iVES OENO One vine & wine

DOI: https://doi.org/10.20870/oeno-one.2026.60.3.9987

OPEN ACCESS

VITICULTURE ORIGINAL RESEARCH ARTICLES

# Large-scale SIFT-MS volatile fingerprinting of *Vitis vinifera* berries grown under greenhouse semi-controlled conditions

Thomas Baerenzung dit Baron[1], Amal Chikri[1], Olivier Rodrigues[1], Alban Jacques[1], Aurélie Roland[2], Flora Tavernier[3], Charles Romieu[3], Vincent Segura[3], Dominique This[3], Roberto Bacilieri[3], Agnès Doligez[3], Thierry Simoneau[4], Aude Coupel-Ledru[4], Olivier Geffroy[1,*]

[1] Univ Toulouse, Ecole d'Ingénieurs de Purpan, PPGV, Toulouse, France
[2] SPO, Univ Montpellier, INRAE, Institut Agro, Montpellier 34060, France
[3] AGAP Institut, Université de Montpellier, CIRAD, INRAE, Institut Agro Montpellier, 34398 Montpellier, France
[4] LEPSE, Université de Montpellier, INRAE, Institut Agro, 34060 Montpellier, France

**Article number: 9987**

**correspondence:*
olivier.geffroy@purpan.fr

*Associate editor:*
Georgia Lytra

*Received:*
27 March 2026

*Accepted:*
22 August 2026

*Published:*
16 September 2026

This article is published under the **Creative Commons licence** (CC BY 4.0).

*Use of all or part of the content of this article must mention the authors, the year of publication, the title, the name of the journal, the volume, the pages and the DOI in compliance with the information given above.*

## ABSTRACT

Selected ion flow tube mass spectrometry (SIFT-MS) is a promising high-throughput technology to measure volatile organic compounds (VOCs) in real-time without the need for complex sample preparation or chromatographic separation. It could be useful to phenotype individuals based on their berry volatilome. However, no large-scale study has been conducted so far under greenhouse semi-controlled environmental conditions. This work aimed to assess the SIFT-MS volatilome of 141 *Vitis vinifera* varieties grown under these conditions and selected from a diversity panel distributed across three genetic groups (TE, Table East; WE, Wine East; WW, Wine West). Two-berry samples were analysed dynamically using $O_2^+$ as the reagent ion and a 40 min equilibrium phase at 70 °C. Variability in normalised SIFT-MS data per berry weight among replicate plants of the same variety was low despite notable differences in sugar concentration. Samples clustered into only two groups, with Cluster 1 (C1, n = 91) showing lower abundances for all *m/z* and more TE varieties, and Cluster 2 (C2, n = 50) exhibiting a greater richness in volatiles and including a larger proportion of WW varieties. This simple binary pattern may reflect the biological continuum of the studied panel, which may be particularly highlighted under controlled-environment growth conditions. Compared to C1, varieties in C2 also showed higher sugar content and smaller berry size. These preliminary findings highlight the potential of SIFT-MS for large-scale phenotyping of berries and genetic studies of volatile fingerprint, while calling for further validation, complementary GC-MS approaches, and investigation across berry maturation.

**KEYWORDS:** selected ion flow tube mass spectrometry, volatilome, genetic diversity panel, biological variability, clustering

## INTRODUCTION

Viticulture is facing growing environmental challenges, particularly increased drought stress driven by climate change, alongside rising societal pressure to reduce pesticide use (Mwaka *et al.*, 2024; van Leeuwen *et al.*, 2024). Breeding cultivars is one of the most effective strategies to cope with these issues, typically requiring intraspecific or interspecific crossings and a long selection process (Gambetta *et al.*, 2020; Pirrello *et al.*, 2023). While marker-assisted selection can be routinely applied to develop new varieties with tolerance to major fungal diseases such as downy mildew (*Plasmopara viticola*) and powdery mildew (*Erysiphe necator*) (Pirrello *et al.*, 2023), developing tolerant varieties to drought remains more difficult due to the multigenic nature of this trait (Gambetta *et al.*, 2020). Consequently, selecting for such complex traits requires long, time-consuming phenotyping steps (Töpfer *et al.*, 2011). Besides tolerance, berry aroma composition is another key character to take into consideration in the selection process as it contributes to wine quality and sensory perception (Zhu *et al.*, 2016).

The concentrations in aroma compounds can be as low as the ng $L^{-1}$ or ng $kg^{-1}$ range, meaning that sample preparation and analyte preconcentration are often as critical as the analytical step itself (Marín-San Román *et al.*, 2020). Regardless of the extraction method, analysis is mostly performed by gas chromatography (GC) or multidimensional gas chromatography (GC-GC) coupled with various detectors, including mass spectrometry (MS), flame ionisation detection (FID) or olfactometry (O) (Welke *et al.*, 2022). These GC approaches share several limitations, including long analysis times (45 min to 2 h) and often complex method development (Chin & Marriott, 2015). Chemical ionisation mass spectrometry (CIMS) techniques such as selected ion flow tube mass spectrometry (SIFT-MS) have been reported as alternatives to minimise sample preparation and enable gas-phase analysis within a few minutes (Smith & Španěl, 2005). This instrument relies on soft ionisation using up to eight reagent ions in the most recent systems ($H_3O^+$, $NO^+$, $O_2^+$, $NO_3^-$, $NO_2^-$, $O^-$, $O_2^-$, and $OH^-$) and allows analysis of sample headspace with determination of ion abundances. It can be operated either in scan mode, covering the full mass range of expected ions, or in selected ion monitoring (SIM) mode, in which specific masses are targeted to achieve quantification at the parts-per-billion (ppb) level (Smith & Španěl, 2005). The reagent ions are generated by a microwave plasma, and ion switching occurs within milliseconds.

The instrument has previously been applied in i) environmental studies, including air quality monitoring and the detection of hazardous chemicals in ambient air and soils, ii) healthcare for human breath analysis and the identification of disease-related biomarkers, and iii) food science for the detection of spoilage and for product quality assessment (Smith *et al.*, 2025).

Despite its broad range of applications, the use of SIFT-MS to analyse grape berries, wine, and other alcoholic beverages remains very limited. In that field, SIFT-MS has recently been applied to analyse twenty-three grape varieties collected from a germplasm collection (Baerenzung dit Baron *et al.*, 2022). This study demonstrated that the use of $O_2^+$ as a reagent ion allowed rapid discrimination of grape varieties based on their production levels of monoterpenols, C13-norisoprenoids, volatile phenols, small alcohols, and aldehydes. Interestingly, in most cases, varieties clustered according to their parentage relationships. Another field study conducted under warm and dry climatic conditions using the same model showed that the SIFT-MS volatilome remained stable from the second phase of maturation (Geffroy *et al.*, 2022). This stability facilitates sample collection, which can therefore rely solely on phenological stage and elapsed time after mid-veraison. More recently, it has been shown that this instrument can be used to assess the aroma potential of berries treated with β-glucosidase, and that hybrids genetically distant from *Vitis vinifera* (*e.g.*, Clinton, *Vitis riparia* × *Vitis labrusca*) can be easily distinguished (Baerenzung dit Baron *et al.*, 2026).

Such high-throughput technology could be an effective tool for phenotyping the large numbers of seedlings produced during breeding programs. During early growth stages, these individuals are typically grown in pots in a greenhouse, providing semi-controlled conditions. At later developmental stages, such conditions allow for the precise application of environmental stresses, particularly water stress, which is highly relevant in studies targeting adaptation strategies to climate change. To date, no research has examined the SIFT-MS volatilome of berries produced under these conditions or assessed the instrument's ability to discriminate between varieties.

This study aimed to investigate the SIFT-MS volatilome, defined here as the set of *m/z* signals detected in the berry headspace from a large set of *Vitis vinifera* cultivars (n = 141) selected from a diversity panel and grown under greenhouse conditions, leveraging the high-throughput phenotyping capacity of SIFT-MS as an attractive alternative to conventional targeted approaches for large-scale screening purposes. Due to the limited availability of grape material, no more than two berries were analysed per sample.

## MATERIALS AND METHODS

### 1. Plant material and sample collection

The five-year-old own-rooted plants used for this experiment were the same as those used and described in Coindre *et al.* (2026), who studied leaf ecophysiological traits. One hundred and forty-one genotypes were selected for their berry availability from a panel of 279 designed to represent the worldwide genetic diversity of *Vitis vinifera* (Nicolas *et al.*, 2016). The list of these genotypes, their belonging to three previously described subpopulations (*e.g.*, Wine West, WW; Wine East, WE; and Table East, EE), their accession ID (Vassal), and their skin colour are shown

in Table S1. It should be noted that such subpopulation differentiation, derived from microsatellite-based genetic diversity analyses, does not fully correspond to use patterns or geographic locations.

They were grown from May to August 2022 in a greenhouse in individual 9-L pots with a substrate composed of a mixture of loamy soil and organic compost on the PhenoArch high-throughput phenotyping platform at the Montpellier Plant Phenotyping Platforms (Cabrera-Bosquet *et al.*, 2016). Water was supplied at 1.05 g $H_2O$ $g^{-1}$ dry soil to maintain a moderate water deficit, simulating early field drought conditions with moderate effects on leaf functioning. For seven cultivars, namely Alexandroouli, Bou Khanzir noir, Camaraou noir, Corbeau, Kolontar, Morrastel, and Pagadebiti (Table S1), berries were sampled from two different plants over the same sampling period, each providing one pooled sample of two berries, to provide a preliminary assessment of the robustness of the SIFT-MS volatilome to inter-plant biological variability, not an exhaustive estimate of this variability for all studied genotypes. Such cultivars were selected on the sole basis of the availability of plant duplicates within the experimental setup.

Air temperature and relative humidity were continuously monitored using HC2-SH sensors (Rotronic, Bassersdorf, Switzerland), while photosynthetically active radiation (PAR) was measured using SKP 215 sensors (Skye Instruments, Llandrindod Well, United Kingdom), with instruments installed at ten locations throughout the greenhouse. Environmental conditions recorded during the June–July period, which corresponds to maturation, were 17 mol $day^{-1}$ for PAR, 1.58 kPa for average maximum daily vapour pressure deficit and 19.0 to 29.3 °C for mean daily air temperature. Exactly twenty-one days after the onset of veraison, as determined for each genotype, two berries per genotype were collected. This choice of two berries per sample was motivated by strong and deliberate experimental constraint given that the plant material came from single potted plants whose primary purpose was ecophysiological phenotyping (Coindre *et al.*, 2026) rather than berry production in quantity. The two berries collected per genotype were pooled into a single sample prior to headspace SIFT-MS acquisition. Because of genetic variability in phenology, berry harvest spanned eight periods: P1 (12–14 June), P2 (21–23 June), P3 (28 June–1 July), P4 (4–8 July), P5 (11–15 July), P6 (19–22 July), P7 (26–29 July), and P8 (3 August). All berries were stored at –18 °C in 50 mL Falcon tubes prior to analysis.

## 2. Chemicals

Ultra-pure water from a Milli-Q system (Millipore, Burlington, Massachusetts, USA) was used in the study. $N_2$ was used as the SIFT-MS carrier gas and to compensate for the pressure drop induced by aspiration (Alphagaz, Air Liquide, 99.9999 %, Paris, France). Certified gas standards for SIFT-MS calibration were provided by Air Liquide America Speciality Gases LLC (Plumsteadville, PA, USA). Tartaric acid (≥ 99.0 %), sodium tartrate (≥ 99.0 %), sodium hydroxide (≥ 98 %), and sodium chloride (≥ 99.5 %) were purchased from Thermo Fisher Scientific (Illkirch, France).

## 3. Basic berry parameters

For each variety, the two berries were weighed after thawing, pooled and homogenised together into a single sample using an Ultra-Turrax T18 molecular grinder (IKA, Staufen, Germany). The decision to pool the two berries was made to obtain a satisfactory analytical signal given the limited amount of material available. The resulting must was analysed for soluble solids content (°Brix) using a digital MA885 Wine Refractometer (Milwaukee, Wisconsin, USA) and for pH using a PHM 210 MeterLab pH meter (Radiometer, Copenhagen, Denmark).

## 4. Sample preparation for SIFT-MS analysis

Following measurement of basic berry parameters, fifteen mL of tartrate buffer (pH 3.2) were added per 3 g of berries, and samples were subjected to a second grinding step to ensure complete tissue disruption. The tartrate buffer (pH 3.2) was prepared by dissolving tartaric acid in ultra-pure water and adjusting the pH with NaOH. Extracts were transferred to 100 mL Drechsel-style gas washing bottles (Servilab, Le Mans, France). Sodium chloride was then added (4.5 g per 3 g of berries, corresponding to 300 $g \cdot L^{-1}$), and the mixture was sonicated for 15 min and heated at 70 °C for 40 min under continuous rotation (100 rpm). Under these conditions, without the use of exogenous β-glycosidase enzymes, SIFT-MS measures free volatile compounds rather than bound forms or precursors.

## 5. SIFT-MS analysis

Following this equilibrium period, the bottle containing the extract was connected on one side to a VOICE 200 Ultra instrument (Syft Technologies, Christchurch, New Zealand) and on the other side to a nitrogen source to enable dynamic analysis. The instrument was operated using nitrogen as the carrier gas in full scan mode (*m/z* range: 15–250) with $O_2^+$ as previously proposed (Baerenzung dit Baron *et al.*, 2022). Samples were introduced through a calibrated capillary at 0.3 $Torr \cdot L \cdot s^{-1}$. Inlet and aspiration flow rates were balanced to minimise vacuum. The internal SIFT-MS flow rate was set at 2.0 $Torr \cdot L \cdot s^{-1}$ with the reaction between the analytes and the selected precursor occurring in the flow tube, maintained at 119 °C and 0.06 kPa. Instrumental repeatability and reproducibility over 3 consecutive days were estimated at 7 % and 10 %, respectively. For each sample, four full mass scans were recorded, of which the last three were retained for analysis.

Data acquisition and analysis were performed using LabSyft 1.6.2 software (Syft Technologies). A calibration procedure was performed before each instrument start to ensure linearity, range, accuracy, and precision by injecting a mix of certified gas standards.

In comparison with the previously published method (Baerenzung dit Baron *et al.*, 2022), the modifications in container volume (100 mL *vs* 1 L), temperature (40 °C *vs* 70 °C), and sampling mode (dynamic *vs* static) were implemented based on preliminary experiments to get

sufficient volatiles in the headspace, detectable and repeatable signal from a limited amount of material (2 berries in most cases *vs* 50 g). Triplicate analysis of a thawed *Vitis vinifera* L. cv. Syrah grape sample using the two systems showed a significant correlation ($p$-value < 0.001) for abundances measured in the *m/z* range 100-250, with Pearson and Spearman correlation coefficients of 0.44 and 0.60, respectively. This indicates a strong monotonic relationship between the two methods, meaning that compounds with high abundance in one model tend to have high abundance in the other, even though the relationship is not strictly linear or proportional. The absence of a stronger relationship may reflect the higher concentrations at 70 °C in the headspace of heavier, less volatile compounds, such as sesquiterpenes (Copolovici & Niinemets, 2015) (data not shown).

### 6. Data treatment & statistics

Averaged SIFT-MS data across the three technical replicates were pre-processed by removing masses with *m/z* below 100 and abundances below the noise threshold, following the methodology described by Baerenzung dit Baron *et al.* (2022). Signal intensities for 128 different *m/z* ratios, collectively referred to hereafter as the berry volatilome or volatile fingerprint, were standardised by berry weight to enable comparison between cultivars. These *m/z* values correspond to raw, unannotated SIFT-MS signals rather than to formally identified volatile compounds or characterised ion fragments. For each cultivar with plant duplicates, Spearman correlation coefficients and relative differences (median and standard deviation) between replicates were calculated to assess biological variability in the SIFT-MS data following recommended practices for mass spectrometry-based metabolomics data (Ghosh *et al.*, 2021). For variability in basic berry parameters, absolute differences were calculated for sugar concentration and pH, while relative differences were used for berry weight.

The data, including those averaged across replicates for the variety concerned, were then subjected to Principal Component Analysis (PCA), followed by Hierarchical Clustering Analysis (HCA) using Euclidean distance and Ward's method, a common approach for multivariate analysis and clustering (Hair, 2011). The optimal number of clusters was chosen to maximise the silhouette and Calinski–Harabasz indices. One-way ANOVA and chi-square test followed by a Marascuilo *post hoc* test were performed to detect whether abundances, basic berry parameters, and the proportion of genotypes belonging to the three genetic subpopulations collected over the eight sampling periods significantly differed between the clusters identified by HCA. For ANOVA performed on abundances, $p$-values were adjusted for multiple testing using the false discovery rate (FDR) method. A heatmap was also generated from the dataset, an approach previously implemented for SIFT-MS data for overall visualisation purposes (Dryahina *et al.*, 2016), with hierarchical clustering applied to both samples and variables. All data analyses were performed using XLSTAT 3.1 software (Addinsoft, Paris), except for PCA and the heatmap, which were created using ClustVis and DataMap (Ge, 2025; Metsalu & Vilo, 2015).

## RESULTS AND DISCUSSION

### 1. Basic berry parameters and SIFT-MS data for the 141 genotypes and their variability across biological replicates

Average basic berry parameters determined on the sampled berries for the 141 genotypes provide an overview of grape maturity at the time of harvest and enable contextualisation of SIFT-MS data. They were 14.1 °Brix ± 3.2 for sugar concentration, 3.64 ± 0.33 for pH, and 3.60 g ± 1.78 for berry weight (Table S1). To our knowledge, no other study has been conducted under a comparable experimental context and sampling timing. However, the sugar concentrations obtained are consistent with those reported in a field trial, where berries sampled 30 days after veraison showed sugar concentrations between 17 and 18.5 °Brix (Geffroy *et al.*, 2014). These results highlight an overall low level of maturation, which might be the consequence of the early sample collection that occurred 21 days after the onset of veraison. The decision to sample individual berries within this post-veraison delay was supported by an earlier field study demonstrating that such a period was adequate for sampling single berries within their active period of sugar accumulation, when sugar concentration rapidly increases from 200 to 1000 mM (*i.e.*, from 3.5 to 18.5 °Brix) (Tavernier *et al.*, 2025), as confirmed here. The discrepancies in sugar concentration across cultivars may be largely explained by the unknown delay between the veraison of the first berry and the sampled one. However, this sampling procedure warrants that samples did not reach the berry shrivelling period after the cessation of phloem import (Savoi *et al.*, 2021), which is marked by abrupt transcriptomic and metabolic changes (Tavernier *et al.*, 2025). It should be noted that, although this sampling time point was not specifically investigated for VOC dynamics, it reflects a trade-off between physiological relevance and experimental constraints, as the berries were also used for additional phenotypic measurements.

The relatively high average berry weight, a trait influenced not only by genotype but also by environmental factors such as water availability (Dai *et al.*, 2011), with maximum values reaching 12.2 g for Angoor Kalan, suggests that water was not limiting under the conditions of this experiment. It should be noted that pH values may have been biased and overestimated, because berries were stored at –18 °C prior to analysis (Santesteban *et al.*, 2013). However, such a process is expected to remove proportionally more $K^+$ than tartaric acid, leading to a decrease in pH. The relatively high pH values observed may therefore result from intense tissue disruption during homogenisation and/or elevated potassium levels, potentially linked to fertilisation conditions.

These low sugar levels may have influenced the SIFT-MS data collected, which have previously been shown to remain relatively stable during the later stages of maturation, particularly from 28 days after mid-veraison under field conditions, although under specific sample handling conditions (Geffroy *et al.*, 2022). However, direct comparisons of timing are difficult, as the earlier study assessed mixed berry populations, whereas our work focused on two berries. While a formal study under greenhouse conditions would be needed to evaluate the effect of post-veraison timing fully, preliminary insights can be drawn from the metrics used to assess biological variability, as presented in Table 1.

The seven replicate plants enabled assessment of the biological variability for basic berry parameters and SIFT-MS data (Table 1). Sugar concentration showed greater variability for Camaraou and Pagadebiti, with an absolute difference of 4.9 °Brix between replicates. However, despite this variability in sugar content, Pagadebiti exhibited low variability across most other metrics, particularly those related to SIFT-MS signal intensities. This apparent decoupling between sugar concentration and SIFT-MS measurements suggests that, within a given genotype, the berry volatilome is only weakly influenced by sugar concentration within the studied range.

It must be pointed out that Alexandroouli also displayed low overall variability, except for pH. The higher variability observed for this parameter suggests that pH values should be interpreted with caution, as discussed previously.

Interestingly, the variability among plants was low for the SIFT-MS data, as reflected by high Spearman correlation coefficients ranging from 0.9632 for Kolontar to 0.9916 for Pagadebiti, along with low relative differences, with median values varying between 16.0 % for Pagadebiti and 44.0 % for Camaraou. Although the moderate inter-plant variability observed for the seven duplicated cultivars suggests reasonable reproducibility of the SIFT-MS signal despite the limited number of berries analysed per sample, results obtained at the individual cultivar level should be interpreted with caution, as two berries cannot fully capture the biological variability inherent to a given genotype. A higher number of biological replicates would be required to support any fine-scale varietal characterisation. Overall, this reproducibility across biological replicates complements the analytical reliability of each individual measurement, which itself results from averaging three consecutive technical mass scans per sample after discarding the first scan for signal stabilisation, with an estimated instrumental repeatability of 7 %

## 2. PCA and HAC clustering results

Principal Component Analysis (PCA) was performed on the full set of 128 *m/z* ratios retained after filtering, in order to explore in an unsupervised manner whether genotypes could be structured into distinct groups based on their overall SIFT-MS volatilome. Hierarchical Clustering Analysis (HCA), applied to the resulting PCA coordinates, identified two genotype clusters (C1 and C2), presented in Figures S1 and S2 and summarised in Table S1. A heatmap was additionally generated directly from the full dataset, independently of the PCA/HCA workflow, providing a complementary visualisation of *m/z* abundance patterns across samples. Minor differences in sample groupings observed between the two approaches likely reflect the heatmap's greater sensitivity to a small number of dominant *m/z* values, whereas PCA/HCA integrates variance across the full set of variables. Surprisingly, given the genetic variability of the studied varieties and the expected distinct composition in aroma compounds arising from it (Robinson *et al.*, 2014), only two homogeneous clusters, hereafter referred to as Cluster 1 (C1, n = 91) and Cluster 2 (C2, n = 50), were identified by PCA followed by HAC. In previous fieldwork conducted on 23 varieties sampled from a germplasm collection, the same method identified four distinct clusters, highlighting more

**TABLE 1.** Within-cultivar variability of basic berry parameters and SIFT-MS data between biological replicates (n = 2). Metrics are expressed in absolute difference for sugar concentration and pH, and in relative difference for berry weight. RD stands for relative difference.

| Full cultivar name | Basic berry parameters | | | SIFT-MS abundance data | | |
|---|---|---|---|---|---|---|
| | Sugar content (°Brix) | pH (pH unit) | Berry weight (%) | $R_{Spearman}$ | RD (%) Median | RD (%) Standard deviation |
| Alexandroouli | 0.6 | 0.61 | 12.5 | 0.9882 | 18.6 | 17.7 |
| Bou Khanzir noir | 2.0 | 0.03 | 47.9 | 0.9763 | 27.4 | 20.9 |
| Camaraou noir | 4.9 | 0.13 | 26.0 | 0.9849 | 44.0 | 25.0 |
| Corbeau | 0.0 | 0.01 | 33.1 | 0.9787 | 22.8 | 25.0 |
| Kolontar | 3.7 | 0.13 | 33.9 | 0.9632 | 41.9 | 32.9 |
| Morrastel | 1.7 | 0.15 | 11.5 | 0.9656 | 30.6 | 26.2 |
| Pagadebiti | 4.9 | 0.23 | 8.6 | 0.9916 | 16.0 | 15.9 |

pronounced differences within the dataset (Baerenzung dit Baron *et al.*, 2022). Rather than indicating a lack of discrimination, the observed binary structure may reflect the existence of a biological continuum within the studied dataset. Such a continuum is consistent with the genetic design of the panel specifically selected to represent a wide range of *Vitis vinifera* diversity (Nicolas *et al.*, 2016). For genetically diverse populations, metabolomic traits often vary along gradients rather than forming discrete classes (Hall, 2006). This continuum can also be observed in the PCA plot with confidence ellipses for the two clusters slightly overlapping (Figure 1A). In addition, the relatively homogeneous environmental conditions (temperature, light, water, and nitrogen supply) under which plants were grown may have revealed a limited level of genetic variability in the berry volatilome, as environmental effects known to influence aroma biosynthesis (van Leeuwen *et al.*, 2020) were minimised. Furthermore, extraction under buffered conditions when preparing samples for SIFT-MS analysis may have contributed to reducing additional sources of variability.

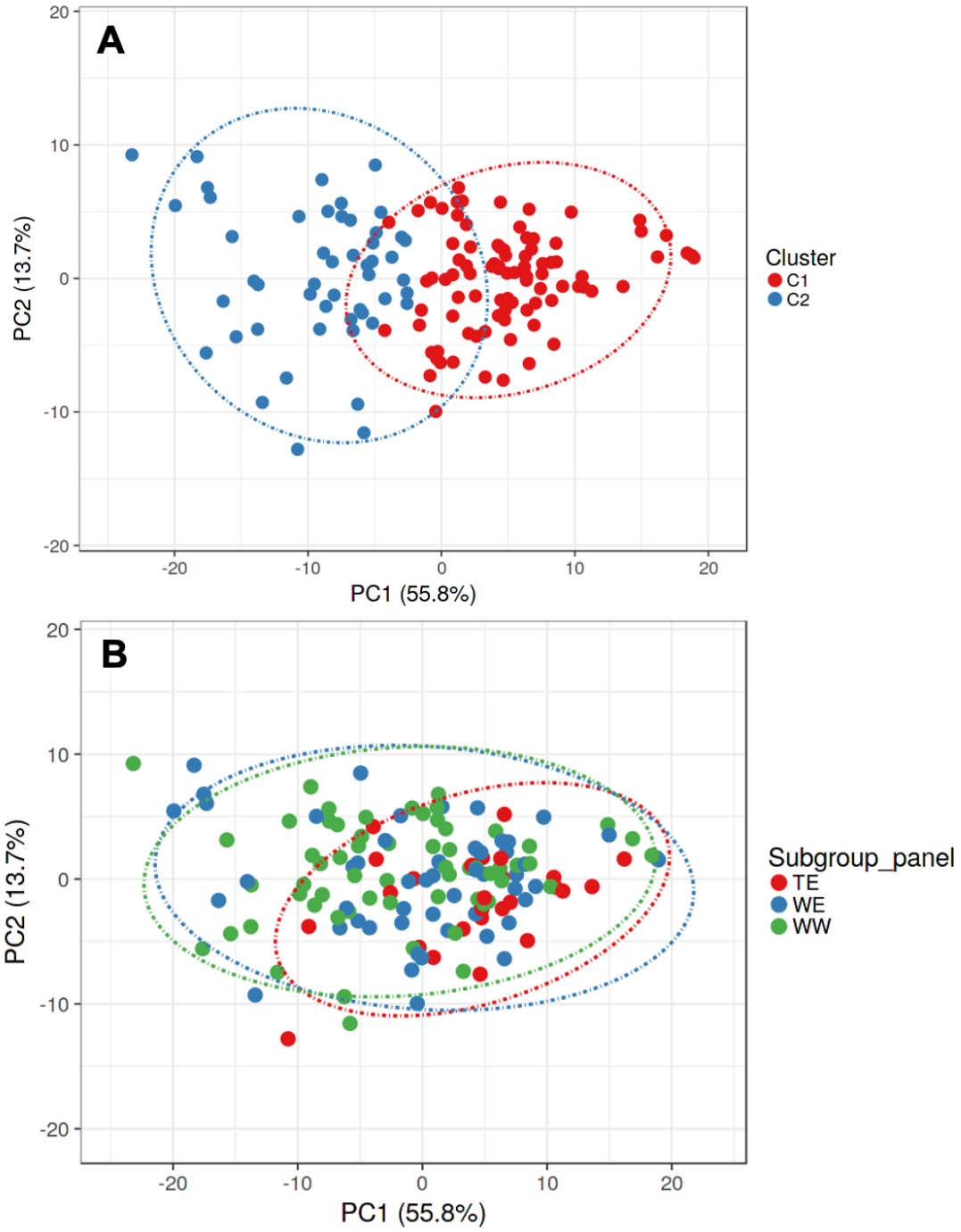



**FIGURE 1.** Factor scores with 95 % confidence ellipse for a principal component analysis (PCA) performed on the SIFT-MS abundance data using $O_2^+$ (n = 141). Ellipses represent groups by A) cluster identified through PCA followed by HAC and B) subgroup panel.

It can be pointed out that the obtained PCA provides a reliable representation of the sample distribution, with PC1 and PC2 explaining 69.5 % of the total variance. PC1 showed strong positive loadings for most *m/z* variables, suggesting that it captures overall signal intensity variation across samples, while PC2 differentiated groups of *m/z* values with contrasting contributions, reflecting compositional differences in the volatilome.

## 3. SIFT-MS spectral results for the two clusters identified

The mean SIFT-MS spectrum obtained for the two clusters (Figure 2) enabled the visualisation of differences in *m/z* abundance between the two clusters. Overall, the two spectra show comparable general patterns, with abundances showing the highest values at low *m/z* (100–150), gradually decreasing toward higher *m/z*, with *m/z* above 200 showing the lowest signal. Overall, this pattern is consistent with previous studies (Baerenzung dit Baron *et al.*, 2022; Geffroy *et al.*, 2022), which demonstrated the ability of SIFT-MS to discriminate between low and high producers of volatile compounds, although the peaks observed around *m/z* 196-201 have not been previously described.

SIFT-MS signals observed with $O_2^+$ can be related to molecular ions ($M^+$) formed by charge transfer, fragments or water adducts ($M^+ \cdot (H_2O)_n$) (Smith & Španěl, 2005). The peaks around *m/z* 196–201 could correspond to molecular ions of sesquiterpenes, oxygenated terpenols (*e.g.*, geranyl acetate or linalyl acetate), or C13-norisoprenoids, some free compounds found in grapes (Robinson *et al.*, 2014) with nominal masses in this *m/z* range whose relatively low volatility due to higher molecular mass may have been increased by change in methodology notably in equilibrium temperature (70 °C *vs* 40 °C).

For the 128 *m/z* ratios studied, with the exception of *m/z* 114 for which no significant difference was observed, abundances were significantly higher (*adjusted* $p < 0.001$, FDR) for C2 with an average 1.91-fold increase (± 0.20) compared to C1. Among the most cluster-specific signals, *m/z* 140, 149, 105, 191, and 159 showed the strongest differential abundances, potentially reflecting variety-specific compounds that may be tentatively associated with fragments or molecular ions of terpenoids or norisoprenoids, as reported in previous SIFT-MS studies (Baerenzung dit Baron *et al.*, 2022). In contrast, *m/z* 114, 214, and 172 were the most stable abundances, suggesting signals related to core metabolites common to both clusters.

The molecular identification of these SIFT-MS markers remains challenging. Each observed signal may arise from a molecular ion, a fragment, or a rearrangement product, and compounds of the same chemical family with similar masses can produce identical fragment ions (Baerenzung dit Baron *et al.*, 2022). Even standards provide limited guidance, as they cannot capture all fragmentation pathways or overlapping signals. Moreover, SIFT-MS signals can appear ± 1–3 *m/z* from the nominal value due to multiple product ions and isotopologue contributions, and unit mass resolution can shift peaks relative to theoretical masses (Smith & Španěl, 2005). Further work using GC-MS would be valuable, although such an approach would be challenging to implement at such a large scale. In the same way, the assessment of genetic associations between these cluster-specific signals and molecular markers characterising the diversity panel would be useful to further investigate these relationships.

## 4. Sub-population analysis and further discussion

The proportion of samples by skin colour, subgroup panel, and sampling period (P1 to P8) for the two clusters (Table 2) enabled the discussion of the observed clustering. Interestingly, C2 included a larger proportion of varieties from the WW subgroup and a lower proportion from the TE subgroup (Table 2), suggesting that wine-oriented varieties may exhibit higher overall concentrations of aroma compounds, although these subpopulation assignments do not fully correspond to use categories. However, this pattern is not clearly visible in the heatmap (Figure S2); the PCA plot, where samples tend to be grouped by subgroup (Figure 1B), is consistent with this assumption and reveals a similar pattern for the WE subgroup. A one-way ANOVA using panel subpopulation as a qualitative factor revealed that 39 *m/z* ratios were significantly different (*adjusted* $p < 0.05$, FDR), with TE generally showing lower abundances, WW higher values, and WE intermediate levels (results not shown).

A first explanation could be that centuries of wine grape selection favoured genotypes with higher concentrations of aroma compounds, whereas breeding of table grapes primarily targeted other traits. A second hypothesis relates to differences in berry maturity and higher average sugar concentration observed in C2 varieties (15.1 °Brix ± 3.0 *vs* 13.6 °Brix ± 3.3 for C1, $p < 0.01$), consistent with higher pH values (3.77 ± 0.27 *vs* 3.57 ± 0.34 for C1, $p < 0.01$), which could have enhanced the richness in aroma compounds. However, this explanation appears unlikely, as it has been demonstrated that most *m/z* signals decreased during maturation, possibly due to volatile emission or the accumulation of bound, non-detectable

**TABLE 2.** Proportion (expressed as %) of samples by skin colour, subgroup panel, and sampling period (from P1 to P8) for the two clusters identified by PCA followed by HCA, and significance (*P*) calculated from the chi-square test.

| Cluster | White | Red | TE | WE | WW | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C1 | 52 | 43 | 24 | 42 | 34 | 2 | 4 | 5 | 15 | 35 | 27 | 9 | 1 |
| C2 | 56 | 44 | 8 | 30 | 62 | 0 | 2 | 10 | 16 | 12 | 40 | 16 | 4 |
| *P* | 0.620 | 0.896 | 0.018 | 0.168 | <0.01 | 0.291 | 0.462 | 0.319 | 0.903 | <0.01 | 0.127 | 0.197 | 0.253 |

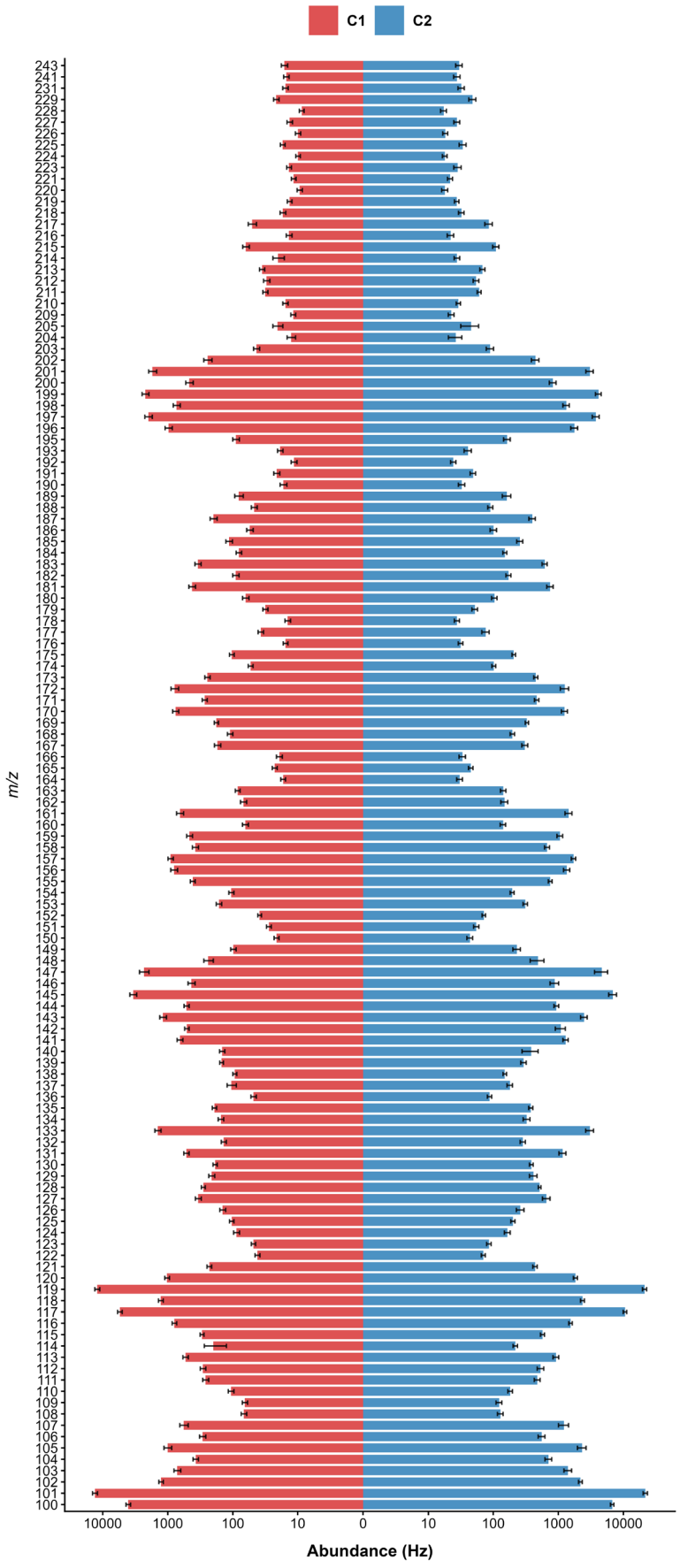



**FIGURE 2.** Mean SIFT-MS spectrum obtained with $O_2^+$ reagent ion for all the sampled varieties combined in C1 (n = 91) and C2 (n = 50). Error bars are 95 % confidence intervals. The y-axis is displayed on a logarithmic scale.

forms (Geffroy *et al.*, 2022). Moreover, this interpretation is inconsistent with our previous discussion based on variability metrics, which suggested only a limited influence of sugar concentration on SIFT-MS volatile fingerprints. A third and more plausible explanation lies in the significant difference in average berry weight between clusters (2.52 g ± 0.78 for C2 *vs* 4.18 g ± 1.90 for C1, $p < 0.001$). Indeed, the TE subpopulation has already been shown to exhibit higher berry weights than WW (Nicolas *et al.*, 2016), a pattern that was confirmed in our study, where berry weight decreased significantly ($p < 0.001$) from TE (5.46 ± 2.64 g) to WE (3.60 ± 1.35 g) and WW (2.81 ± 0.86 g). As SIFT-MS data were standardised per berry weight, smaller berries are likely to contain a larger proportion of skin per unit mass, which may have induced higher signals. Indeed, the berry skin is known to be particularly rich in phenolic and aroma compounds (Urcan *et al.,* 2016). This latter hypothesis does not fully contradict the first one, as historical selection for smaller berries in wine grapes may have favoured higher concentrations of these compounds.

To obtain preliminary insights, although the number of seeds was not recorded in our experiment despite being a key parameter known to influence berry weight (Dai *et al.*, 2011), berry volume (V) was estimated for each variety based on sugar concentration and corresponding density values reported in OIV tables (OIV, 2021). Berry diameter (d) was then calculated using the formula $V = \pi d^3/6$. From this estimation, berry skin surface area (S) was derived using $S = \pi d^2$, allowing the determination of skin surface per unit mass. This variable significantly discriminated between clusters ($p < 0.001$), with values of 3.02 $cm^2\ g^{-1}$ ± 0.42 for C1 and 3.50 $cm^2\ g^{-1}$ ± 0.37 for C2. Overall, C2 exhibited values approximately 1.16-fold higher than C1. However, this difference is markedly smaller than the 1.91-fold increase observed in SIFT-MS abundances across all *m/z*. This suggests that skin surface area alone cannot fully account for the differences in abundances between clusters. Together, these results support the hypothesis of enhanced overall volatile compound production in varieties belonging to the C2 cluster, which includes a greater proportion of WW varieties.

It is also worth noting that cluster C1 included a significantly higher proportion of varieties sampled at P5, whereas the greatest proportion of C2 were harvested later at P6. This pattern suggests that C1 varieties reached maturity earlier in the season, which is consistent with previous observations for table grape varieties (Nicolas *et al.*, 2016) that are more frequently represented in C1. Therefore, C1 varieties may have also been exposed to cooler conditions during maturation. The absence of significant differences between clusters in the distribution of varieties across the other sampling periods, together with the lack of clear clustering in the heatmap (Figure S2), further supports the hypothesis that environmental conditions had a limited impact on the dataset, as expected under semi-controlled conditions. Consistent with this, a one-way ANOVA using sampling period as a qualitative factor revealed a significant but limited effect on sugar concentration ($p < 0.05$), driven by higher values at P6 (15.4 ± 3.0 °Brix) and P7 (15.4 ± 3.7 °Brix) compared to P3 (11.8 ± 3.2 °Brix), with no significant differences among the remaining periods. This moderate increase likely reflects the progressive rise in greenhouse temperature over the June-August period. pH also varied significantly across periods, with the highest values at P1 (4.10 ± 0.00, n = 2) and the lowest, as for sugar concentration, at P3 (3.36 ± 0.46). However, unlike sugar concentration, pH did not follow a comparable temperature-related trend across the season, suggesting that this period effect likely reflects differences in the specific set of genotypes sampled within each period. Such effects are unlikely to have affected the SIFT-MS-based clustering, given the weak coupling between sugar concentration and volatilome already noted among replicate plants (Table 1).

Finally, it should be noted that the thermal treatment applied (70 °C, 40 min) to obtain a detectable and reproducible headspace signal from a limited amount of berry material may have altered the native volatilome of fresh berries, notably through partial thermal hydrolysis of esters, degradation of heat-sensitive sulphur compounds, and the possible formation of newly generated compounds via Strecker/Maillard-type reactions. As a result, the SIFT-MS volatile fingerprints obtained here are not strictly representative of fresh berry aroma. However, since all samples underwent a rigorously identical thermal treatment, the relative comparisons between cultivars and clusters which form the core of the present results remain valid.

## CONCLUSION

This large-scale study provides the first comprehensive assessment of the SIFT-MS berry volatilome across 141 varieties grown under semi-controlled greenhouse conditions. Interestingly, despite some varieties showing substantial differences in sugar content, the overall variability in SIFT-MS data among duplicate plants of the same variety was low. The samples clustered into two groups, namely C1 (n = 91) characterised by lower *m/z* abundances and a higher representation of TE subgroup varieties, and C2 (n = 50) richer in volatiles and including a larger proportion of WW varieties. Varieties in C2 also exhibited higher sugar content and particularly smaller berry size. Although semi-controlled growth conditions are particularly well suited to enhance the detection of genotype-driven differences by minimising environmental variability, the clear binary grouping observed may be specific to the panel studied and reflect an underlying biological continuum. It should also be acknowledged that the experimental conditions used in this study, a semi-controlled greenhouse environment, potted plants, and young, five-year-old vines, differ substantially from those encountered in conventional field viticulture, which represents an important limitation to bear in mind when interpreting and generalising the present results. These findings should be considered preliminary and interpreted with caution at the individual cultivar level, given the limited number of berries analysed per sample, and a higher number of biological replicates would be required for any fine-scale varietal characterisation. It should also be kept in mind that the thermal treatment applied to samples (70 °C, 40 min) means that

the SIFT-MS volatile fingerprints reported here do not strictly reflect the volatile composition of fresh berries. Further studies are needed to evaluate the dynamics of volatilome evolution during berry maturation under greenhouse conditions, and to complement these results through GC-MS approaches. Overall, this work highlights the potential of this high-throughput approach for large-scale phenotyping and future genetic studies aimed at deciphering the determinants of aroma compound abundance.

## ACKNOWLEDGEMENTS

The authors would like to thank the Plastivigne flagship project of the "key challenge Vinid'Occ", a project funded by the Occitanie Region and led by the University of Montpellier (France), and the G2WAS project (ANR-19CE20-0024) for supporting the greenhouse experiment. They also acknowledge colleagues from the LEPSE laboratory for their technical support.

## REFERENCES

Baerenzung dit Baron, T., Yobregat, O., Jacques, A., Simon, V., & Geffroy, O. (2022). A novel approach to discriminate the volatilome of Vitis vinifera berries by selected ion flow tube mass Spectrometry analysis and chemometrics. *Food Research International*, *157*, 111434. https://doi.org/10.1016/j.foodres.2022.111434

Baerenzung dit Baron, T., Rigou, P., Ducasse, M-A., Cesson, A., Usta, MA., Mouret, J-R., Jacques, A., & Geffroy, O. (2026). SIFT-MS volatilome fingerprinting for rapid screening of aroma precursor potential in grapes: a case study on *Vitis vinifera* and hybrid varieties. *Applied Food Research*, in press.

Cabrera-Bosquet, L., Fournier, C., Brichet, N., Welcker, C., Suard, B., & Tardieu, F. (2016). High-throughput estimation of incident light, light interception and radiation-use efficiency of thousands of plants in a phenotyping platform. *New Phytologist*, *212*(1), 269-281. https://doi.org/10.1111/nph.14027

Chin, S.-T., & Marriott, P. J. (2015). Review of the role and methodology of high resolution approaches in aroma analysis. *Analytica Chimica Acta*, *854*, 1-12. https://doi.org/10.1016/j.aca.2014.06.029

Coindre, E., Boulord, R., Chir, L., Freitas, V., Ryckewaert, M., Laisné, T., Bouckenooghe, V., Lis, M., Cabrera-Bosquet, L., & Doligez, A. (2026). Robustness of high-throughput prediction of leaf ecophysiological traits using near infrared spectroscopy and poro-fluorometry. *The Plant Phenome Journal*, *9*(1), e70059. https://doi.org/10.1002/ppj2.70059

Copolovici, L., & Niinemets, Ü. (2015). Temperature dependencies of Henry's law constants for different plant sesquiterpenes. *Chemosphere*, *138*, 751-757. https://doi.org/10.1016/j.chemosphere.2015.07.075

Dai, Z. W., Ollat, N., Gomès, E., Decroocq, S., Tandonnet, J.-P., Bordenave, L., Pieri, P., Hilbert, G., Kappel, C., & van Leeuwen, C. (2011). Ecophysiological, genetic, and molecular causes of variation in grape berry weight and composition: a review. *American Journal of Enology and Viticulture*, *62*(4), 413-425. https://doi.org/10.5344/ajev.2011.10116

Dryahina, K., Sovová, K., Nemec, A., & Španěl, P. (2016). Differentiation of pulmonary bacterial pathogens in cystic fibrosis by volatile metabolites emitted by their in vitro cultures: *Pseudomonas aeruginosa*, *Staphylococcus aureus*, *Stenotrophomonas maltophilia* and the *Burkholderia cepacia* complex. *Journal of Breath Research*, 10(3), 037102. https://doi.org/10.1088/1752-7155/10/3/037102

Gambetta, G. A., Herrera, J. C., Dayer, S., Feng, Q., Hochberg, U., & Castellarin, S. D. (2020). The physiology of drought stress in grapevine: towards an integrative definition of drought tolerance. *Journal of Experimental Botany*, *71*(16), 4658-4676. https://doi.org/10.1093/jxb/eraa245

Ge, X. (2025). DataMap: A Portable Application for Visualizing High-Dimensional Data. *arXiv preprint arXiv:2504.08875*.

Geffroy, O., Baerenzung dit Baron, T., Yobrégat, O., Denat, M., Simon, V., & Jacques, A. (2022). The SIFT-MS fingerprint of *Vitis vinifera* L. cv. Syrah berries is stable over the second part of maturation under warm conditions of climate. *OENO One*, *56*(4), 139-146. https://doi.org/10.20870/oeno-one.2022.56.4.7066

Geffroy, O., Dufourcq, T., Carcenac, D., Siebert, T., Herderich, M., & Serrano, E. (2014). Effect of ripeness and viticultural techniques on the rotundone concentration in red wine made from V itis vinifera L. cv. D uras. *Australian Journal of Grape and Wine Research*, *20*(3), 401-408. https://doi.org/10.1111/ajgw.12084

Ghosh, T., Philtron, D., Zhang, W., Kechris, K., & Ghosh, D. (2021). Reproducibility of mass spectrometry based metabolomics data. *BMC Bioinformatics*, *22*(1), 423. https://doi.org/10.1186/s12859-021-04336-9

Hair, J. F. (2011). Multivariate data analysis: An overview. *International Encyclopedia of Statistical Science*, 904-907.

Hall, R. D. (2006). Plant metabolomics: from holistic hope, to hype, to hot topic. *New Phytologist*, *169*(3), 453-468. https://doi.org/10.1111/j.1469-8137.2005.01632.x

Marín-San Román, S., Rubio-Bretón, P., Pérez-Álvarez, E. P., & Garde-Cerdán, T. (2020). Advancement in analytical techniques for the extraction of grape and wine volatile compounds. *Food Research International*, *137*, 109712. https://doi.org/10.1016/j.foodres.2020.109712

Metsalu, T., & Vilo, J. (2015). ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Research*, *43*(W1), W566-W570. https://doi.org/10.1093/nar/gkv468

Mwaka, O., Mwamahonje, A., Nene, W., Rweyemamu, E., & Maseta, Z. (2024). Pesticides use and its effects on grape production: A review. *Sustainable Environment*, *10*(1), 2366555. https://doi.org/10.1080/27658511.2024.2366555

Nicolas, S. D., Péros, J.-P., Lacombe, T., Launay, A., Le Paslier, M.-C., Bérard, A., Mangin, B., Valière, S., Martins, F., & Le Cunff, L. (2016). Genetic diversity, linkage disequilibrium and power of a large grapevine (*Vitis vinifera* L) diversity panel newly designed for association studies. *BMC Plant Biology*, *16*(1), 74. https://doi.org/10.1186/s12870-016-0754-z

OIV (2021). *Compendium of international methods of analysis of wines and musts.* Organisation Internationale de la Vigne et du Vin, Paris, France.

Pirrello, C., Magon, G., Palumbo, F., Farinati, S., Lucchin, M., Barcaccia, G., & Vannozzi, A. (2023). Past, present, and future of genetic strategies to control tolerance to the main fungal and oomycete pathogens of grapevine. *Journal of Experimental Botany*, *74*(5), 1309-1330. https://doi.org/10.1093/jxb/erac487

Robinson, A. L., Boss, P. K., Solomon, P. S., Trengove, R. D., Heymann, H., & Ebeler, S. E. (2014). Origins of Grape and Wine Aroma. Part 1. Chemical Components and Viticultural Impacts. *American Journal of Enology and Viticulture*, *65*(1), 1-24. https://doi.org/10.5344/ajev.2013.12070

Santesteban, L. G., Miranda, C., & Royo, J. B. (2013). Influence of the freezing method on the changes that occur in grape samples after frozen storage. *Journal of the Science of Food and Agriculture*, *93*(12), 3010-3015. https://doi.org/10.1002/jsfa.6133

Savoi, S., Torregrosa, L., & Romieu, C. (2021). Transcripts switched off at the stop of phloem unloading highlight the energy efficiency of sugar import in the ripening *V. vinifera* fruit. *Horticulture Research*, *8*. https://doi.org/10.1038/s41438-021-00628-6

Smith, D., & Španěl, P. (2005). Selected ion flow tube mass spectrometry (SIFT-MS) for on-line trace gas analysis. *Mass Spectrometry Reviews*, *24*(5), 661-700. https://doi.org/10.1002/mas.20033

Smith, D., Španěl, P., Demarais, N., Langford, V. S., & McEwan, M. J. (2025). Recent developments and applications of selected ion flow tube mass spectrometry (SIFT-MS). *Mass spectrometry Reviews*, 44(2), 101-134. https://doi.org/10.1002/mas.21835

Tavernier, F., Motelica-Heino, E., Thomas, M., Herbold, T., Shi, M., Le Cunff, L., Romieu, C., & Segura, V. (2025). In situ near-infrared non-destructive monitoring of sugar accumulation reveals that single berries ripening takes only 20 days. *Quantitative Plant Biology*, 1-25. https://dx.doi.org/10.1017/qpb.2025.10025

Töpfer, R., Hausmann, L., Harst, M., Maul, E., Zyprian, E., & Eibach, R. (2011). New horizons for grapevine breeding. *Fruit, Vegetable and Cereal Science and Biotechnology*, *5*(1), 79-100.

Urcan, D. E., Lung, M.-L., Giacosa, S., Torchio, F., Ferrandino, A., Vincenzi, S., Río Segade, S., Pop, N., & Rolle, L. (2016). Phenolic substances, flavor compounds, and textural properties of three native Romanian wine grape varieties. *International Journal of Food Properties*, *19*(1), 76-98. https://doi.org/10.1080/10942912.2015.1019626

van Leeuwen, C., Barbe, J.-C., Darriet, P., Geffroy, O., Gomès, E., Guillaumie, S., Helwi, P., Laboyrie, J., Lytra, G., Le Menn, N., Marchand, S., Picard, M., Pons, A., Schüttler, A., & Thibon, C. (2020). Recent advancements in understanding the terroir effect on aromas in grapes and wines: This article is published in cooperation with the XIIIth International Terroir Congress November 17-18 2020, Adelaide, Australia. Guests editors: Cassandra Collins and Roberta De Bei. *OENO One*, *54*(4). https://doi.org/10.20870/oeno-one.2020.54.4.3983

van Leeuwen, C., Sgubin, G., Bois, B., Ollat, N., Swingedouw, D., Zito, S., & Gambetta, G. A. (2024). Climate change impacts and adaptations of wine production. *Nature Reviews Earth & Environment*, *5*(4), 258-275. https://doi.org/10.1038/s43017-024-00521-5

Welke, J. E., Nicolli, K. P., Hernandes, K. C., Biasoto, A. C. T., & Zini, C. A. (2022). Adaptation of an olfactometric system in a GC-FID in combination with GCxGC/MS to evaluate odor-active compounds of wine. *Food Chemistry*, *370*, 131004. https://doi.org/10.1016/j.foodchem.2021.131004

Zhu, F., Du, B., & Li, J. (2016). Aroma compounds in wine. *Grape and Wine Biotechnology*, 273-283. https://doi.org/10.5772/65102