



UNRAVELLING REGIONAL TYPICALITY OF AUSTRALIAN PREMIUM SHIRAZ THROUGH AN UNTARGETED METABOLOMICS APPROACH

Sijing Li^{1*}, Leigh Schmidtke¹, Wes Pearson^{1,2}, Leigh Francis², John Blackman¹

¹National Wine and Grape Industry Centre, Charles Sturt University, School of Agricultural and Wine Science, Locked Bag 588, Wagga Wagga, NSW 2678, Australia

²The Australian Wine Research Institute, PO Box 197, Glen Osmond, SA 5064, Australia

*Corresponding author: sli@csu.edu.au

Abstract

Aims: The current study seeks to demonstrate that premium Shiraz wines from different Australian geographic indications (GI) can be distinguished by their volatile compound composition.

Method and Results: Barossa, McLaren Vale, Hunter Valley, Canberra District, Heathcote and Yarra Valley were selected to represent a range of climatic conditions. In each region, three to four wines were chosen by a panel of local winemakers to represent the regional wine styles. Volatile fractions of all wines ($n = 22$) were extracted from 3 bottles, using a solid phase extraction protocol. The extracts were analysed in random order with GC-EI-QTOFMS. Features (a feature is an ion fragment with unique m/z , retention time and intensity) were extracted from the raw MS data and then grouped and deconvoluted, to give 321 'compound' spectra. The feature with the highest intensity in each spectrum was taken to build a classification model using the random forests (RF) algorithm. This model was able to correctly classify all samples according to their GI. Features with lower contributions to the model were gradually eliminated, and 80 features were found to be sufficient to maintain the accuracy of classification. Of these 80 features, 45 were tentatively identified by comparing their mass spectra and Kovats retention indices with either an in-house library or the NIST 14 library. A range of these compounds, including terpenoids, benzenoids, esters, furan derivatives and aliphatic alcohols, have been associated with grape composition, wine making influences and the aging process.

Conclusion: This study showed that Shiraz wines from different GIs have unique volatile 'fingerprints'. These classifications may be associated with the unique terroir of the GI, which includes climatic and production differences. Well-designed processing tools for MS data and robust data mining algorithms served as a powerful combination of techniques to uncover the regional 'fingerprints'.

Significance and Impact of the Study: This study realised the challenging assignment of separating commercial Shiraz wines from six GIs according to their volatile composition. Forming a part of a broader project that include sensory and climate data, it helps to benchmark regional styles of Australian Shiraz wine.

Keywords: Terroir, metabolomics, Australia, Shiraz, climate

Introduction

It is expected that a particular set of sensory and chemical characteristics of a wine arise from the specific region or locality of its production. The 'origin of production' plays an important role in the perception of wine quality by the consumer, since they associate higher quality with wines from more renowned regions (Johnson and Bruwer, 2007). Therefore, the sensory and chemical profiles differentiating wine regions have received international research interest (Such as those reviewed in Alañón *et al.*, 2015).

Volatile compounds are a group of important metabolites in wine, as many of them are associated with specific sensory characters. For example, some fermentation-derived esters, especially fatty acid ethyl esters, were found to contribute to berry fruit aromas in red wine (Escudero *et al.*, 2009); while six-carbon alcohols and methoxypyrazines can impart green and grassy aromas (Benkwitz *et al.*, 2012). These compounds are considered important in wine chemistry and some studies sought to elucidate regional differences by targeting a selection, usually 30 - 60, of these compounds (Canuti *et al.*, 2019; King *et al.*, 2014; Kustos *et al.*, 2019). Another approach to examine the regional differences in wine is to use untargeted analysis in combination with statistical data modelling to capture as much information about the wine volatilome as possible. Several of such attempts have been made on different grape varieties, and all reported more than 100 compounds that showed significant statistical differences between regions and many of these compounds were previously unidentified or rarely reported in wine (Carlin *et al.*, 2016; Robinson *et al.*, 2011; Šuklje *et al.*, 2019; Welke *et al.*, 2012). This is not surprising as more than 800 volatile metabolites have been identified in wine, including isoprenoids, aliphatics, higher alcohols, esters, phenylpropanoids, methoxypyrazines, and volatile sulphur compounds (Robinson *et al.*, 2014). Thus, untargeted analysis presents as an important technique that can offer information in addition to targeted analysis, so as to provide a more comprehensive and holistic overview of the wine volatilome associated with regionality.

Australia has a system of wine geographical indications (GIs), where the demarcation of GI is defined and protected legally as a trade mark. Shiraz is the most important red grape variety in Australia, accounting for 46% of total red grape plantings and 52% of total red wine export by volume in 2019 (Wine Australia, 2019). However, to date, the region-specific chemical composition of Shiraz wine is still relatively unexplored. One study quantitated 48 volatile compounds from two Australian Shiraz wines, one from a cool GI and the other from a warm GI, and observed substantial differences in concentrations of certain compounds between the two wines (Mayr *et al.*, 2014). Nevertheless, larger sample sizes that comprise a greater number of GIs is needed to help elucidate the regional volatile profile of Shiraz wine. Moreover, untargeted analysis of volatilome has not yet been applied to the investigation of Australian Shiraz.

The aim of the current project is to establish the GI-related differences in the Shiraz wine volatilome. To this end, gas chromatography coupled with time-of-flight mass spectrometer (GC-TOFMS), which has faster acquisition speed and sensitivity than the traditional single quadrupole MS, was used for compound detection. In recent years, highly successful open source software has been developed, mostly based on R language, to pre-process raw data generated by mass spectrometer (Fernández-Ochoa *et al.*, 2020). Concurrently, open source software packages written in R language have also been developed to implement complex machine learning statistical methods, such as random forests and support vector machine, for analysing high-dimensional data, such as those generated by MS. This provides opportunities to create a user-customised pipeline in processing complex raw MS data on a single platform, such as the one presented in this manuscript.

Materials and Methods

Six GIs were chosen to represent different climatic conditions within Australia, from cool climate regions such as Yarra Valley to warm regions such as Barossa Valley. In each region, an expert panel was convened to help choose 3 to 4 wines that were representative of the range of styles from each region, from an initial selection of 22 to 26 wines for each region. In total, 22 wines from 2015 and 2016 vintages were selected.

Volatile fractions were extracted in triplicate (3 different bottles), using a solid phase extraction (SPE) protocol (López *et al.*, 2002) in a randomised order. One internal standards (ISTD) was added to each wine before extraction, while another two ISTDs were added to the SPE eluent after extraction. A pooled quality assurance (QA) sample was created by mixing 100 mL of every wine, and one QA was extracted along with 10 samples. The extracts were analysed in a randomised order with GC-QTOFMS, using 70 eV electron impact ionisation. MS was operated in the scan mode to record m/z between 35 and 350. Other GC and MS parameters were reported

previously (Schmidtke *et al.*, 2013). All subsequent data analyses were performed in R (version 3.6.1), using a suite of packages.

XCMS package (Smith *et al.*, 2006; Tautenhahn *et al.*, 2008) was used for the integration and alignment of mass fragments (hereafter referred to as features). CAMERA (Kuhl *et al.*, 2012) was used to group features according to common elution time and peak width, giving “pseudo-spectra”. Extracted ion chromatograms (EIC) for all features provided a visual tool to assess the peak shape and quality, which helped with eliminating false-positives peaks and resolving co-eluting compounds. From each “pseudo-spectrum” only the feature with the highest intensity was retained, resulting in 312 features representing individual compounds. The intensity of all features were normalised by taking the geometric mean of the three ratios with the three added internal standards, and then transformed using Box Cox transformation, mean centred, and scaled by dividing the standard deviation. Random forests (Breiman, 2001) was chosen as the classification model to investigate the underlying pattern in chemical composition associated with different GIs. The R package ‘randomForest’ (version 4.6-14, <http://cran.r-project.org/>) was used to build the RF model. A metric computed by the model, mean decrease Gini (MDG), has been demonstrated to be a robust measure for ranking feature importance in the model (Calle and Urrea, 2010). Features with lower MDG were progressively eliminated until the classification accuracy of the model started to decrease. This step reduced the number of features used by the model from 312 to 80.

Of the 80 features, 18 were identified by comparing with an in-house library containing TOFMS spectrum of analytical grade compounds. Another 27 were tentatively identified by comparing MS spectra contained in the NIST14 library and retention indices reported in the literature. The rest of the features were regarded as ‘unknowns’. A hierarchical cluster analysis was performed using these 80 features, with both features and samples clustered using the complete linkage method based on Euclidean distance. The result was visualised using a heatmap constructed with the ‘ComplexHeatmap’ package (version 2.4.3, <http://cran.r-project.org/>).

Results and Discussion

An initial principal component analysis (PCA) was performed on all 321 features to examine the overall data quality, i.e. to identify outliers and systematic errors. Figure 1 shows that all QAs clustered tightly in the centre of the distribution of the samples; moreover, all triplicates of each wine clustered closely together. PCA provides assurance that the data is sufficiently robust for further statistical modelling of the sample composition.

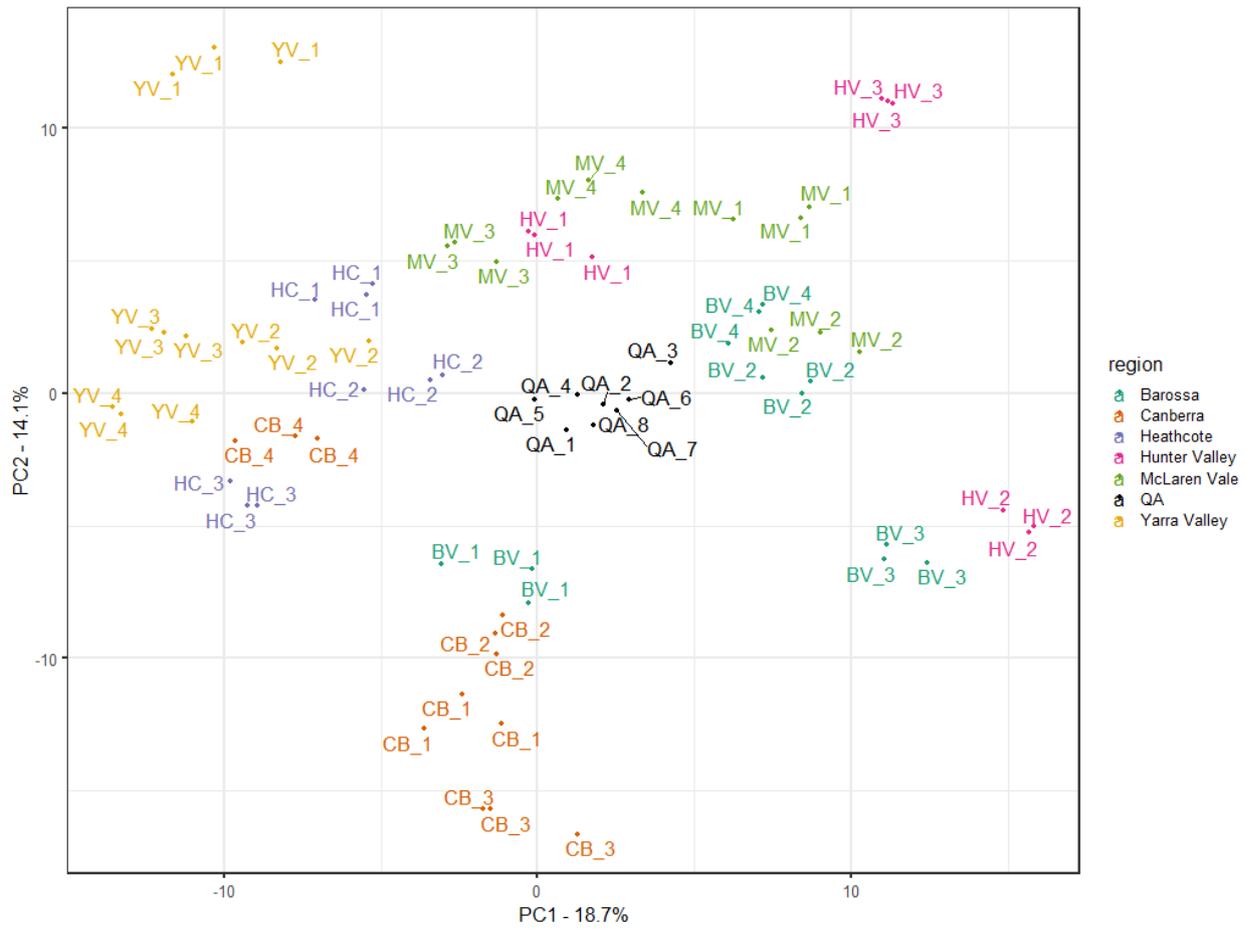


Figure 1. Principal component analysis using the most intense feature from all pseudo-spectra. The sample with the same code corresponds to the independent bottles of the same wine, and colour corresponds to different GIs.

Random forest model showed great ability to discriminate sample based on region (Figure 2A). A clear sample clustering based on GI could be found among samples in the first two dimensions that account for 50.9% of the total variance. The first dimension clearly separates Barossa and McLaren Vale wines from those made in Yarra Valley and Canberra District. Hunter Valley and Heathcote wines are positioned between these two divisions. Hunter Valley wines overlap with Barossa wines whereas Heathcote wines appear to have commonality with Canberra wines along the first dimension. The second dimension separates samples from between Barossa and McLaren Vale, as well as between Yarra Valley and Canberra. By eliminating features with lower importance rankings, a model could be built on only 80 features without compromising the classification accuracy of the original model (Figure 2B).

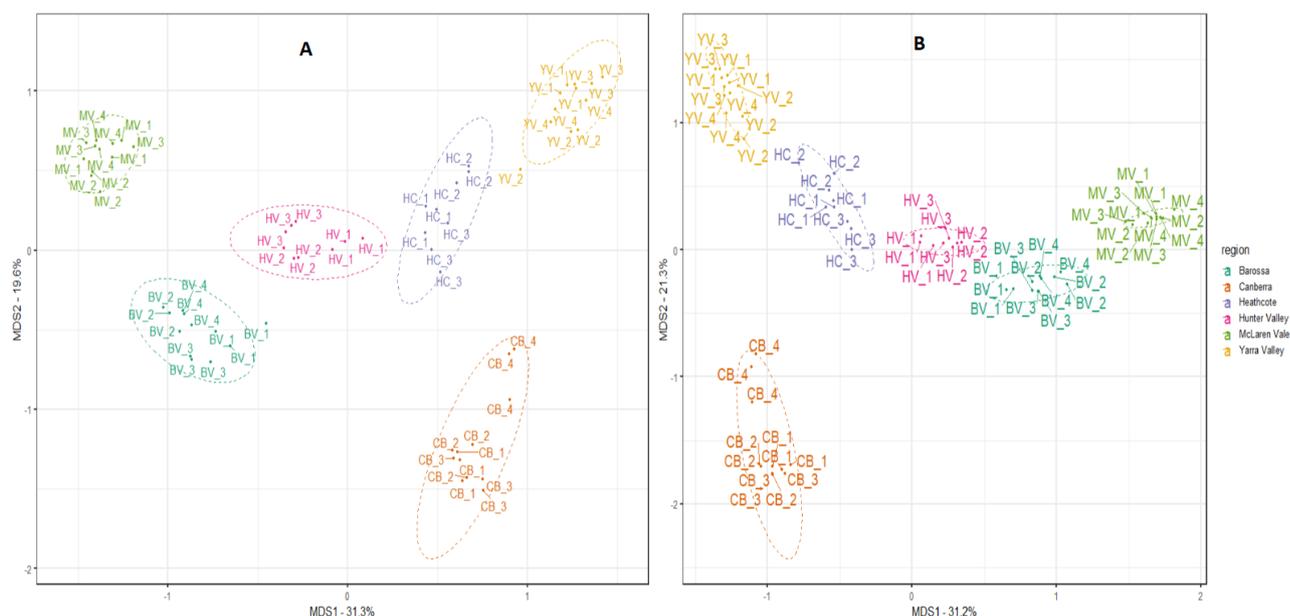


Figure 2. The multi-dimensional scaling plot based on the distance matrix generated by random forests model, using (A) 321 features and (B) 80 features. The sample with the same code corresponds to the independent bottles of the same wine, and colour corresponds to different GIs. The dotted lines represent the 95% confidence intervals based on multivariate *t*-distribution.

The 80 volatiles retained in the RF model were regarded as important markers for regionality and subjected to tentative Identification. The wines were grouped together by a hierarchical cluster analysis based on the relative concentrations of these compounds (Figure 3), revealing some region-specific patterns. Benzyl alcohol is a grape derived compound that was found to be more abundant in grapes from cooler regions (Šuklje *et al.*, 2019). Together with its acetate ester, it was found to be present at higher concentrations in wines from the Canberra District compared to other regions. Another grape derived compound, 1,1,6,-trimethyl-1,2-dihydronaphthalene (TDN), which was reported to be more abundant in wines from warmer regions (Šuklje *et al.*, 2019), was found to be at higher concentration in Hunter Valley wines, followed by Barossa and McLaren Vale. Three fermentation derived ethyl esters, 2-methylbutanoate, ethyl 3-methylbutanoate and ethyl 2-phenylacetate, known to contribute to berry fruit aromas in red wines (Escudero *et al.*, 2007), were found at relatively higher concentrations in McLaren Vale wines. Another two esters, ethyl cinnamate and ethyl dihydrocinnamate showed highest concentration in Yarra Valley wines. Both cinnamates could impart sweet-caramel note in red wine (Escudero *et al.*, 2007). Two age related compounds, ethyl furan-2-carboxylate and ethyl vanillate were found at relatively higher concentrations in wines from Hunter Valley.

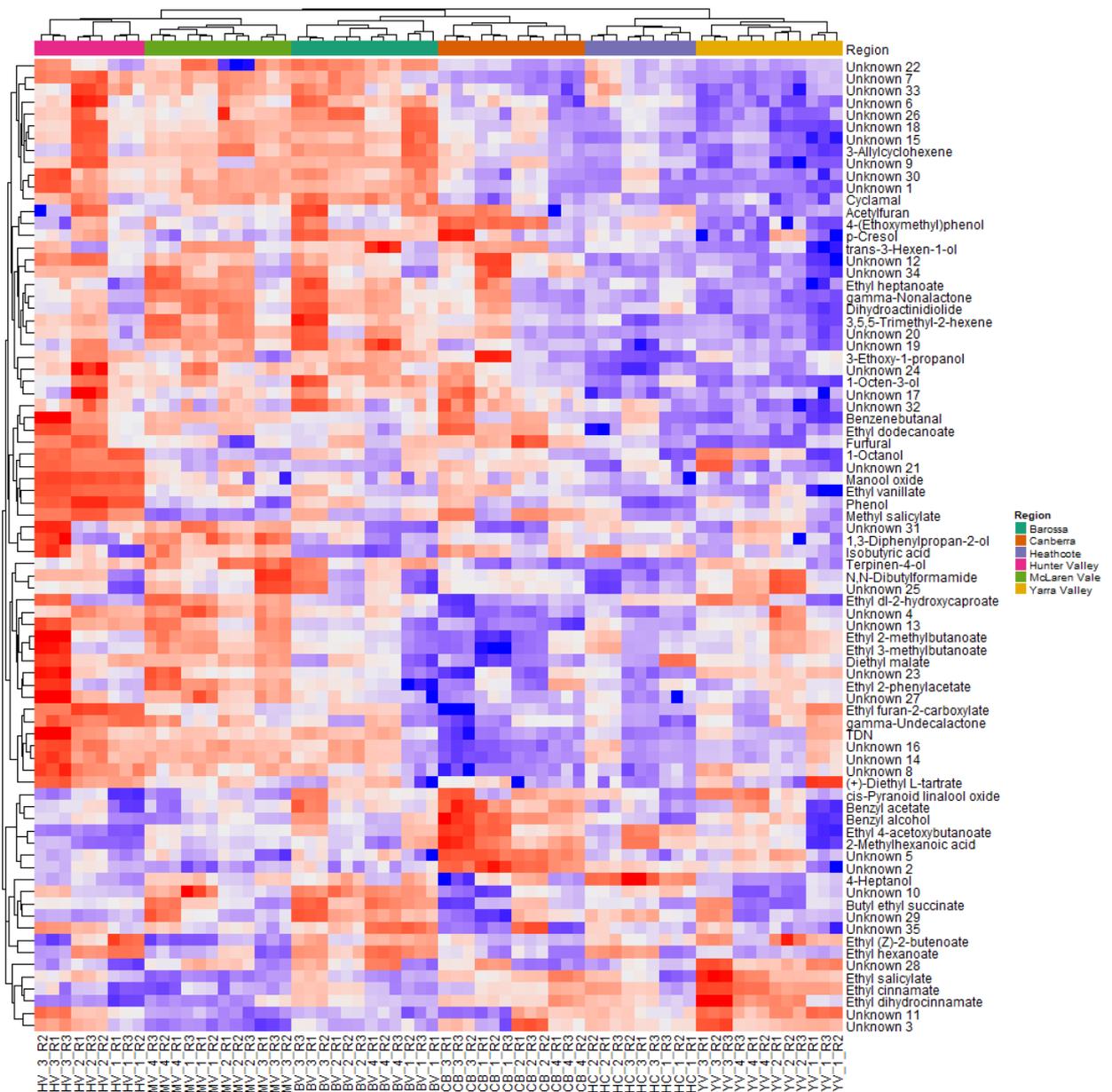


Figure 3. Heatmap using the selected features from RF model. R1, R2 and R3 represent 3 independent bottles of the same wine. Red, white and blue colours represent relative concentration of the compounds from high to low. One of the limitations of the current work is not being able to provide unequivocal identification for a number of compounds that showed regional distinction. However some unknown compounds showed very clear pattern in terms of regionality. Future work to further identify these unknown compounds is required. Furthermore, the results from untargeted analyses of volatome may be interpreted in conjunction with the distinctive regional sensory attributes, in order to explore the relationship between wine volatile composition and regional wine styles.

Conclusions

Commercial wines, despite their complex nature, showed distinctive regional chemical composition, within which some volatile components were identified. A relatively small number of compounds were able to classify the studied wines according to their GI, despite a range of wine styles being deliberately selected. The volatile profiling showed that regional typicality is a multi-faceted concept and encompasses volatiles derived from grapes, fermentation and aging, which are in turn affected by natural elements and human interventions.

Acknowledgements

The authors acknowledge funding from Wine Australia and Charles Sturt University (CSU 1602). We also acknowledge the help in technical aspects provided by Drs. Urska Vrhovsek, Silvia Carlin and Pietro Franceschi from Fondazione Edmund Mach, Italy.

References

- Alañón, ME., Pérez-Coello, MS., Marina, ML.,** 2015. Wine science in the metabolomics era. *TrAC Trends in Analytical Chemistry*, 74: 1-20.
- Benkowitz, F., Tominaga, T., Kilmartin, PA., Lund, C., Wohlers, M., Nicolau, L.** 2012. Identifying the chemical composition related to the distinct aroma characteristics of New Zealand Sauvignon Blanc wines. *American Journal of Enology and Viticulture*, 63: 62-72.
- Calle, ML., Urrea, V.,** 2010. Letter to the eEditor: Stability of rRandom fForest importance measures. *Briefings in Bioinformatics*, 12: 86-89.
- Canuti, V., Frost, S., Lerno, LA., Tanabe, CK., Zweigenbaum, J., Zanoni, B., Ebeler, SE.,** 2019. Chemical characteristics of Sangiovese wines from California and Italy of 2016 vintage. *Journal of Agricultural and Food Chemistry*, 67: 2647-2659.
- Carlin, S., Vrhovsek, U., Franceschi, P., Lotti, C., Bontempo, L., Camin, F., Toubiana, D., Zottele, F., Toller, G., Fait, A., Mattivi, F.,** 2016. Regional features of northern Italian sparkling wines, identified using solid-phase micro extraction and comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry. *Food Chemistry*, 208: 68-80.
- Escudero, A., Campo, E., Fariña, L., Cacho, J., Ferreira, V.,** 2007. Analytical characterization of the aroma of five premium red wines. Insights into the role of odor families and the concept of fruitiness of wines. *Journal of Agricultural and Food Chemistry*, 55: 4501-4510.
- Fernández-Ochoa, Á., Quirantes-Piné, R., Borrás-Linares, I., Cádiz-Gurrea, MDLL., Precisesads Clinical Consortium, Alarcón Riquelme, ME., Brunius, C., Segura-Carretero, A.,** 2020. A case report of switching from specific vendor-based to R-based pipelines for untargeted LC-MS metabolomics. *Metabolites*, 10: 28.
- Johnson, R., Bruwer, J.,** 2007. Regional brand image and perceived wine quality: the consumer perspective, *International Journal of Wine Business Research*, 19: 276-297.
- King, ES., Stoumen, M., Buscema, F., Hjelmeland, AK., Ebeler, SE., Heymann, H., Boulton RB.,** 2014. Regional sensory and chemical characteristics of Malbec wines from Mendoza and California. *Food Chemistry*, 143: 256-267.
- Kuhl, C., Tautenhahn, R., Böttcher, C., Larson, TR., Neumann, S.,** 2012. CAMERA: An integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Analytical Chemistry*, 84: 283-289.
- Kustos, M., Gambetta, JM., Jeffery, DW., Heymann, H., Goodman, S., Bastian, SEP.,** 2020. A matter of place: Sensory and chemical characterisation of fine Australian Chardonnay and Shiraz wines of provenance. *Food Research International*, 130: 108903.
- López, R., Aznar, M., Cacho, J., Ferreira, V.,** 2002. Determination of minor and trace volatile compounds in wine by solid-phase extraction and gas chromatography with mass spectrometric detection. *Journal of Chromatography A*, 966: 167-177.
- Mayr, CM., Geue, JP., Holt, HE., Pearson, WP., Jeffery, DW., Francis, IL.,** 2014. Characterization of the key aroma compounds in Shiraz wine by quantitation, aroma reconstitution, and omission studies. *Journal of Agricultural and Food Chemistry*, 62: 4528-4536.
- Robinson, AL., Boss, PK., Heymann, H., Solomon, PS., Trengove, RD.,** 2011. Development of a sensitive non-targeted method for characterizing the wine volatile profile using headspace solid-phase microextraction comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry. *Journal of Chromatography A*, 1218: 504-517.

Schmidtke, LM., Blackman, JW., Clark, AC., Grant-Preece, P., 2013. Wine metabolomics: Objective measures of sensory properties of Semillon from GC-MS profiles. *Journal of Agricultural and Food Chemistry*, 61: 11957-11967.

Smith, CA., Want, EJ., O'Maille, G., Abagyan, R., Siuzdak, G., 2006. XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry*, 78: 779-787.

Šuklje, K., Carlin, S., Antalick, G., Blackman, JW., Deloire, A., Vrhovsel, U., Schmidtke, LM., 2019. Regional discrimination of Australian Shiraz wine volatome by two-dimensional gas chromatography coupled to time-of-flight mass spectrometry. *Journal of Agricultural and Food Chemistry*, 67: 10273-10284.

Tautenhahn, R., Böttcher, C., Neumann, S., 2008. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics*, 9: 504.

Welke, JE., Manfroi, V., Zanus, M., Lazarotto, M., Alcaraz Zini, C., 2012. Characterization of the volatile profile of Brazilian Merlot wines through comprehensive two dimensional gas chromatography time-of-flight mass spectrometric detection. *Journal of Chromatography A*, 1226: 124-139.

Wine Australia, 2019. 'Variety snapshot 2019 – Shiraz' Accessed on: 29/04/2020, Retrieved from: <https://www.wineaustralia.com/market-insights/variety-snapshots>