



ENOLOGY ORIGINAL RESEARCH ARTICLES

Sensory characterisation using text mining analysis: An approach for a site-typicity description in Chilean Sauvignon blanc wines

Alejandro Cáceres-Mella^{1,*}, Irina Díaz-Gálvez², Claudia Viacava², and Alvaro Peña-Neira³

¹ School of Agronomy, Faculty of Food and Agronomical Sciences, Pontificia Universidad Católica de Valparaíso. Av. San Francisco s/n. Quillota, Chile

² Instituto de Investigaciones Agropecuarias, INIA, Avda. Salamanca s/n, km 105 Ruta 5 Sur, Rengo, Chile

³ Faculty of Agronomical Sciences, Universidad de Chile, Santa Rosa 11315, La Pintana, Santiago, Chile

Article number: 8494



*correspondence:
alejandro.caceres@pucv.cl

Associate editor:
Maria Tiziana Lisanti



Received:
22 January 2025

Accepted:
13 October 2025

Published:
29 October 2025

ABSTRACT

Text mining can be a valuable tool for the sensory description of wines and word clouds provide a graphical representation of sensory descriptors generated by a sensory panel. Using these tools, the sensory characteristics of commercial Sauvignon blanc wines were analysed to assess the influence of three wine producing regions, Casablanca, Leyda, and San Antonio. The wines from San Antonio and Leyda showed higher levels of acidity, both chemically and sensorially, compared to those from Casablanca, which were characterised by greater complexity. Text mining analysis allowed a detailed examination of sensory descriptors to be carried out across the study regions and revealed Casablanca wines to feature a broader diversity of descriptors than those from Leyda and San Antonio. While some descriptors overlapped across the three regions, such as *balanced* and *asparagus*, distinctive sensory profiles were evident. The wines from the three regions were described as having greener and fresher notes, with descriptors such as *vegetal*, *asparagus*, and *herbaceous*, and a mouthfeel characterised by citrus notes and acidity. Using text mining analysis, it was possible to describe the typicity of Chilean Sauvignon blanc wines and link it to their specific *terroir*.



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.

KEYWORDS: Sauvignon blanc, typicity, text mining, sensory analysis, *terroir*, word clouds

INTRODUCTION

Sauvignon blanc is the most widely planted white grapevine variety in Chile, accounting for 11.2 % of the national total; in Central Chile it is mainly produced in the renowned Casablanca and the newer Leyda and San Antonio regions. These regions, generally called “valleys”, come under Decree 464, which establishes the viticultural zoning of Chile, with each valley constituting a Denomination of Origin (D.O.) (SAG, 2023). Located 35–40 km from the Pacific Ocean, the Casablanca Valley has a Mediterranean climate, with four well-defined seasons and refreshing morning mists. Closer to the sea (10–20 km), the San Antonio Valley and Leyda Valley have similar conditions but milder winters and a stronger sea breeze. These three valleys are considered as “cold climate” regions due to the influence of the Pacific Ocean, which varies even within the valleys (Peirano-Bolelli *et al.*, 2022). While climate influences the composition of wine, there are other elements that determine its characteristics and make up what is known as *terroir*. The latter refers to the geographical origin, the grape variety, the management carried out by the producer and the techniques applied in its production, which all define the varietal typicity (Basalekou *et al.*, 2023).

The differentiation of the chemical and sensory composition of Sauvignon blanc wines according to geographical region has been studied using various sensory techniques, with an emphasis on describing sensation through sensory tools, such as descriptive analysis, sorting tasks, Pivot® profile, or typicity rating (Parr *et al.*, 2007; Parr *et al.*, 2010; Green *et al.*, 2011; Mafata *et al.*, 2019); the results are represented in graphs particularly using multivariate techniques, such as principal component analysis (PCA). Even though these techniques are important for the sensory analysis of samples, it is sometimes necessary for the evaluators to express the sensations that wines evoke in terms of aromas, flavours, and mouthfeel sensation using self-generated terminology. This often results in words and short phrases that can be too difficult to analyse to obtain satisfactory results. Sensory methods such as “Free-choice profiling” and “Flash profiling” use self-generated terminology and can be useful for the rapid sensory analysis of samples, but the terms generated can be too difficult to interpret (Lawless & Heymann, 2010). A word-based description that highlights the abundance of sensory descriptors though word clouds could be a valuable tool for classifying the oenological potential of a wine from a particular region. Word clouds can be generated using tools such as data mining, whereby knowledge and patterns is extracted from unstructured text data, such as a free-form description of a wine. The use of text mining and word clouds has not yet been widely adopted in the viticultural field and literature referring to the use of these tools is limited (Pelonnier-Magimel *et al.*, 2020; Alderson *et al.*, 2021; Días Araujo *et al.*, 2021; Gupta & Katarya, 2024). However, these methodologies have been used in wine market studies to understand consumer decision-making and to determine consumer satisfaction (Fu *et al.*, 2022), as well as to determine key factors for defining wine tours

(Barbierato *et al.*, 2021) and to determine the usefulness of online information for wine consumers when choosing a product (Calderon-Monge *et al.*, 2024). To our knowledge, no studies have been reported on the use of text mining and word clouds for classifying Sauvignon blanc wines from different geographical regions in Chile. Therefore, the aim of this study was to provide a sensory description of the sensory typicity of Sauvignon blanc commercial wines from three geographical regions of Chile using the novel approach of text mining analysis and graphical representation through word clouds.

MATERIALS AND METHODS

1. Wine samples

The study was conducted on 20 commercial Sauvignon blanc wines, all produced from the 2020 vintage and originating from three geographical coastal regions in Central Chile. Seven of these wines were from the Denomination of Origin (D.O.) Casablanca, seven from the D.O. Leyda and six from the D.O. San Antonio. The collection for the samples was carried out by purchasing these wines from specialised stores and directly from wineries. Table 1 presents the characteristics of the wines selected for the study.

None of the wines underwent barrel ageing or malolactic fermentation. Three bottles of each type of wine were used, corresponding to three replicates per wine. The climatic data for the three regions under study during the 2020 season can be seen in Table 2.

2. Descriptive sensory analysis procedure

The sensory study was conducted over two sessions at the facilities of the “Incubadora de Innovación para el Vino y la Oliva” (IIVO), Chile. A sensory analysis of the samples was carried out in sensory booths under controlled conditions. The sensory panel consisted of 11 judges (five men and six women) of between 23 and 55 years old, who were trained in the sensory analysis of white and red wines. Prior to the evaluation, a session was dedicated to training the judges in the recognition of the characteristic aromas of Sauvignon blanc wines. For this purpose, different types of fruit, vegetables and herbs were used, as shown in Table 3, and each descriptor was placed in transparent glasses (ISO 3591.1977) so that the judges could smell and become familiar with them. This session lasted one hour and took place the day before the wine tasting. As well as using the aforementioned elements to train the judges to recognise the most typical descriptors of Sauvignon blanc wines, the previous experience of the judges in the sensory evaluation of wines was drawn on to reach a consensus on the sensations perceived in the wines; for example, for the floral and passion fruit descriptors, the sensory panel determined that the sensation would be greater, if the aroma contained notes of orange blossom and jasmine (floral) or passion fruit, which are all typical of this variety of wine. Regarding acidity, the sensory panel concluded that a wine would be perceived as more acidic if it produced a greater sensation of salivation

TABLE 1. Characteristics of the Sauvignon blanc wines employed in the study and their geographical regions.

| Wine code | Geographical region | Vintage | Retail price (US\$) | Type of production | Bottle closure |
|-----------|---------------------|---------|---------------------|-----------------------------|----------------|
| CB1 | Casablanca | 2020 | 8.3 | Organic | Screwcap |
| CB2 | Casablanca | 2020 | 7.8 | Certified sustainable | Screwcap |
| CB3 | Casablanca | 2020 | 11.5 | Certified sustainable | Screwcap |
| CB4 | Casablanca | 2020 | 5.2 | Certified sustainable | Screwcap |
| CB5 | Casablanca | 2020 | 20.9 | Organic/Biodynamic | Screwcap |
| CB6 | Casablanca | 2020 | 6.3 | Conventional | Screwcap |
| CB7 | Casablanca | 2020 | 20.9 | Certified sustainable | Screwcap |
| SA1 | San Antonio | 2020 | 7.3 | Certified sustainable | Screwcap |
| SA2 | San Antonio | 2020 | 7.3 | Certified sustainable | Screwcap |
| SA3 | San Antonio | 2020 | 23.4 | Certified sustainable | Cork |
| SA4 | San Antonio | 2020 | 6.3 | Organic (Ecocert certified) | Screwcap |
| SA5 | San Antonio | 2020 | 5.2 | Certified sustainable | Screwcap |
| SA6 | San Antonio | 2020 | 11.5 | Organic/Biodynamic | Screwcap |
| LE1 | Leyda | 2020 | 14.6 | Certified sustainable | Cork |
| LE2 | Leyda | 2020 | 9.4 | Certified sustainable | Screwcap |
| LE3 | Leyda | 2020 | 6.3 | Certified sustainable | Screwcap |
| LE4 | Leyda | 2020 | 20.8 | Certified sustainable | Screwcap |
| LE5 | Leyda | 2020 | 16.7 | Certified sustainable | Screwcap |
| LE6 | Leyda | 2020 | 9.4 | Certified sustainable | Screwcap |
| LE7 | Leyda | 2020 | 7.8 | Certified sustainable | Screwcap |

TABLE 2. Climatic data from January to March in the areas of Casablanca, Leyda and San Antonio.

| | D.O. Casablanca | D.O. Leyda | D.O. San Antonio |
|-------------------------------------|-----------------|------------|------------------|
| Maximum temperature (°C) | 27.7 | 21.3 | 21.7 |
| Minimum temperature (°C) | 8.9 | 12.1 | 9.5 |
| Thermal oscillation (°C) | 18.8 | 9.1 | 12.2 |
| Relative humidity (%) | 71.3 | 84.7 | 80.7 |
| Precipitation (mm) | 0.0 | 0.0 | 0.0 |
| Solar radiation (W/m ²) | 856.3 | 924.3 | 643.2 |
| Maximum wind speed (m/s) | 2.3 | 3.9 | 3.7 |

Data obtained from the Agroclimatic systems AGROMET (Red Agroclimática Nacional) of Chile.

and discomfort on the taste buds, and less acidic if this response was not elicited. In the case of colour, the consensus was that a wine with less colour would be considered a bright wine with a pale-yellow colour and greenish notes, whereas a wine with golden yellow notes would be considered to have more colour. Finally, given that young wines were used, the sensory panel agreed that a wine with fewer descriptors would be considered less complex, whereas a wine with a greater number of sensory descriptors would be considered more complex. All the judges were informed about the details of the

evaluation through a leaflet containing information regarding their participation and rights. All members of the trained panel signed an informed consent. The study was conducted in accordance with the tenets of the Declaration of Helsinki (World Medical Association, 2013) and was approved by the Pontificia Universidad Católica de Valparaíso Bioethics and Biosecurity Committee (BIOPUCV-B 443-2021). The judges were instructed that the aim of the evaluation was to analyse samples of Sauvignon blanc wines. Thirty millilitres of wine was served in standard tasting glasses (ISO 3591.1977)

TABLE 3. Aromatic descriptors of the Sauvignon blanc wines and associated reference material used as an example odour.

| Groups | Aromas | Descriptors | Reference material |
|----------|------------------|-----------------|-------------------------|
| Vegetal | Fresh vegetables | Leaves/Stems | Green leaves |
| | | Grass | Freshly-cut grass |
| | | Green capsicum | Green capsicum slice |
| | | Green pepper | Green pepper slice |
| | | Tomato stalk | Tomato stalks |
| | | Boxwood/cat pee | Boxwood leaves |
| | Canned/cooked | Green beans | Fresh green beans |
| | | Asparagus | Canned asparagus |
| | | Green olives | Canned green olives |
| | | Artichoke | Canned artichoke |
| Citric | Citrus fruits | Grapefruit | Grapefruit slices |
| | | Lemon | Lemon slices |
| | | Orange | Orange slices |
| | | Lemon peel | Lemon peel |
| Tropical | Tropical fruits | Pineapple | Pineapple slice |
| | | Melon | Melon piece |
| | | Banana | Banana piece |
| | | Cherimoya | Cherimoya piece |
| Others | Stone fruits | Apricot | Fresh apricot slice |
| | | Peach | Fresh peach slice |
| | Pome fruits | Green apple | Fresh green apple slice |

labelled with three-digit codes in a completely randomised order, and water and unsalted crackers were provided between samples for palate cleansing. The wines were served at a temperature of 10°–12 °C. The judges followed guidelines for the sensory evaluation, which was in two parts. The first part of the evaluation consisted of a structured sensory analysis, in which the judges evaluated the wines through a descriptive sensory analysis using nine attributes: one visual attribute (colour), six olfactory attributes (herbal aromas, green pepper, passion fruit, tropical fruits, citrus fruits, and floral aromas) and two mouthfeel attributes (acidity and complexity in the mouth), all derived from available literature on the sensory outcomes of this variety (Tsai *et al.*, 2022; Parr *et al.*, 2007; Lund *et al.*, 2009; Green *et al.*, 2011). These descriptors were provided in a sensory evaluation form using a discrete scale of 1 to 10, where 1 indicates low intensity of the descriptor and 10 indicates high intensity (Lawless & Heymann, 2010). The second part of the evaluation consisted in an unstructured sensory description of the wines. The judges were asked to freely note any olfactory and gustatory descriptors they recognised in the wines using the descriptors from the previous training session (Table 3) and sensory language based on their individual perceptions

of the wine. In this part of the evaluation, the judges were asked to preferably write down a single word that described the aromas and flavours they perceived in the samples; more than one word or short phrases (no more than three to four words) were also acceptable, but long phrases were to be avoided.

3. Text mining and generation of word clouds using R software

The information collected from the unstructured description of the wines was transcribed into an Excel table to organise the words and/or short phrases. Once organised, they were translated into English, taking special care to preserve the meaning of each phrase and/or word. In the second phase of the analysis, a reprocessing and cleaning of the words were conducted for the short phrases, where redundant words, connectors between words (articles, prepositions, conjunctions), and similar words related to the perceived sensation (*e.g.*, acid, acidity) were removed or unified. For instance, after processing the sentence “tasty and fresh in mouth with a citric profile and notes of flowers and white asparagus”, which had been written by one of the judges and translated into English, it read as “tasty,

fresh, citric, floral, asparagus”. Once the text for each group of wines was finalised, it was analysed through text mining with the RStudio software (version 2023.12.1, Posit Software, Boston, MA, USA) using the following libraries: “tm” (Feinerer *et al.*, 2008), “snowballc” (Bouchet-Valat, 2023), “wordcloud” (Fellows, 2018), “RColorBrewer” (Neuwirth, 2014), “ggplot2” (Wickham, 2016), and “wordcloud2” (Lang, 2018). The R analysis processed the text by removing punctuation, stopwords, and whitespace, and converting the text to lowercase. Subsequently, the software enabled frequency analysis of the terms and construction of word clouds for the visual description of the analysed wines.

4. Wine chemical analyses

The analyses of volatile acidity (g acetic acid/L), titratable acidity (g tartaric acid/L), residual sugar (g glucose/L), alcohol content (% v/v), and pH were carried out using the analytical methods recommended by the International Organisation of Vine and Wine (OIV) (OIV, 2012).

To determine total phenols, absorbance at 280 nm was measured using UV spectrophotometry, with gallic acid as the standard, according to the methodology of Glories (1984) on a UV-1280 UV-Vis spectrophotometer (Shimadzu, Kyoto, Japan). These analyses are presented in Table 4.

5. Statistical analyses

For the chemical and sensory descriptive analysis, the data were subjected to the Shapiro–Wilk test to assess their normality and the Bartlett test for the homogeneity of variances. These were followed by one way analysis of variance (ANOVA) and LSD test with a significance level of 95 % ($p < 0.05$). For the analysis of the unstructured evaluation, especially the word cloud analysis, a frequency table was constructed, and the statistical tests Chi-square and adjusted residuals analysis were applied. All the statistical analyses of the results were conducted using the R statistical software (4.2.2 version) and RStudio statistical software (2023.12.1 version, Posit Software, Boston, MA, USA).

TABLE 4. Analytical parameters for Sauvignon blanc wines from three geographical regions.

| Wine codes | TP (mg GAE/L) | VA (g/L) | TA (g/L) | RS (g/L) | Ethanol (% vol) | pH |
|------------|------------------|---------------|---------------|---------------|--------------------|---------------|
| CB1 | 128.91 ± 0.64 | 0.22 ± 0.01 | 6.10 ± 0.09 | 0.35 ± 0.05 | 13.87 ± 0.29 | 2.76 ± 0.03 |
| CB2 | 118.55 ± 0.31 | 0.25 ± 0.01 | 5.90 ± 0.17 | 0.95 ± 0.05 | 13.70 ± 0.10 | 2.89 ± 0.03 |
| CB3 | 118.75 ± 0.89 | 0.37 ± 0.03 | 5.60 ± 0.09 | 1.00 ± 0.05 | 13.90 ± 0.26 | 2.86 ± 0.02 |
| CB4 | 129.01 ± 1.63 | 0.21 ± 0.02 | 5.85 ± 0.00 | 0.35 ± 0.05 | 12.70 ± 0.10 | 2.78 ± 0.01 |
| CB5 | 129.42 ± 0.99 | 0.32 ± 0.01 | 6.00 ± 0.00 | 0.45 ± 0.05 | 14.00 ± 0.17 | 2.64 ± 0.01 |
| CB6 | 118.96 ± 1.52 | 0.41 ± 0.03 | 5.40 ± 0.00 | 1.10 ± 0.00 | 14.00 ± 0.17 | 2.85 ± 0.03 |
| CB7 | 137.53 ± 2.47 | 0.41 ± 0.02 | 6.15 ± 0.00 | 1.00 ± 0.00 | 13.97 ± 0.12 | 2.92 ± 0.03 |
| | 125.78 ± 2.76 ab | 0.31 ± 0.03 a | 5.86 ± 0.10 b | 0.74 ± 0.13 a | 13.73 ± 0.18 a | 2.81 ± 0.04 a |
| SA1 | 112.08 ± 1.11 | 0.33 ± 0.02 | 6.65 ± 0.09 | 0.35 ± 0.05 | 13.33 ± 0.29 | 2.82 ± 0.01 |
| SA2 | 176.41 ± 2.13 | 0.59 ± 0.01 | 7.30 ± 0.09 | 0.70 ± 0.00 | 13.70 ± 0.46 | 2.75 ± 0.02 |
| SA3 | 153.12 ± 4.19 | 0.31 ± 0.01 | 6.75 ± 0.15 | 0.95 ± 0.05 | 14.03 ± 0.25 | 2.95 ± 0.01 |
| SA4 | 129.11 ± 2.78 | 0.25 ± 0.03 | 6.00 ± 0.00 | 0.60 ± 0.00 | 13.73 ± 0.42 | 2.89 ± 0.06 |
| SA5 | 129.03 ± 17.83 | 0.25 ± 0.01 | 7.14 ± 0.20 | 0.93 ± 0.04 | 13.97 ± 0.06 | 2.78 ± 0.04 |
| SA6 | 127.78 ± 1.54 | 0.21 ± 0.03 | 6.75 ± 0.00 | 0.50 ± 0.00 | 14.17 ± 0.25 | 2.86 ± 0.02 |
| | 137.92 ± 9.38 a | 0.32 ± 0.06 a | 6.76 ± 0.18 a | 0.67 ± 0.10 a | 13.82 ± 0.12 a | 2.84 ± 0.03 a |
| LE1 | 103.46 ± 10.67 | 0.30 ± 0.00 | 6.75 ± 0.00 | 0.55 ± 0.05 | 13.77 ± 0.21 | 2.88 ± 0.02 |
| LE2 | 121.21 ± 1.75 | 0.37 ± 0.03 | 6.85 ± 0.09 | 0.30 ± 0.00 | 12.73 ± 0.06 | 3.03 ± 0.03 |
| LE3 | 150.05 ± 4.09 | 0.40 ± 0.01 | 5.85 ± 0.00 | 4.35 ± 0.05 | 13.63 ± 0.15 | 2.92 ± 0.02 |
| LE4 | 131.47 ± 0.81 | 0.39 ± 0.01 | 6.60 ± 0.00 | 1.50 ± 0.00 | 13.37 ± 0.15 | 2.90 ± 0.02 |
| LE5 | 89.41 ± 3.71 | 0.38 ± 0.01 | 7.05 ± 0.15 | 0.50 ± 0.10 | 13.90 ± 0.20 | 2.80 ± 0.03 |
| LE6 | 113.62 ± 0.62 | 0.45 ± 0.01 | 6.90 ± 0.00 | 0.40 ± 0.00 | 13.50 ± 0.40 | 2.90 ± 0.01 |
| LE7 | 106.85 ± 1.11 | 0.38 ± 0.00 | 6.70 ± 0.09 | 1.03 ± 0.06 | 13.27 ± 0.21 | 2.95 ± 0.01 |
| | 116.58 ± 7.52 b | 0.38 ± 0.02 a | 6.67 ± 0.15 a | 1.23 ± 0.54 a | 13.45 ± 0.15 a | 2.91 ± 0.03 a |

Values are expressed as mean ± standard deviation ($n = 3$). TP = total phenols; VA = volatile acidity; TA = titratable acidity; RS = reducing sugars; GAE = gallic acid equivalent. Different letters denote statistical differences according to the LSD test ($p < 0.05$).

RESULTS

The results of the chemical analysis of the wines are shown in Table 4. Differences in total phenols (TP) were found between the regions, the wines of San Antonio showing the highest TP concentration. The other parameter that showed differences was titratable acidity (TA), with the wines from Leyda and San Antonio exhibiting higher acidity.

The results of the descriptive analysis of the wines can be seen in Figure 1. Among all the parameters evaluated, differences were only observed in two parameters, acidity and complexity: the wines from Leyda were perceived as having the highest acidity and those from Casablanca as having the highest complexity.

Figure 2 shows a description of the wines from the three different geographical areas in the form of a word cloud. Regarding the Casablanca wines (Figure 2A), the word cloud was constructed using 26 different terms, those appearing more than once being *persistent* (5), *fresh* (4), *herbaceous* (4), *tasty* (4), *citric* (3), *balanced* (3), *mature* (2), *asparagus* (2), *graphite* (2), *bitter* (2), *green pepper* (2), and *green* (2), totalling 49 words. Figure 2B shows the descriptors for the Leyda wines. This word cloud was constructed using 23 sensory terms, and the most frequent descriptors were *acidic* (6), *citric* (4), *light* (4), *tropical* (3), *bitter* (2), *fatty* (2), *juicy* (2), *complex* (2), *balanced* (2), *tasty* (2), *good* (2), and *pineapple* (2), totalling 44 words. Finally, Figure 2C shows the descriptors for the San Antonio

wines. The word cloud was constructed using 24 different terms, and the most frequent descriptors were *asparagus* (8), *acidic* (6), *balanced* (3), *saline* (3), *sweet* (2), *olives* (2), *herbaceous* (2), *aromatic* (2), *mineral* (2), and *persistent* (2), totalling 46 words.

The statistical analysis revealed that some words are more frequent than expected within a uniform distribution. Figure 3 shows that, for the Casablanca region, the most frequent descriptors are *persistent*, *fresh*, *herbaceous*, *tasty*, *citric*, and *balanced*. However, only the descriptor *persistent* is statistically different (Table S1). In the Leyda region, the most frequent words were *acidic*, *citric*, *light*, and *tropical*, but only *acidic* was statistically different (Table S2). In the case of San Antonio, the most frequent descriptors were *asparagus*, *acidic*, *saline*, and *balanced*, with *asparagus* and *acidic* being statistically different (Table S3). In general, clear differences can be seen between the descriptors perceived in the wines by the judges, and the statistical analysis suggests that the statistically significant words could be key words for the sensory description of Sauvignon blanc wines from each of the respective regions.

DISCUSSION

Sensory evaluation is a tool for understanding and quantifying the various sensations that wines can produce, which are linked to their organoleptic characteristics. A group of expert evaluators rated aspects related to the visual characteristics

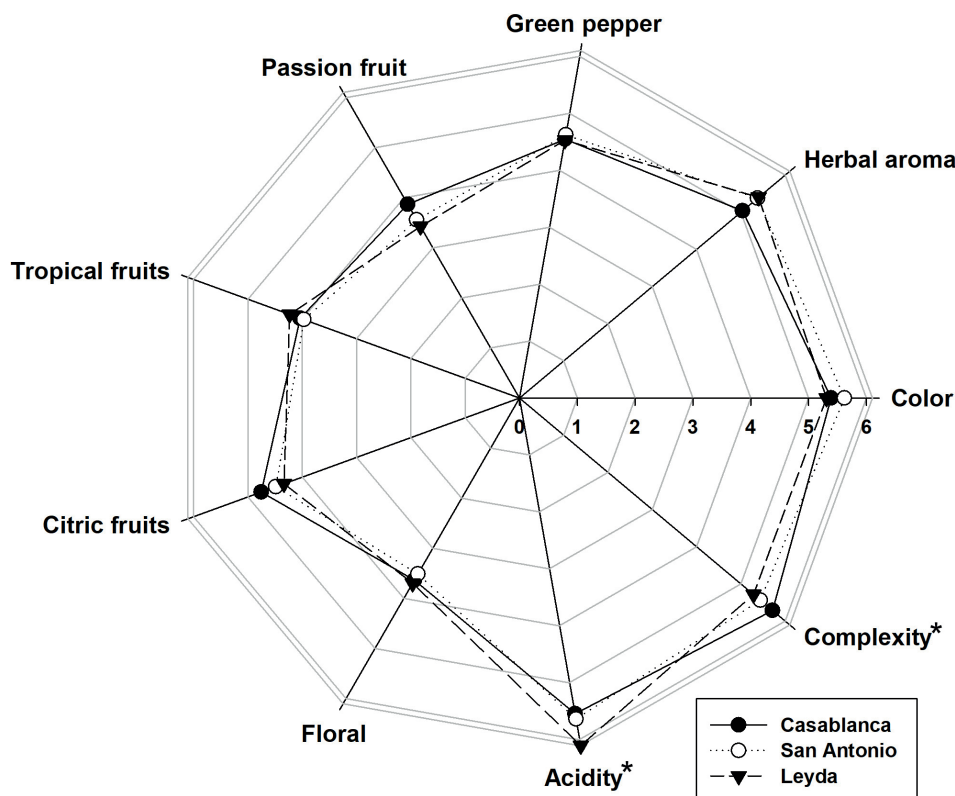


FIGURE 1. Spider plot of the sensory parameters evaluated by participants in Sauvignon blanc wines from three geographical regions. * Denotes statistically significant differences according to the LSD test ($p < 0.05$).

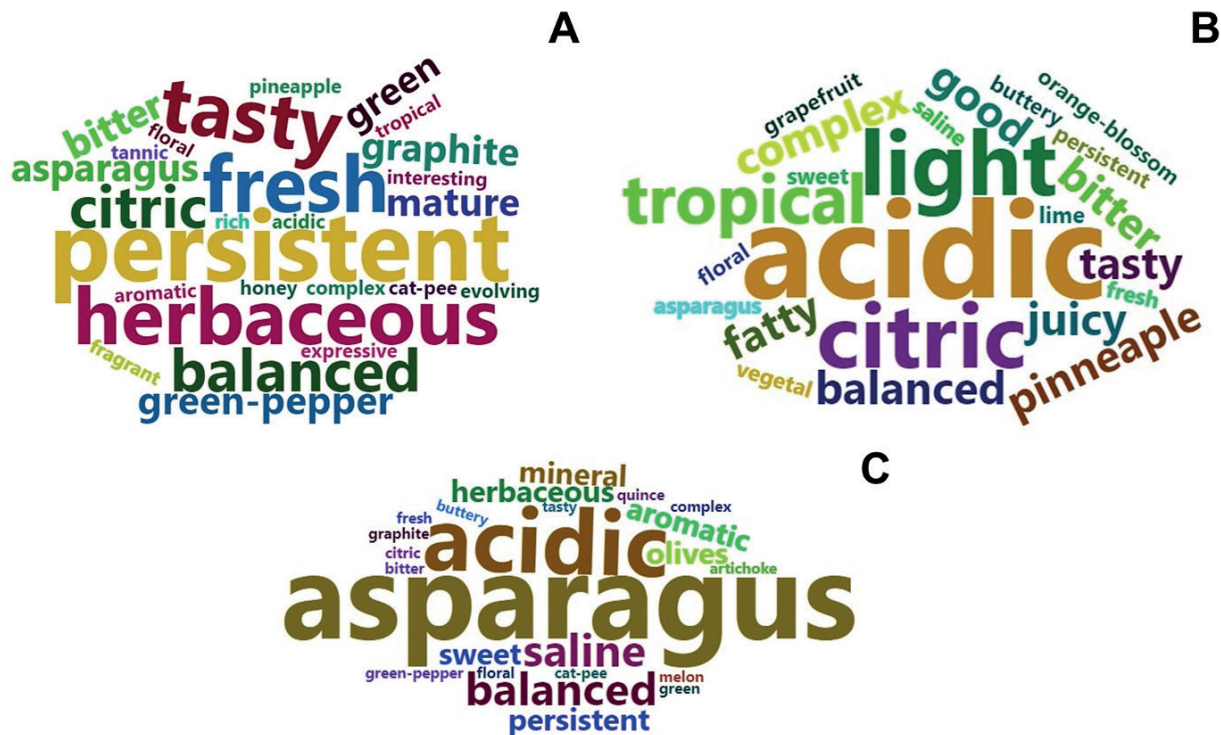


FIGURE 2. Word clouds constructed by participants in the sensory evaluation of Sauvignon blanc wines from three geographical regions in the central zone of Chile. A) Casablanca Valley, B) Leyda Valley, and C) San Antonio Valley. The size of the text represents the frequency that the respective descriptor was used by the judge: the larger the text the higher the frequency. Colour has no specific meaning in this figure.

of the wine, in addition to quantifying the sensations arising from the perceived aromas and flavours, as well as other sensations related to body and persistence, among others, which together can describe a type of wine (Lawless & Heymann, 2010).

The results showed that the participants were able to discriminate between the wines in terms of certain sensations perceived in the descriptive analysis, such as acidity and complexity. Acidity was perceived the more in the wines from Leyda and San Antonio, which is consistent with the higher titratable acidity present in these wines (Table 4). These differences may be due to the climatic differences in these geographical areas (Table 1), as San Antonio and Leyda are closer to the Pacific Ocean, therefore experiencing a much greater coastal influence than Casablanca. The lower temperatures during the grape ripening stage in these regions reduces respiration in the berries, which in turn preserves their organic acids (Rojas *et al.*, 2024). Acidity is a desirable sensation for white wines and is thus an important attribute in wines from Leyda and San Antonio (Volschenk *et al.*, 2006). As for complexity, it is a sensation that consumers use to describe the quality of a wine, which encompasses various components. For example, wines may exhibit multiple aromas or even have a single distinctive element, such as flavour intensity (Schlich *et al.*, 2015). In this study, the wines from Casablanca exhibited greater complexity, followed by the wines from San Antonio and Leyda in that order. Although complexity can be due to the different types of aromas in a

wine or to aromas that appear at different moments during tasting, in the present study, a lack of differentiation in other sensory parameters, such as the aroma descriptors in the wines, can make it difficult to pinpoint the reason for these differences.

In the investigation by Rojas *et al.* (2024), in which the wines were chemically analysed for their concentration of phenols and volatile compounds, the wines from Casablanca showed a higher abundance of esters, C₁₃ norisoprenoids and terpenes, which are associated with citric and tropical notes, whereas Leyda exhibited more floral aromas, and San Antonio greener aromas like vegetables and freshly cut grass. This could be an explanation for the greater complexity associated with the wines from Casablanca. Despite this, the statistical techniques that enabled better visualisation of the perceived sensations using text mining revealed the wines to be represented by quite distinct descriptors across the three analysed regions (Figure 2).

The wines from Casablanca were described by the panellists using a total of 49 words comprising 26 different terms, the most frequent terms (≥ 3 times) being *persistent*, *fresh*, *herbaceous*, *tasty*, *citric*, and *balanced*. The wines from Leyda were described using 23 terms out of a total of 44 words, those most frequently mentioned (≥ 3 times) being *acidic*, *citric*, *light*, and *tropical*. Meanwhile, the wines from San Antonio were described using 24 sensory descriptors, accounting for a total of 46 words, the most frequent (≥ 3 times) being *asparagus*, *acidic*, *saline*, and *balanced*.

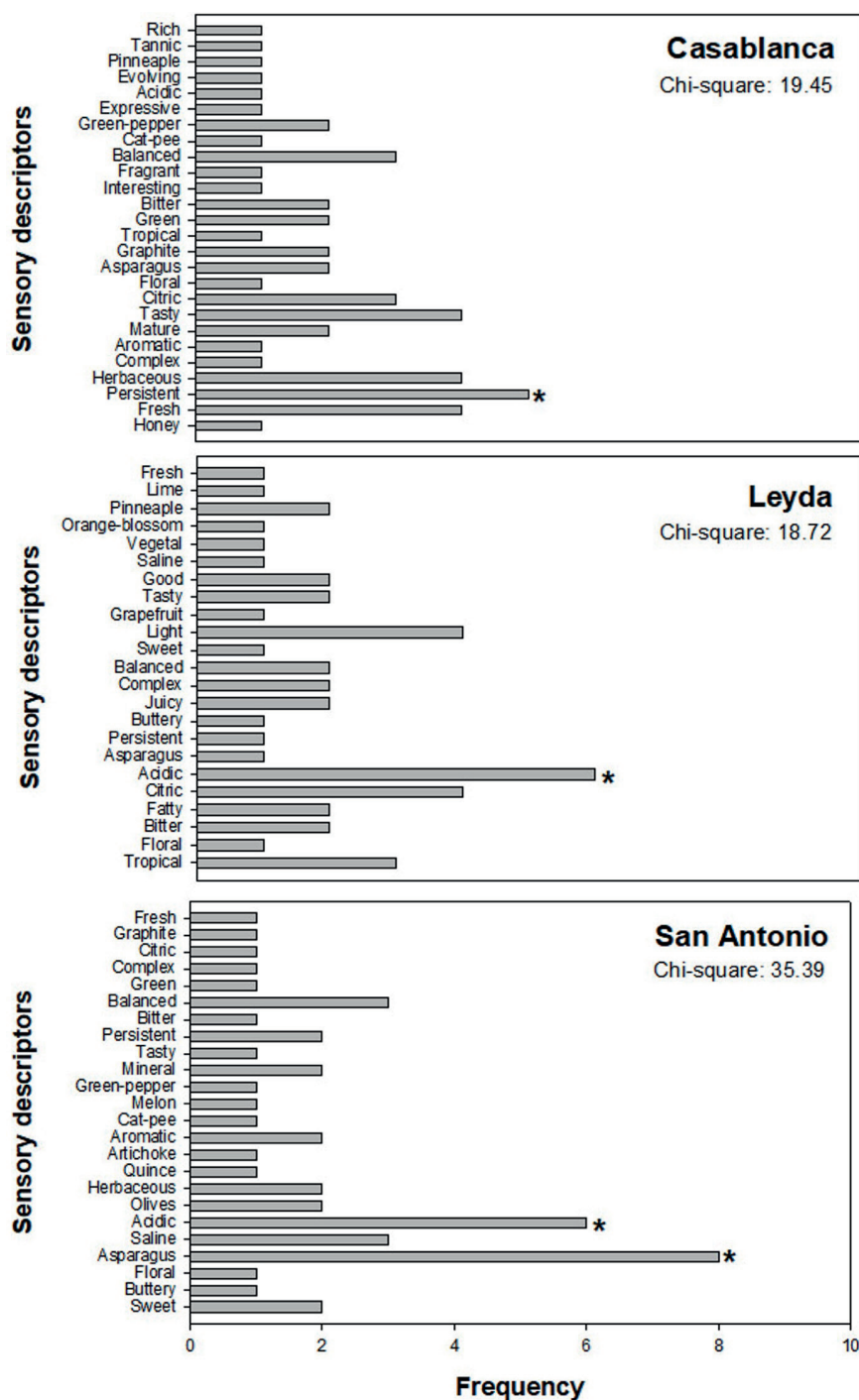


FIGURE 3. Frequency graphs for the words presented in the word clouds. * Denote significant difference according to the Chi-square test ($p < 0.05$).

Regarding complexity, Casablanca was associated with more descriptors than the wines from the other two geographical areas, which may corroborate the finding that the wines from Casablanca were the most complex.

In general, Sauvignon blanc is a variety characterised by citrus and tropical aromas (Lund *et al.*, 2009). It has been observed that aromas mainly depend on geographical area; for example, the results of Green *et al.*, (2011) indicate that wines from cooler geographical areas have more vegetal aromas, whereas warmer areas yield tropical fruit aromas.

The wines from San Antonio, and especially Leyda, exhibited descriptors associated with fresher and more vegetal aromas, which can be explained by their proximity to the Pacific Ocean and the lower daytime temperature in these regions (Table 2). Although the descriptive analysis only revealed differences in terms of wine complexity and acidity, the word clouds generated using the R software allowed the sensations produced by the wines to be represented and thus differentiated. The word clouds show clear differences between the wines in the terms used, although there are terms common to all the wines due to their

being made from the same grape variety and retaining certain typicality across the regions. For example, it can be seen that the wines from Leyda and San Antonio were perceived as having higher acidity, both sensorially and chemically: both wines were described using words like *acidic* and *citric* due to their proximity to the Pacific Ocean and *saline* due to the greater coastal influence than in the Casablanca region. It is striking that the term *saline* was used as a descriptor for these wines. The perception of saltiness by a trained panel is related to the concentration of salts in the wine (Walker *et al.*, 2023), but this would need to be confirmed by further study. All three regions feature the descriptor *balanced*, which is a positive attribute in wines. The descriptor *asparagus* also appears in all three regions, although San Antonio shows a higher frequency of this descriptor (Figure 3). In previous research on the sensory typicality of Sauvignon blanc wines these wines have been described using descriptors such as *fruity* and *vegetal* due to the presence of volatile thiols and methoxypyrazines, and descriptors like *boxwood*, *passion fruit*, *green pepper*, *asparagus*, and *vegetal* contributed to the typicality of New Zealand Sauvignon blanc wines (Parr *et al.*, 2013). Other research has differentiated between wines from the northern and southern Hemispheres, with more fruity notes in wines from the northern Hemisphere (passion fruit, grapefruit citrus notes, and tropical notes such as pineapple and guava), and greener notes in those from the southern Hemisphere (green pepper, grass, vegetal) (Mateo-Vivaracho *et al.*, 2010; Green *et al.*, 2011; Parr *et al.*, 2013). The findings of the present study validate the typicality of Chilean Sauvignon blanc wines, given that they exhibit a greater presence of green notes, like those observed in New Zealand Sauvignon blanc wines. This similarity can be attributed to the influence of the Humboldt Current, an oceanic current from Antarctica that cools the coastal vineyards of Central Chile and is located in the southern Hemisphere, corroborating the findings of other authors.

Even though we worked with commercial wines in this study, which can be affected by different winemaking techniques, clones, and other factors, it is important to note that wines of this variety are recognised for their pale-yellow colour, high acidity and distinct aromatic typicality (Peirano-Bolelli *et al.*, 2022). To achieve these characteristics, it is necessary for wineries to use similar winemaking techniques, prioritising very ripe berries containing a high concentration of organic acids and the use of antioxidants to prevent wine browning, as well as minimising the contact between the solid parts of the grape and the must to avoid high extraction of phenolic compounds and thus preserve the pale yellow colour. These attributes therefore also appear in our results, especially the greener notes, which could indicate that the wines from the three regions have in common a certain sensory typicality, with some differences that can be explained by the inherent variability of the three different *terroirs*.

CONCLUSION

Text mining and word clouds were used to describe the sensory characteristics of Sauvignon blanc wines from three geographical regions of Chile, revealing that Casablanca wines were the most complex and diverse. Additionally, we observed that the wines from Casablanca, Leyda and San Antonio shared some descriptors, but they also showed regional differences and demonstrated that the typicality of Chilean Sauvignon blanc wines is characterised by the greener notes in the wines, such as *asparagus*, *vegetal*, and *green pepper*, among others. This methodology allowed us to determine the typicality of the wines based on their respective organoleptic attributes via the words the evaluators used to describe the wine and the size of the word in the word cloud, which indicates its importance for that wine. The relative frequency allows us to better interpret the size of the word and therefore its relevance to the description of a wine. Word cloud analysis revealed a wider and more diverse set of sensory descriptors characterising Sauvignon blanc wines. These findings support the development of an expanded descriptor lexicon that can be applied in future sensory evaluation to more accurately capture and describe the typicality of Chilean Sauvignon blanc wines.

Future studies that explore subzone variations, key chemical compounds, and seasonal impacts could further refine the understanding of typicality and contribute to producing a more detailed classification of these wines.

ACKNOWLEDGEMENTS

The authors would like to thank oenologist Diego Rivera from Vina Garcés Silva and the wineries in the Leyda and San Antonio areas for providing the wines for this study. This research was funded by “Agencia Nacional de Investigación y Desarrollo” (ANID Chile), Fondecyt Iniciación Fund, grant number 11180265.

REFERENCES

- Alderson, H., Liu, C., Mehta, A., Suresh Gala, H., Rutendo, N., Chen, Y., Zhang, Y., Wang, S., & Serventi, L. (2021). Sensory profile of kombucha brewed with New Zealand ingredients by focus group and word clouds. *Fermentation*, 7(3), 100. <https://doi.org/10.3390/fermentation7030100>
- Barbierato, E., Bernetti, I., & Capecci, I. (2021). Analyzing TripAdvisor reviews of wine tours: an approach based on text mining and sentiment analysis. *International Journal of Wine Business Research*, 34(2), 212-236. <https://doi.org/10.1108/IJWBR-04-2021-0025>
- Basalekou, M., Tataridis, P., Georgakis, K., & Tsintonis, C. (2023). Measuring wine quality and typicality. *Beverages*, 9(2), 41. <https://doi.org/10.3390/beverages9020041>
- Bouchet-Valant, M. (2023). SnowballC: Word stemming based on the snowball C++library. R package version 0.7.1.
- Calderon-Monge, E., Ripollés-Matallana, V., Baruque-Zanón, B., & Porrás Alfonso, S. (2024). Big data analysis of Spanish wine consumers reviews. *International Journal of Wine Business Research*. <https://doi.org/10.1108/IJWBR-12-2023-0084>

- Días Araujo, L., Parr, W., Grose, C., Hedderley, D., Masters, O., Kilmartin, P., & Valentin, D. (2021). In-mouth attributes driving perceived quality of Pinot noir wines: Sensory and chemical characterisation. *Food Research International*, 149, 110665. <https://doi.org/10.1016/j.foodres.2021.110665>
- Feinerer, I., Hornik, K., & Meyer, D. (2008). Text mining infrastructure in R. *Journal of Statistical Software*, 25, 1-54. <https://doi.org/10.18637/jss.v025.i05>
- Fellows, I. (2018). Wordcloud: Word clouds. R package version 2.6.
- Fu, W., Choi, E. K., & Kim, H. S. (2022). Text mining with network analysis of online reviews and consumers' satisfaction: A case study in Busan wine bars. *Information*, 13(3), 127. <https://doi.org/10.3390/info13030127>
- Glories, Y. (1984). La couleur des vins rouges, 2eme Partier. Mesure, origine et interpretation. *Connaissance de la Vigne et du Vin*, 18, 253–271. <https://doi.org/10.20870/oenone.1984.18.4.1744>
- Green, J., Parr, W., Breitmeyer, J., Valentin, D., & Sherlock, R. (2011). Sensory and chemical characterisation of Sauvignon blanc wine: Influence of source of origin. *Food Research International*, 44, 2788–2797. <https://doi.org/10.1016/j.foodres.2011.06.005>
- Gupta, G., & Katarya, R. (2024). A computational approach towards food-wine recommendations. *Expert Systems with Application*, 238 Part A, 121766. <https://doi.org/10.1016/j.eswa.2023.121766>
- Lang, Y. (2018). Wordcloud2: Crate word cloud by “HTMLWidget”. R package version 0.2.1.
- Lawless, H. & Heymann, H. (2010). *Sensory Evaluation of Food*. New York, USA: Springer. <https://doi.org/10.1007/978-1-4419-6488-5>
- Lund, C., Thompson, M., Benkwitz, F., Wohler, M., Triggs, C., Gardner, R., Heymann, H., & Nicolau, L. (2009). New Zealand Sauvignon Blanc distinct flavour characteristics: Sensory, chemical, and consumer aspects. *American Journal of Enology and Viticulture*, 60, 1–12. <https://doi.org/10.5344/ajev.2009.60.1.1>
- Mafata, M., Brand, J., Panzeri, V., Kidd, M., & Buica, A. (2019). A multivariate approach to evaluating the chemical and sensorial evolution of South African Sauvignon blanc and Chenin blanc wines under different bottle storage conditions. *Food Research International*, 125, 108515. <https://doi.org/10.1016/j.foodres.2019.108515>
- Mateo-Vivaracho, L., Zapata, J., Cacho, J., & Ferreira, V. (2010). Analysis, occurrence and potential sensory significance of five polyfunctional mercaptans in white wines. *Journal of Agricultural and Food Chemistry*, 58(18), 10184–10194. <https://doi.org/10.1021/jf101095a>
- Neuwirth, E. (2014). RColor Brewer: Color brewer palettes. R package version 1.1-2.
- OIV. (2012). Compendium of International Methods of Wine and Must Analysis; Volume III, OIV: Paris, France.
- Parr, W. V., Green, J. A., White, K. G., & Sherlock, R. R. (2007). The distinctive flavour of New Zealand Sauvignon blanc: Sensory characterisation by wine professionals. *Food Quality and Preference*, 18(6), 849–861. <https://doi.org/10.1016/j.foodqual.2007.02.001>
- Parr, W. V., Valentin, D., Green, J. A., & Dacremont, C. (2010). Evaluation of French and New Zealand Sauvignon wines by experienced French wine assessors. *Food Quality and Preference*, 21(1), 56–64. <https://doi.org/10.1016/j.foodqual.2009.08.002>
- Parr, W., Schlich, P., Theobald, J.C., & Harsch, M.J. (2013). Association of selected vitivicultural factors with sensory and chemical characteristics of New Zealand Sauvignon blanc wines. *Food Research International*, 53, 464–475. <https://doi.org/10.1016/j.foodres.2013.05.028>
- Peirano-Bolelli, P., Heller-Fuenzalida, F., Cuneo, I.F., Pena-Neira, A., & Cáceres-Mella, A. (2022). Changes in the composition of flavonols and organic acids during ripening for three cv. Sauvignon Blanc clones grown in a cool-climate valley. *Agronomy*, 12(6), 1357. <https://doi.org/10.3390/agronomy12061357>
- Pelonnier-Magimel, E., Mangiorou, P., Darriet, P., de Revel, G., Joudes, M., Marchal, A., Marchand, S., Pons, A., Riquier, L., Teissedre, P-L., Thibon, C., Lytra, G., Tempère, S., & Barbe, J-C. (2020). Sensory characterisation of Bordeaux red wines produced without added sulfites. *Oeno One*, 54(4), 687-697. <https://doi.org/10.20870/oenone.2020.54.4.3794>
- Rojas, J., Viacava, C., Ubeda, C., Peña-Neira, A., Cuneo, I., Kuhn, N., Cáceres-Mella, A. (2024). Chemical characterization of Sauvignon blanc wines from three cold-climate-growing areas of Chile. *Foods*, 13(13), 1991. <https://doi.org/10.3390/foods13131991>
- SAG. (2023). Servicio Agrícola y Ganadero. Catastro Vitícola Nacional. (accessed on 20 March 2024). Available online: <https://www.sag.gob.cl>
- Schlich, P., Medel Maraboli, M., Urbano, C., & Parr, W. V. (2015). Perceived complexity in Sauvignon blanc wines: Influence of domain-specific expertise. *Australian Journal of Grape and Wine Research*, 21(2), 168–178. <https://doi.org/10.1111/ajgw.12129>
- Tsai, P. C., Araujo, L. D., & Tian, B. (2022). Varietal aromas of sauvignon blanc: impact of oxidation and antioxidants used in winemaking. *Fermentation*, 8(12), 686. <https://doi.org/10.3390/fermentation8120686>
- Volschenk, H., Van Vuuren, H., & Viljoen-Bloom, M. (2006). Malic acid in wine: Origin, function and metabolism during vinification. *South African Journal of Enology and Viticulture*, 27(2), 123–136. <https://doi.org/10.21548/27-2-1613>
- Walker, R. R., Holt, H., Blackmore, D. H., Pearson, W., Clingeleffer, P. R., & Francis, L. (2023). Salt concentration and salty taste perception in ‘Chardonnay’ and ‘Shiraz’ wines from own roots and different rootstocks under saline irrigation. *Vitis*, 62(4), 151-162. <https://doi.org/10.5073/vitis.2023.62.151-162>
- Wickam, H. (2016). Ggplot2: elegant graphics for data analysis. Springer-Verlag New York.
- World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*, 310, 2191-2194. <https://doi.org/10.1001/jama.2013.281053>