# CLUSTERING WINE AROMATIC COMPOSITION OF *VITIS VINIFERA* GRAPEVINE VARIETIES

Marc PLANTEVIN[1,4*], Cécile THIBON[2,3], Philippe DARRIET[2,3], Georgia LYTRA[2,3], Jean-Christophe BARBE[2,3], Justine GARBAY[2,3], Julien LECOURT[4], Cornelis van LEEUWEN[1]

*[1]EGFV, Univ. Bordeaux, Bordeaux Sciences Agro, INRAE, ISVV, F-33882 Villenave d'Ornon, France*
*[2]Univ. Bordeaux, Bordeaux INP, INRAE, OENO, UMR 1366, ISVV, F-33140 Villenave d'Ornon, France*
*[3]Bordeaux Sciences Agro, Bordeaux INP, INRAE, OENO, UMR 1366, ISVV, F-33170 Gradignan, France*
*[4]Pôle Scientifique, Bernard Margez Grands Vignobles, 33000 Bordeaux, France*

*Corresponding author: marc.plantevin@inrae.fr

**Context and purpose of the study**
Climate change is likely to impact wine typicity across the globe, raising concerns in wine regions historically renowned for the quality of their terroir. Amongst several changes in viticultural practices, replacing some of the planting material (i.e clones, rootstocks and cultivars) is thought to be one of the most promising potential levers to be used for adapting to climate change. But the change of cultivars also involves the issue of protecting the region's wine typicity. In Bordeaux (France), extensive research has been conducted on identifying meridional varieties that could be good candidates to help guard against the effects of climate change while less research has been done concerning their impacts on Bordeaux wine typicity. Thus, the present study aims to characterize the aromatic composition of a large pool of *Vitis vinifera* cultivars through the analyses of some impacting aromatic compounds. Then, aromatic composition of traditional-Bordeaux varieties and non-Bordeaux varieties are compared.

**Materials and Methods**
A 2-hectares plot of 84 cultivars was planted in 2013, in the Médoc wine region (Bordeaux, France) within the vineyards of a wine estate. Amongst this very large collection of cultivars, a pool of 25 red varieties was isolated, including traditional Bordeaux varieties and potential candidates for introduction in the Bordeaux varietal mix. Each of those varieties has been separately vinified since 2018 in 2hL stainless steel tanks, close to commercial wine production conditions. 41 major aroma compounds were then quantified in each variety for each vintage (from three to five vintages per cultivar) by gas chromatography and mass spectrometry (GC-MS). Statistical analyses, including hierarchical clustering analysis (HCA) and principal component analysis (PCA) were then performed on this unique dataset for aroma profile characterization and to discriminate and isolate varieties according to their aromatic profile.

**Results**
As expected, analyses resulted in a strong varietal characterization of the different wines with a significant vintage effect on some of the aroma compounds. Of the 41 aroma compounds analyzed, a subset appears to explain a large part of the Bordeaux wines aromatic composition. Clustering of cultivars was possible, and Bordeaux cultivars grouped well together into a unique cluster. Interestingly, a few non-traditional Bordeaux cultivars were close to some of the classical Bordeaux varieties in both the HCA and PCA analyses. These results enhanced the idea that some non-native cultivars could be introduced in the Bordeaux cultivar mix while maintaining global wine typicity. This methodology could help other established wine regions to identify varieties that could be potential candidates for adaptation to climate change.

**Keywords**: wine typicity, aroma compounds, climate change, GC-MS, *Vitis vinifera*

## 1. Introduction

Temperature is of major importance in the final berry and wine composition (Keller, 2010), which leads to modification in wine composition as a result of global warming (van Leeuwen & Darriet, 2016). Some of these changes are well documented and already observed by growers and researchers, especially the decrease of total acidity and the increase of alcohol content (Mira de Orduña, 2010). The impact of climate change on aroma compounds has also been investigated (Pons et al., 2016). Some compounds responsible for vegetative aromas in wines, such as 2-methoxy-3-isobutylmethoxypyrazine (IBMP), are known to significantly decrease as a result of higher temperatures (Falcao et al., 2007). Conversely, concentrations of compounds responsible for cooked-fruit aromas in red wines (such as lactones and furan compounds) increase with higher temperatures (Allamy et al., 2018). Other molecules have contrasting results in the scientific literature, such as $C_{13}$ – norisoprenoids and monoterpenes, which could benefit from an increase of temperature up to a certain point from which concentrations decrease after reaching this tipping point (Pons et al., 2016).
Terroir expression, i.e. wine typicity related to origin, relies on a subtle balance between many compounds (van Leeuwen & Seguin, 2006). The changes in wine composition induced by higher temperatures may lead to a loss of typicity in established wine regions. Amongst the different options for adaptation to climate change, the change of cultivars is thought to be one of the most promising levers. Varietal characterization of wine aromatic composition has been widely studied in the past (Câmara et al., 2004; Ferreira et al., 2000; Versini et al., 1994) and the results of these studies can be helpful to find cultivars with similar typicity. The present study aims at characterizing differences and similarities of the aromatic composition of 25 red cultivars over 3 to 5 vintages in the Bordeaux winegrowing region. This is the first step for the introduction of new varieties in the Bordeaux varietal mix that would be better suited to a warmer climate, while preserving some of the Bordeaux wines typicity. The second step will be to confirm these results by sensorial analyses.

## 2. Material and Methods

### Experimental Design

The experimental vineyard is located in Château La Tour Carnet in Saint-Laurent Médoc, 33112, France (45.15°N, 0.79°W) on a parcel planted in 2014 within the Haut-Médoc AOP. It includes 25 red traditional and non-traditional Bordeaux varieties (all *Vitis vinifera*). Each row is made of one cultivar with 150 vines per row. The vineyard is dry-farmed with all varieties being double Guyot-pruned and trained with a vertical shoot positioning trellis on a sandy/loam soil. The planting density is 8 333 vines/ha.

### Harvest and vinification

Varieties were harvested when they reached optimum maturity, assessed by standard enological parameters and berry tasting. All varieties were de-stemmed and vinified in 2 or 4 hL stainless steel tanks. Yeast assimilable nitrogen in the musts was adjusted to 250 mg/L and $SO_2$ was added at a level of 4g/Hl before fermentation. All wines were fermented using Zymaflore® FX10 dry active yeast from Laffort (33270, Floirac, France). Malo-lactic fermentation was induced using Lactoenoes® B7 Direct from Laffort. At the end of both fermentations, free $SO_2$ was adjusted at 40 mg/L.

**Aromatic compound quantification**
Due to different climatic hazards, wines from some varieties were not obtained in all five vintages. 24 varieties were harvested in 2018 and 2019, 13 in 2020, 5 in 2021 and 25 in 2022. Analyses of the 2022 vintage was still in progress at the time of writing. Each wine (taken directly from the bottle) was analyzed once. Two methods were used to quantify aromatic compounds. 26 esters were quantified according to the method developed in Antalick et al., 2010. Following, 15 aromatic compounds that belong to the families of thiols, lactones, aldehydes, furanones and $C_{13}$-norisoprenoids, were quantified using the method developed in Thibon et al., 2015.

**Statistics**
The analyses were run on R Studio version 2022.12.0 using Dplyr package for descriptive analysis, Stats package for hierarchical clustering analysis (HCA), FactoMineR package for principal component analysis (PCA) and RandomForest package for random forest algorithm.

## 3. Results and Discussion

### *3.1. Vintage effect*
A strong vintage effect was shown. Each vintage seems well-characterized by PCA, except for 2021 due to a low number of data points (Figure 1). 2019 contributes more to the positive side of the second dimension, mostly explained by $C_{13}$-norisoprenoids and lactones compounds (data not shown), while 2018 and 2020 are similarly negatively contributing to this second dimension. However, these later vintages are opposed along the first dimension (largely explained on the positive side by higher alcohol acetates), with 2018 having a positive contribution unlike 2020. This findings, in accordance with the literature (Vilanova et al., 2007), show the importance of taking into account more than one vintage in varietal aromatic composition characterization.

### *3.2. Year-to-year stability of varietal aromatic composition*
To characterize the aromatic composition of the cultivars, the consistency of varieties in having high or low concentrations of some specific compounds, was investigated over three vintages from 2018 to 2020. To buffer vintage variation in aromatic signature, we ranked the varieties from 1 to 24 in 2018 and 2019 and from 1 to 14 in 2020 as a function of their concentration for each compound (1 being the variety with the highest concentration of the compound). Table 1 shows six varieties which turned out to have a stable composition for 12 compounds (considered as constantly in the highest or lowest values in the three vintages). Interestingly, it must be noted that Marselan has the highest concentration of 3 sulfanylexan-1-ol (3SH) (and higher values of ß-Damascenone) in the three vintages, an impacting compound in the fruity character of red Bordeaux cultivars (Blanchard et al., 2004). It is also very interesting that Merlot and Cabernet-Sauvignon are both expressing the two highest values of ethyl phenylacetate in the three vintages. This aromatic ester, responsible for flowery/rose aroma nuances, is synthetized through a slow esterification process as the ethyl esters of branched aliphatic acids (Antalick et al., 2010), such as ethyl 2-methylpropanoate, ethyl-2-methylbutanoate and ethyl pentanoate. These compounds are also constantly present in higher concentrations in the three vintages in wines made from Merlot and Cabernet-Sauvignon.

### *3.3. Hierarchical Clustering Analyses of the varieties (HCA)*
To overcome issues with multicollinearity, we performed an HCA (Figure 2) of the 24 varieties based on their coordinates on the 5 first dimensions of the PCA, explaining 64.8% of the total variance. Except Cot (also known as Malbec, a traditional Bordeaux variety), all the varieties from Bordeaux are grouped together in one large cluster, highlighted in red. Cabernet franc,

Merlot and Cabernet-Sauvignon are closely related with a fourth variety, originally from the south-west of France: Duras. It's also interesting to note that Marselan and Grenache are both in the same cluster. Indeed, Marselan is a cross of Cabernet-Sauvignon and Grenache which may explain aromatic composition similarities due to genetic links. The use of a random forest machine learning algorithm allowed us to identify 5 compounds (hexyl acetate, ethyl butyrate, ß-damascenone, ethyl octanoate and isoamyl octanoate) that explain a large part of the difference between the two largest clusters in the Figure 2. These results need to be confirmed with the 2022 vintage.

## 4. Conclusion

The analyses of aromatic compounds of wines made from 24 red varieties grown in Bordeaux over multiple vintages enabled to characterize their aromatic signature. After confirming potential bias induced by a strong vintage effect, we show that the aromatic signature of Bordeaux varieties could potentially be explained by some esters of branched aliphatic acids and some aromatic esters. Multi-variate analysis enabled clustering varieties as a function of their similarities in terms of aroma compounds. Using this approach, Bordeaux varieties clustered together. Machine learning algorithms would allow to obtain a deeper insight in the clustering of the varieties.

## 5. Acknowledgments

This study was supported by the scientific department of Bernard Magrez – Grands Vignobles and Château La Tour Carnet.

## 6. Litterature

**Allamy, L., Darriet, P., & Pons, A.** (2018). Molecular interpretation of dried-fruit aromas in Merlot and Cabernet Sauvignon musts and young wines: Impact of over-ripening. Food Chemistry, 266, 245–253

**Antalick, G., Perello, M.-C., & de Revel, G.** (2010). Development, validation and application of a specific method for the quantitative determination of wine esters by headspace-solid-phase microextraction-gas chromatography–mass spectrometry. Food Chemistry, 121(4), 1236–1245

**Blanchard, L., Darriet, P., & Dubourdieu, D.** (2004). Reactivity of 3-Mercaptohexanol in Red Wine: Impact of Oxygen, Phenolic Fractions, and Sulfur Dioxide. American Journal of Enology and Viticulture, 55(2), 115–120

**Câmara, J. S., Herbert, P., Marques, J. C., & Alves, M. A.** (2004). Varietal flavour compounds of four grape varieties producing Madeira wines. Analytica Chimica Acta, 513(1), 203–207

**Falcao, L. D., de Revel, G., Perello, M. C., Moutsiou, A., Celso Zanus, M., & Bordignon-Luiz, M. T.** (2007). A Survey of Seasonal Temperatures and Vineyard Altitude Influences on 2-Methoxy-3-isobutylpyrazine, C13-Norisoprenoids, and the Sensory Profile of Brazilian Cabernet Sauvignon Wines. J. Agric. Food Chem., 55, 3605–3612.

**Ferreira, V., López, R., & Cacho, J. F.** (2000). Quantitative determination of the odorants of young red wines from different grape varieties. Journal of the Science of Food and Agriculture, 80(11), 1659–1667

**Markus Keller.** (2010). The Science of Grapevines—1$^{st}$ Edition (Academic Press).

**Mira de Orduña, R.** (2010). Climate change associated effects on grape and wine quality and production. Food Research International, 43(7), 1844–1855

**Pons, A., Allamy, L., Schüttler, A., Rahut, D., Thibon, C., & Darriet, P.** (2016). What is the expected impact of climate change on wine aroma comounds and their precursors in grape? OENO One, 51(2), 141–146

**Thibon, C., Pons, A., Mouakka, N., Redon, P., Méreau, R., & Darriet, P.** (2015). Comparison of electron and chemical ionization modes for the quantification of thiols and oxidative compounds in white wines by gas chromatography–tandem mass spectrometry. Journal of Chromatography A, 1415, 123–133

**van Leeuwen, C., & Darriet, P.** (2016). The Impact of Climate Change on Viticulture and Wine Quality. Journal of Wine Economics, 11(1), 150–167

**van Leeuwen, C., & Seguin, G.** (2006). The concept of terroir in viticulture. Journal of Wine Research, 17(1), 1–10

**Versini, G., Orriols, I., & Dalla Serra, A.** (1994). Aroma components of Galician Albarino, Loureira and Godello wines. Vitis, 33, 165–170

**Vilanova, M., Cortés, S., Santiago, J. L., Martinez, C., & Fernandez, E.** (2007). Aromatic Compounds in Wines Produced During Fermentation: Effect of Three Red Cultivars. International Journal of Food Properties, 10, 867–875

## 7. Figures and Tables :

| Cepage | Name | Rank in 2018 | Rank in 2019 | Rank in 2020 |
|---|---|---|---|---|
| Merlot | Ethyl 2-methylpropanoate | 1 | 3 | 2 |
| | Ethyl 3-methylbutanoate | 1 | 3 | 1 |
| | Ethyl Phenylacetate | 1 | 2 | 1 |
| Marselan | Ethyl Dihydrocinnamate | 2 | 3 | 1 |
| | 3 Sulfanylhexan-1-ol | 1 | 1 | 1 |
| | β-damascenone | 2 | 3 | 1 |
| Cabernet franc | Butyl Acetate | 2 | 3 | 2 |
| | Ethyl Pentanoate | 20 | 20 | 13 |
| Cabernet-Sauvignon | Ethyl-2-methylbutanoate | 2 | 1 | 2 |
| | Ethyl Phenylacetate | 2 | 1 | 2 |
| Carménère | Phenylacetaldehyde | 2 | 3 | 2 |
| Tannat | Ethyl-2-sulfanylacetate | 21 | 20 | 11 |

**Table 1:** Varieties with consistent ranking in impacting compound concentration in 2018, 2019 and 2020. Wines from 24 varieties were analyzed in 2018 and 2019, 14 in 2020.

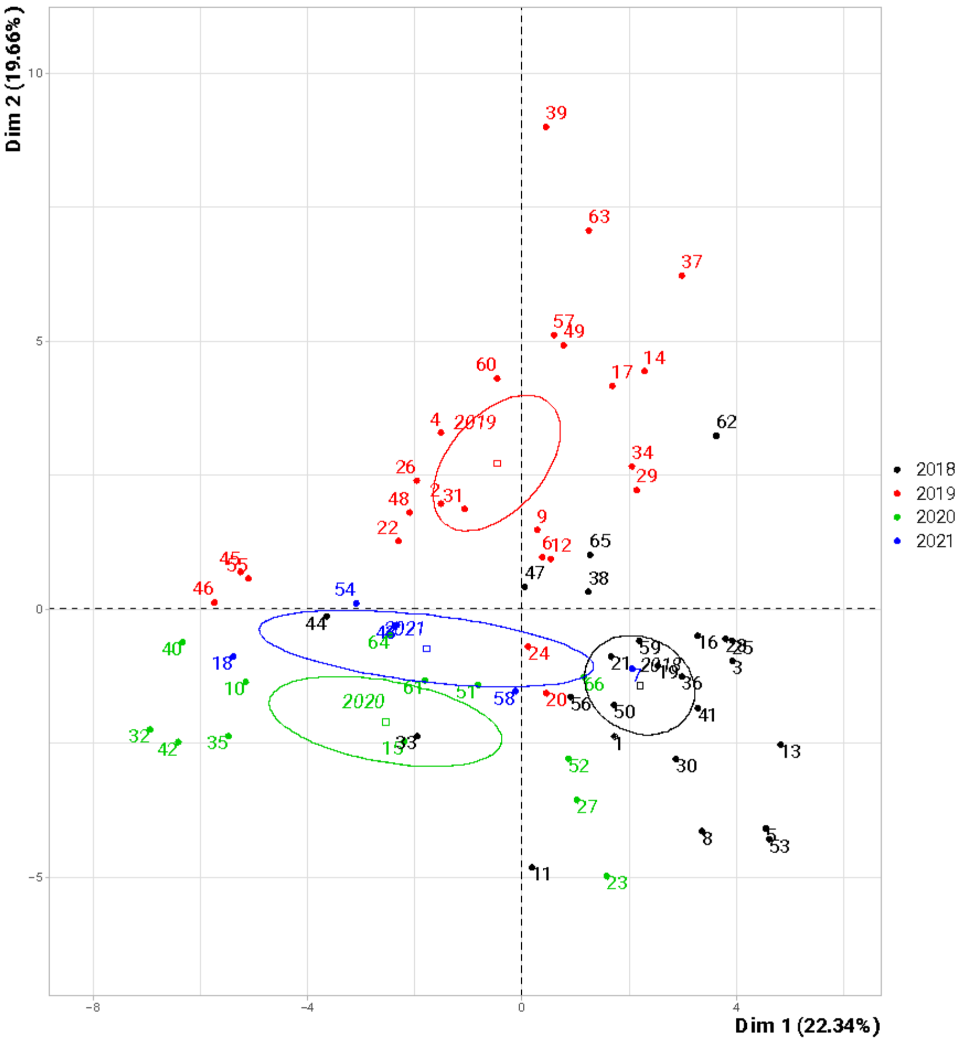


**Figure 1:** Principal Component Analysis (PCA) of the 61 samples. The two dimensions account for 42% of the total variance. Ellipses with confidence level at 95% were drawn around categories of vintage with barycenters for each vintage shown as squares.

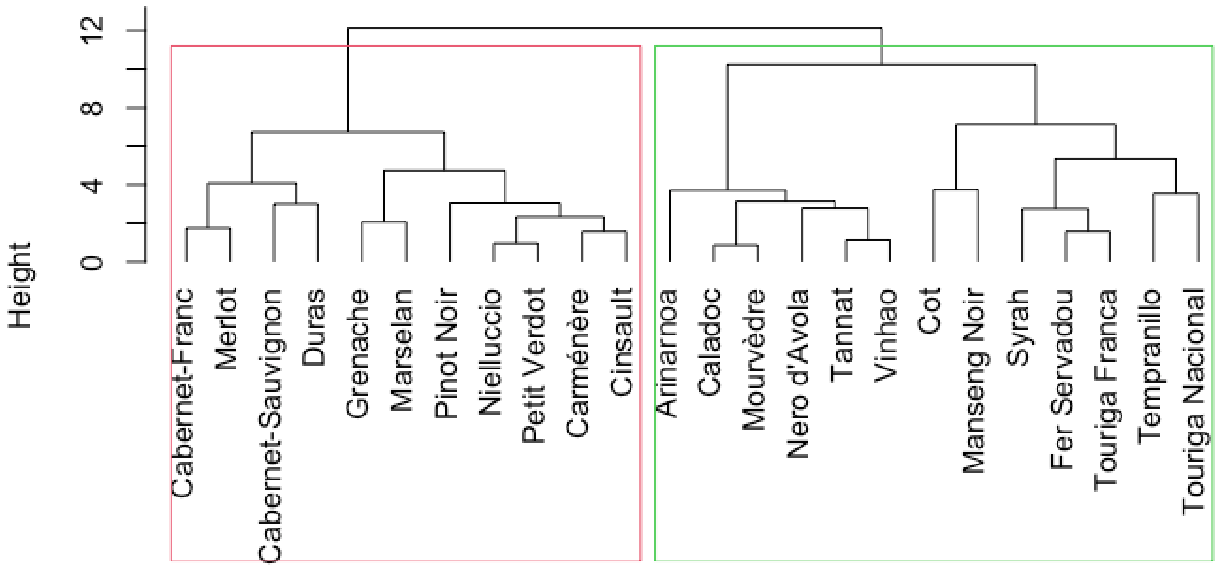


**Figure 2:** Hierarchical Clustering Analyses (HCA) of the 24 varieties based on their coordinates on the 5 first dimensions of the Principal Component Analyses to avoid biases induced with multicollinearity issues. The distances between individuals were calculated using Euclidean distance measure and clusters were formed using the Ward method (agglomerative coefficient of 0.82).