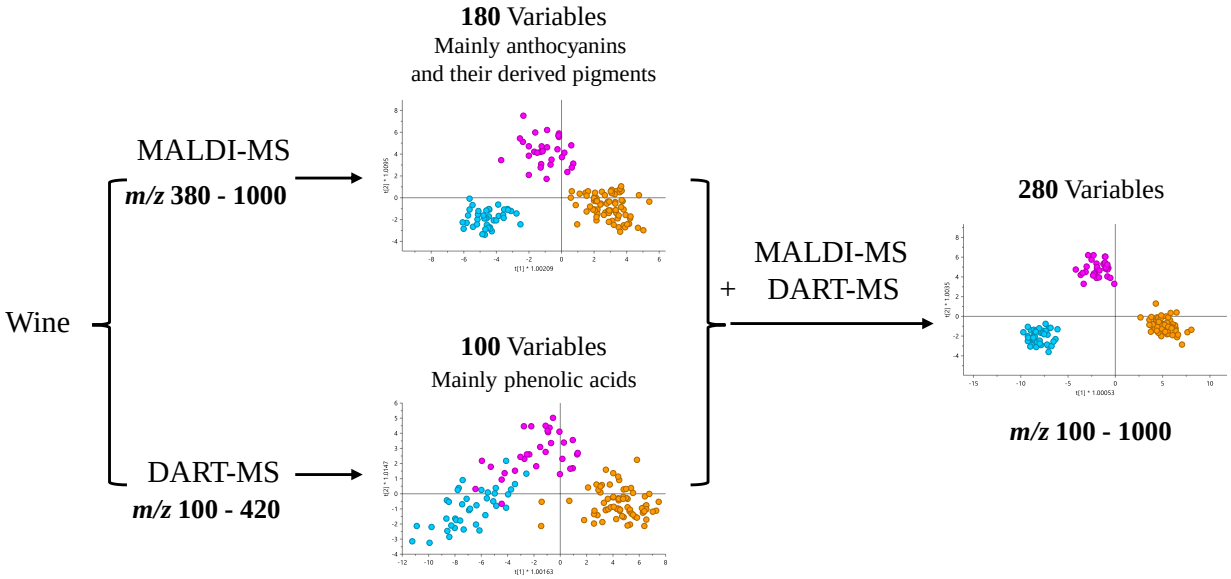

Fig. 3. Scheme of combining spectral results from MALDI-MS and DART-MS.

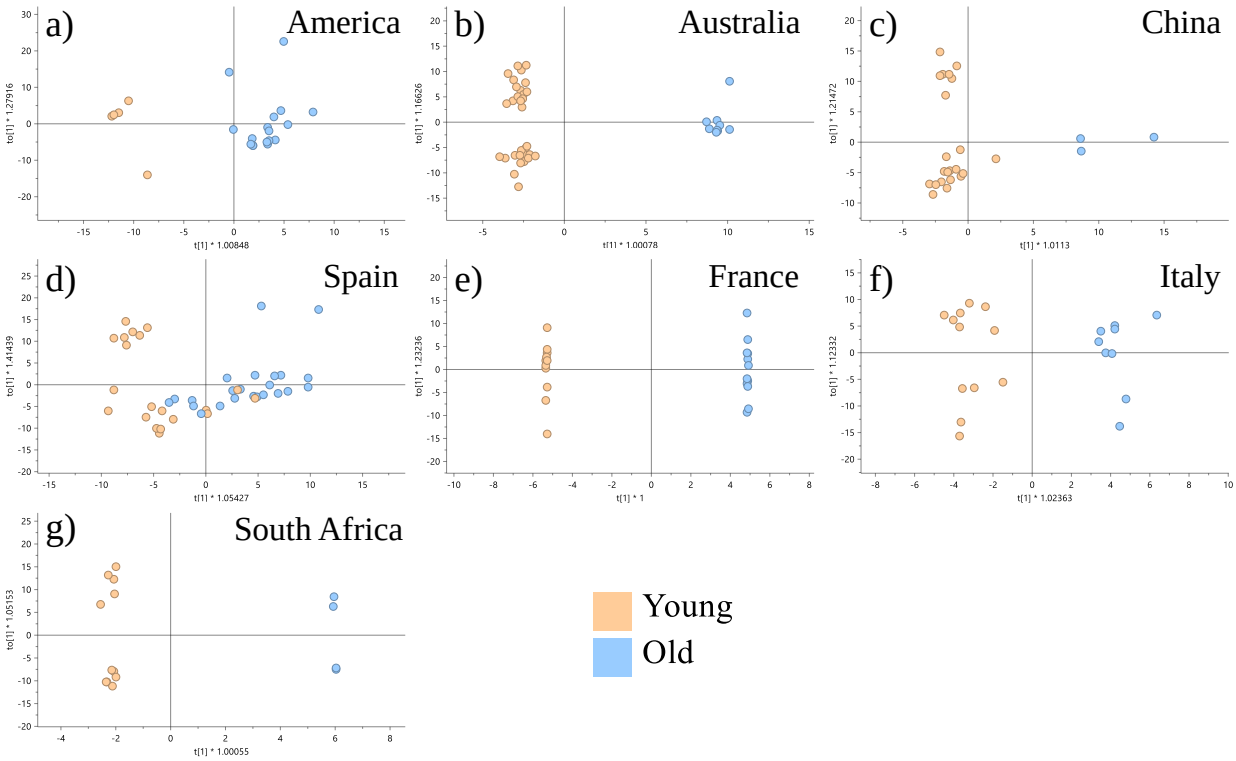

Fig. 4. OPLS-DA classification of the young group and the old group from different countries.

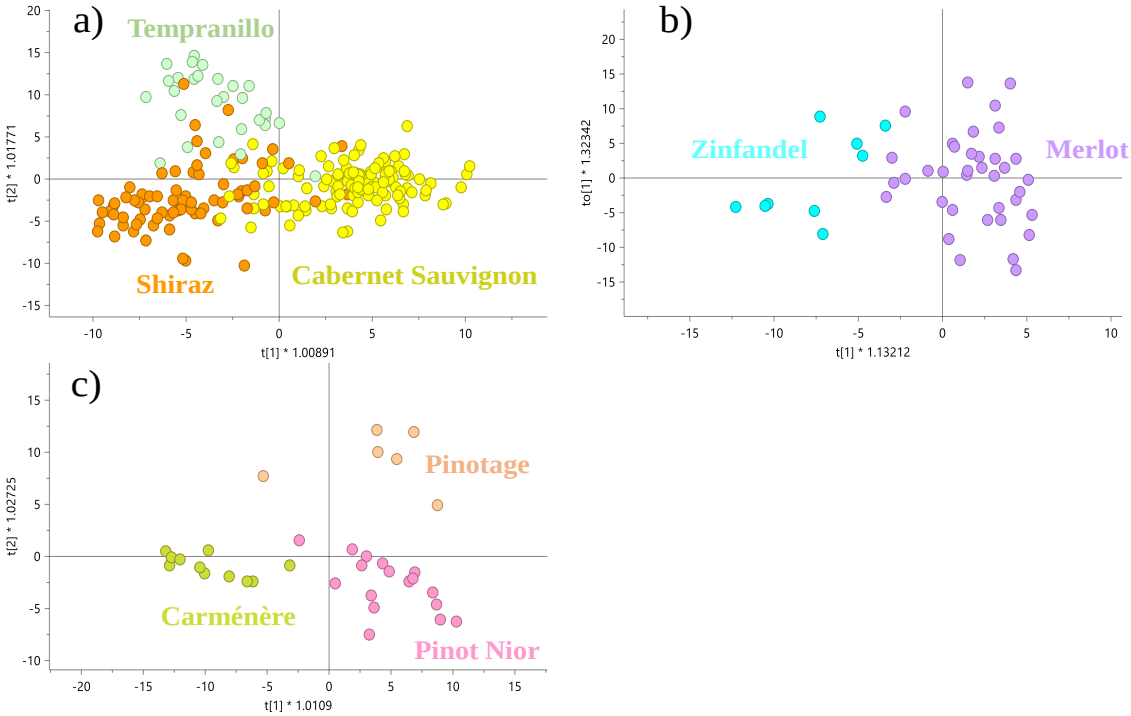

Fig. 5. OPLS-DA classification of grape varieties from a) Tempranillo, Shiraz and Cabernet Sauvignon, b) Zinfandel and Merlot and c) Carménère, Pinotage and Pinot Nior.

Table 1. The overall correct classification rates (%) of the OPLS-DA models.

	Calibration			Validation		
	MALDI	**DART**	**MALDI+DART**	**MALDI**	**DART**	**MALDI+DART**
Model 1	96.9	86.7	98.3	91.2	70.7	89.0
Model 2	100.0	96.5	100.0	86.3	74.0	93.2
Model 3	100.0	95.8	100.0	90.7	89.8	96.3
Model 4	99.3	99.3	100.0	82.1	83.5	88.5
America	100.0	78.1	100.0	81.3	87.5	93.8
China	100.0	100.0	100.0	85.7	71.4	100.0
France	100.0	100.0	100.0	82.1	100.0	100.0
Italy	100.0	95.0	100.0	68.4	86.8	68.4
Spain	98.4	100.0	100.0	90.3	64.5	90.3
Australia	100.0	100.0	100.0	86.5	75.7	91.9
Chile	100.0	94.9	100.0	84.2	79.0	94.7
South Africa	100.0	90.3	100.0	88.2	64.7	94.1