Background

The exposure of wine grapes (*Vitis vinifera* [*V. vinifera*]) to smoke from wildland fire or prescribed burns changes the sensory profile of the berry (*i.e.*, the grape). More specifically, wine made from smoke-exposed berries shows an increased incidence of 'smoky', 'ashy', 'burnt meat' and 'Band-Aid' sensory attributes, all of which are undesirable in a quality product.[1–4] Chemically these negative sensory descriptors are associated with a specific class of compounds called volatile phenols (VP). This phenomenon is particularly problematic for the wine industry in the Okanagan Valley given the frequent occurrence of wildland fires during the growing season. However, it is also important for the global wine industry, as many key growing regions are also located in close proximity to fire-prone regions. For instance, recent reports suggest that the economic impact of wildland fire on the Australian wine industry during the 2009 growing season was $299 million.[5] It is expected that this issue will increase in relevancy, as climate change models are suggesting an increase in the frequency of wildland fires in key wine growing regions (*e.g.*, California, British Columbia, Australia).[5]

Lignin, which accounts for 20-30% of the dry weight of wood, leads to the formation of a variety of VP during combustion. Many of these combustion products are known to correlate with the negative sensory descriptors associated with smoke-exposed berries (Figure 1). However, a subset of VP may also be present endogenously in the berry, where they are found in free (aglycone) and sugar-bound forms (glycosides), with the concentration of the glycosides typically much higher than the aglycones. Adding to the complexity of this problem is the fact that phenolic glycosides (VP-glycosides) may be enzymatically or chemically hydrolyzed during fermentation and aging. As such, despite possessing no sensory properties, VP-glycosides represent a 'sensory potential' that can influence the sensory profile of wine, even years after bottling. Existing methods (using VP and their glycosides) to quantify the risk associated with using smoke-exposed berries are only 50 – 80% predictive of negative sensory attributes in wine, leaving vineyards and wine producers at considerable financial risk.[6] My research aims to obtain a detailed assessment of the chemical composition of smoke-exposed berries (including and beyond VP and their glycosides), which will facilitate the development of a more accurate model for predicting wine quality issues, as well as inform remedial and preventative strategies.

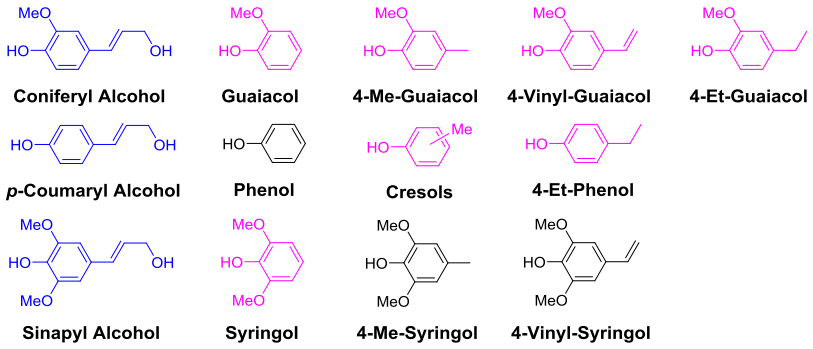

Figure 1: The core units of lignin (blue) and examples of combustion products with demonstrated relevance to the negative sensory properties of smoke-exposed *V. vinifera* berries. Nine of these VP (purple), including three cresols, were assessed quantitatively (*vide infra*).

Experimental Design and Sample Analysis

To assess the impact of smoke on the chemical composition of *V. vinifera* berries, a series of controlled field experiments were conducted. Using a custom-built enclosure that housed nine vines, four

commercial varietals (Merlot, Pinot Noir, Cabernet Sauvignon and Cabernet Franc) were exposed to simulated wildland fire smoke (Figure 2). In an effort to ensure equal exposure for all vines, only the middle five vines were sampled as 'smoked' berries. A separate block of five vines per varietal were used as a control condition (*i.e.*, no smoke exposure). For each condition (smoked *versus* control) and each vine (5/condition) a series of time points were collected from immediately preceding smoke-exposure through until commercial maturity (Figure 2). Each sample was processed to as whole berry homogenate (HMG) and free-run juice (FRJ) to mimic the raw materials for red white wine production, respectively. Finally, a subset of time-points for some varietals were split into two fractions that were either washed or unwashed before processing as HMG and FRJ.

Cultivar	# Samples	# Time Points	Total Samples
Merlot	60	6	180
Cabernet Sauvignon	60	6	120
Pinot Noir	50	5	100
Cabernet Franc	50	5	160

Figure 2: The enclosure used to expose vines to simulated wildland fire smoke, showing the outside (top left), inside (top middle) and inside during smoke exposure (top right). The sample collection and processing scheme resulted in 560 total samples collected from the 2016 growing season. This includes HMG, FRJ, washed and unwashed samples.

In addition to the samples outlined above, a second set was collected by sampling 50-60 vines per varietal over an area corresponding to 1-2 acres. This was done to quantify endogenous levels of key VP in these varietals, which would facilitate a rigorous statistical comparison between control and smoke-exposed berries. These samples were collected at commercial maturity, which corresponded to the last time-point for each varietal. The GPS coordinates of each sample were collected to enable repeated analysis of the same vines across multiple years and to assess the presence of trends as a function of location.

To quantify the concentration of VP known to contribute to the negative sensory attributes of wine made from smoke-exposed berries (Figure 1), targeted analysis of nine VP was conducted. This quantitative analysis was performed on berry extracts using gas chromatography-mass spectrometry (GC-MS). This analytical methodology separates VP in the time domain (GC), then uses a sensitive and specific detector (MS) to ensure the correct signal corresponding to the desired VP are accurately quantified (Figure 3).

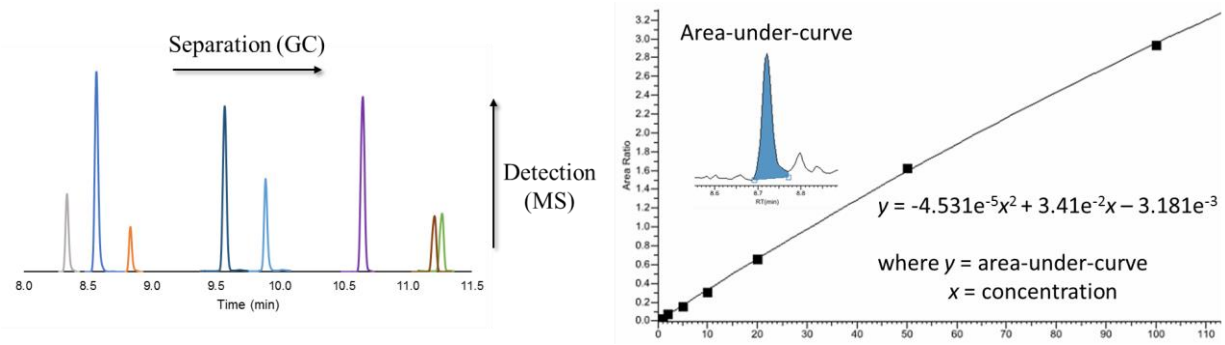

Figure 3: Gas chromatography-mass spectrometry (GC-MS) involves separation in the time domain (GC), with each peak in the above data color-coded to a single compound (left). This is followed by sensitive detection (MS) that is specific to each time-response pair (left). Integration of the area-under-the-curve for each compound and subsequent comparison to a calibration function results in accurate quantitative analyses (right).

To date there has been a myopic focus on VP and their glycosides to predict quality issues in wine made using smoke-exposed berries. While successful and informative, this approach has obvious limitations given that VP and their glycosides are only 50-80% predictive of wine quality issues. Improving this predictive accuracy requires a broad comparison of the chemical composition of smoke-exposed and control berries. The use of mass-spectrometry-based non-target screening workflows (*i.e.*, metabolomics) will facilitate this characterization. My approach uses ultra-high pressure liquid chromatography (uHPLC) to separate compounds prior to detection using MS. Conceptually this is similar to the GC-MS approach described above. However, in this workflow MS detection is non-targeted, often producing in excess of 10,000 unique masses per sample that need to be mined for significance. Moreover, rather than producing a simple-to-interpret VP concentration the output from non-targeted analysis is qualitative, producing a mass that the analyst must assign significance to. This is most often achieved by assigning an empirical chemical formula, or by performing a statistical comparison using a given accurate mass (Figure 4). As a final layer of complexity, the MS used in this study produces masses accurate to the fourth decimal place, but each measurement also has an uncertainty associated with it (± 2-5 parts-per-million). This creates a mass binning problem that must be addressed before significance can be assigned.

Data Analysis

The desired output from quantitative GC-MS analysis is a table of VP concentrations that need to be correlated back to specific experimental conditions (*vide supra*). Based on the tested sample matrix for control *versus* smoked-exposed wine grapes, GC-MS analysis will generate a total of 5,220 (580 samples x 9 VP) unique VP concentrations that need to processed into meaningful results summaries. As well, the samples associated with specific GPS coordinates (1,980 unique VP values) need to be visualized to assess spatial trends and determine summary statistics to establish baseline levels of VPs in control vines. To manage these large data sets and enable data from future studies to be easily integrated into an efficient data management system, I will build a series of relational databases in Access that will contain the quantitative GC-MS results and the associated metadata. Using SQL queries for data reduction and analysis, I then propose to take queried data sets and export them into R for statistical analyses (mean, standard error of the mean, t-test with unequal variance, *etc.*) and to Tableau and R for data visualization. As well, I intend to use R to model left censored data (where some samples quantitate below a defined threshold) when calculating summary statistics for the GPS data. This process will require an evaluation

of the degree of censoring per VP and varietal, determining the nature of each distribution (*e.g.*, normal, log-normal, *etc*.) and finally, applying the appropriate statistical tests to obtain values for the censored data.

To facilitate the generation of maps in Tableau I will create a conversion algorithm in Excel to change GPS coordinates in degrees:minutes:seconds to decimal degrees, which is the format required by Tableau. After generating the appropriate background map in Tableau, I will use the converted GPS coordinates to map the concentration of VPs over the areas surveyed.

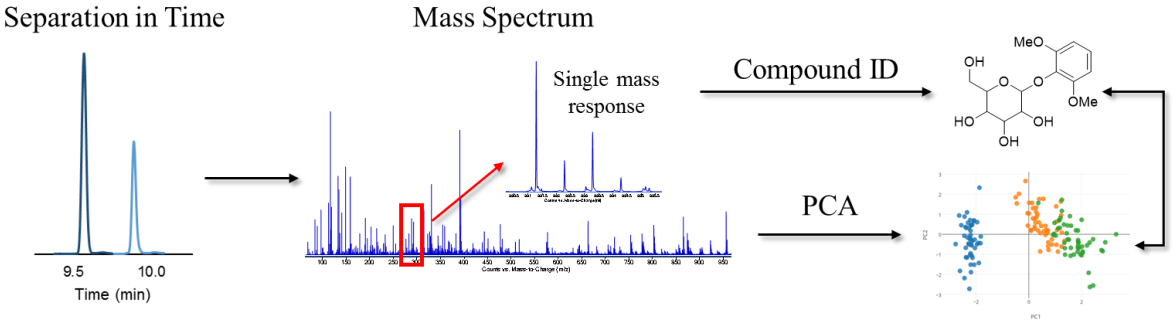

Figure 4: A simplified representation of a potential non-targeted (metabolomics) screening workflow. Following separation in the time domain a series of mass spectra are generated. The spectra can be mined for significance (*e.g.*, principle component analysis [PCA]), after which key mass responses can be targeted for compound identification, or *vice versa*.

A non-targeted analytical workflow produces data that requires a much different data analysis stream than quantitative analysis (Figure 4). Given the complexity associated with mining uHPLC-MS data, which involves identifying and binning relevant MS responses that correlate to chromatographic peaks (as per Figure 4), instrument vendor software will be used (MassHunter, Agilent Technologies) in tandem with custom-built Excel tools to facilitate data reduction.

On the front end of data reduction, Excel will be used to provide a list of masses that MassHunter should search for. This type of workflow is referred to as 'known-unknown' screening. Providing this list requires the calculation of exact masses for a series of chemical formulae. Most MS vendor software packages enable the calculation of the exact mass for a single chemical formula. However, when a series of formula conversions are required the user is left to do this one-by-one. To improve the efficiency of this process I will build an exact mass calculator in Excel. This tool will take a chemical formula (with a defined set of elements), parse out the quantity of each element and return the exact mass. As well, I will build an Excel template that will perform simple *in silico* chemical conversions to aid in the generation of a combinatorial database for VP-glycosides.

Results and Discussion

Targeted Data Analysis

Excel was used for unit conversion and parsing to take GPS coordinates in degrees:minutes:seconds (dd° mm.sss') to decimal degrees, which is the format required by Tableau (Figure A4). In this formula the LEFT() and MID() functions were used to parse out the degrees, minutes and seconds from the coordinates obtained from a GPS device (*e.g.*, 49° 50.423'). The parsed values were converted to decimal degrees as part of the same function. The final step involved an IF() statement that checked the direction of each coordinate and assigned a negative value if the direction was West or South.

Using the GPS coordinates[1] (in decimal degrees), the ability of Tableau to efficiently visualize the spatial distribution of VPs was evaluated for four grape varietals from two vineyards in the Okanagan Valley. To add interest to the final visualization, the addition of satellite images in lieu of the stock Tableau base maps was explored. Adding a satellite image in Tableau currently requires the use of Mapbox Studio as the source of satellite imagery. Going through this process yielded a more interesting and informative map, as the rows at each vineyard can be seen underlying the VP data (Figure A4). After importing the quantitative results from GC-MS analysis into Tableau a dashboard was created to enable simultaneous review of all varietals (Figure A5). Each map in the dashboard was generated with the varietal coded by color and the amount of VP coded by size. To enable visual assessment of data trends VP concentrations were binned to improve size discrimination. From figure A5 it is apparent that pinot noir has a higher guaiacol concentration than the other three varietals. It remains to be seen if this elevated level of guaiacol results in an increased susceptibility to sensory issues in wine following pinot noir grapevine smoke-exposure. Moreover, it could be argued that summary statistics for each VP in each varietal (*e.g.*, box-whisker plot) would be a more efficient data comparison. This argument seems valid for peer-reviewed literature. However, for graphically representing this data in conferences and presentations the use of the Tableau map in tandem with summary statistics visually guides the audience as to how the data should be interpreted, which is a more effect means of conveying large amounts of data with complex relationships.

To organize the quantitative results, a series of relational databases were created in Access (Figure A6). These relational databases were interrogated using SQL queries to generate reduced data sets. It is my intention to take the reduced data sets and use R Studio to calculate appropriate summary statistics and a combination of R Studio and Tableau to generate publication quality summary figures. As an example of this approach, I used an SQL query to tabulate the percent of left-censored data for each varietal (see smokeVolatiles.mdb for query code). The percentage of left-censored data, the distribution of the data (*vide infra*) and the number of replicates determine the statistical methods that can be used to model censored data. Choosing an accurate censorship model enables the calculation of summary statistics with minimal bias, as might be obtained from the more common practice of arbitrarily substituting one of several values (*e.g.*, zero, detection limit/2, *etc*.) for censored data. After saving the query results in Access, I linked Tableau to the Access database and generated a summary plot indicating the suggested course of action for each data set by color code (Figure A7). Displaying the data in this

[1] Note that the results presented in this update are not intended to be the final evaluation of the spatial distribution data. More accurately, they are being used in their current (raw) form as placeholders to develop and demonstrate the visualization and data reduction schemes outlined above. Much of these results are left-censored and require further statistical treatment before strong conclusions can be reached. It is my intent to have this analysis completed in R for my final report

way was an efficient summary for an audience unfamiliar with my methodologies, as it clearly indicates how the GPS data were treated statistically to produce a meaningful set of summary statistics.

Non-Targeted Data Analysis

To support 'known-unknown' screening a parsing tool was created to do batch calculations of exact masses given a list of chemical formulae as input (Figure A1). This tool used the FIND() and ISERROR() functions to index the constituent elements from each formula string. The ISERROR() function was used to return '0' if no elements of given type were found, which was required to enable correct referencing in subsequent steps. After indexing, an iterative IF() statement was generated to pull out the number of atoms of each element. The IF() iterations locate the index for each element in the previous step using the MID() function, then define the length of each number sequence using the ISNUMBER() function. The number of IF() iterations used limited the maximum number of each element to 999, which is an acceptable range for the type of analyses conducted as part of my PhD research. The final piece of this tool was the calculation of the monoisotopic mass, which involved the sum of absolute references to accurate masses for each element multiplied by the quantity of each element parsed from the input formulae.

A combinatorial database of VPs and sugars known to be involved in glycoside formation in wine grapes was constructed using the exact masses of the constituent components. These masses were batch-calculated using the Excel tool described above. The glycoside chemical formulae were calculated by referencing the parsed chemical formulae of each component of the glycoside and adding up the constituent elements via a series of VLOOKUP() functions in tandem with concatenation of the resulting sums (Figure A2). The exact mass of the glycosides were calculated using VLOOKUP() to find the exact mass of the constituents and sum them. For the calculation of chemical formulae and exact masses, the loss of H_2O during the formation of a new glycosidic bond was accounted for.

The combinatorial database was imported directly into the MassHunter software to interrogate uHPLC-MS data. For this analysis, a series of control and smoke-exposed Cabernet Franc wine grapes were used. As an example of the output from MassHunter for this analysis, a schematic of the final workflow is shown in Figure A3. The first step required input in the form of exact masses from the combinatorial glycoside database. The MassHunter software then utilizes a proprietary algorithm (with exact masses as input) to identify potential matches in uHPLC-MS data. From the Cabernet Franc data analyzed, a variety of putative VP-glycosides were identified at elevated levels in the smoke-exposed sample set when contrasted to the control wine grapes (data not shown). Of particular interest was the strong match for syringyl-glucuronide, as the glucuronic acid family of sugars has not been reported in the literature. If this finding can be confirmed, it would mean that current methods for assessing the impact of smoke-exposure neglect an entire class of glycosides. It remains to be determined if such glycosides are present at high enough levels in smoke-exposed wine grapes to influence sensory perception in the resulting wine.

Bibliography

[1] R. Ristic, A. L. Fudge, K. A. Pinchbeck, R. De Bei, S. Fuentes, Y. Hayasaka, S. D. Tyerman, K. L. Wilkinson. Impact of grapevine exposure to smoke on vine physiology and the composition and sensory properties of wine. *Theor. Exp. Plant Physiol.* **2016**, *28*, 67.

[2] R. Ristic, K. A. Pinchbeck, A. L. Fudge, Y. Hayasaka, K. L. Wilkinson. Effect of leaf removal and grapevine smoke exposure on colour, chemical composition and sensory properties of Chardonnay wines. *Aust. J. Grape Wine Res.* **2013**, *19*, 230.

[3] K. R. Kennison, M. R. Gibberd, A. P. Pollnitz, K. L. Wilkinson. Smoke-derived taint in wine: The release of smoke-derived volatile phenols during fermentation of Merlot juice following grapevine exposure to smoke. *J. Agric. Food Chem.* **2008**, *56*, 7379.

[4] K. R. Kennison, K. L. Wilkinson, H. G. Williams, J. H. Smith, M. R. Gibberd. Smoke-derived Taint in Wine : Effect of Postharvest Smoke Exposure of Grapes on the Chemical Composition and Sensory Characteristics of Wine Smoke-derived Taint in Wine : Effect of Postharvest Smoke Exposure of Grapes on the Chemical Composition and Senso. *J. Agric. Food Chem.* **2007**, *55*, 10897.

[5] M. P. Krstic, D. L. Johnson, M. J. Herderich. Review of smoke taint in wine: Smoke-derived volatile phenols and their glycosidic metabolites in grapes and vines as biomarkers for smoke exposure and their role in the sensory perception of smoke taint. *Aust. J. Grape Wine Res.* **2015**, 537.

[6] M. Parker, P. Osidacz, G. A. Baldock, Y. Hayasaka, C. A. Black, K. H. Pardon, D. W. Jeffery, J. P. Geue, M. J. Herderich, I. L. Francis. Contribution of several volatile phenols and their glycoconjugates to smoke-related sensory properties of red wine. *J. Agric. Food Chem.* **2012**, *60*, 2629.