

Quantitative metabolomics unveils the impact of agricultural practices in grape metabolome

Ilce G. Medina-Meza^{1†*}, Joshua VanderWeide^{2†}, Cristobal Torres-Palacios³ and Paolo Sabbatini⁴

¹ Department of Biosystems and Agricultural Engineering, Michigan State University, East Lansing, Michigan 48824, United States; ilce@msu.edu

² Department of Horticulture, Michigan State University, East Lansing, Michigan 48824 and United States; vande732@msu.edu

³ Department of Biosystems and Agricultural Engineering, Michigan State University, East Lansing, Michigan 48824, United States; cristobal.torres.palacios@gmail.com

⁴ Department of Horticulture, Michigan State University, East Lansing, Michigan 48824, United States; sabbatin@msu.edu

* Correspondence: ilce@msu.edu; Tel.: 517-884-1971

† These authors equally contributed to the manuscript

Abstract: Integration of multiple data set in agricultural and food practices is critical for decision-making, particularly if affecting product qualitative characteristics that influence producers and consumers decisions on production strategies and purchases. Herein, we apply a multidimensional data analysis to evaluate grape chemical composition obtained via high-resolution metabolomics and vine growth characteristics following the application of early leaf-removal (ELR), a canopy management technique implemented in cool climate viticulture. The application of discriminant analysis using a supervised PLS (sPLS-DA) algorithm along with MANOVA reveals that ELR enhanced the concentration of several secondary metabolites, without compromising other berry attributes pivotal for the vinification process, such as sugar content. Overall, this study paves the way for highly effective integrated strategies (metabolomics and agricultural practices) that link laboratory analysis with vineyard management decisions.

Keywords: *Vitis vinifera*; metabolomics; chemometrics; leaf removal; secondary metabolites; agricultural practices; multivariate analysis

1. Introduction

Metabolomics has become an increasingly powerful molecular phenotyping technique for a broad variety of studies, along with established omics technologies for high throughput DNA sequencing (genomics), protein profiling (proteomics) and mRNA expression (transcriptomics), by disposing causality between genetic polymorphisms and variations in concentration of metabolites in biological samples (1-3). Moreover, metabolomics allows for the exploration of the nexus among gene-environment-agriculture and their interactions. Quantitative metabolomics data analysis and interpretation pose numerous statistical challenges; this is particularly evident when *omics* data from different sources are integrated (4). Routinely in food and agricultural fields, data interpretation is critical for decision-making, as choosing agronomical practices or processing technologies that can better satisfy both economic and qualitative needs. Since the major objective of quantitative metabolome profiling is to determine differences in metabolites concentrations from comparative samples, the size, weight, or volume of individual samples may significantly impact the relative quantification results. Moreover, data processing is a critical step where normalization of metabolic data minimizes the impact of unwanted biases and experimental variance in any subsequent analysis. One critical pre-processing step is the normalization of the metabolomic data. Most common normalization methods include constant noise, linear baseline (5), constant sum (6) and probabilistic

quotient approaches (7). Generally, univariate statistical methods, as *t*-test and analysis of variance (ANOVA), are unable to catch interactions among metabolic features, and may fail to obtain critical biological information. Thus, these challenges can be overcome by multivariate approaches which model features as a set, and complement results obtained from univariate methods. Quantitative multivariate metabolomics relies on several computational algorithms, including unsupervised and supervised statistical methods, as principal component analysis (PCA) and partial-least squares (PLS), respectively (4, 8). Both approaches share common characteristics: they a) reduce the number of variables to a few relevant features, and b) provide easy-to-interpret model assisted by graphic representations. Although extensively used in chemometrics studies (9–11), they are still marginally adopted in quantitative agriculture metabolomic studies.

Recently, we investigated extensively the effect of an innovative cluster-zone leaf removal technique on grape technological maturity, by collecting grapevine canopy, cluster, and yield parameters in addition to quantifying a number of secondary metabolites across two growing seasons (12). We found evidences that early leaf removal (ELR) applied around grapevine bloom, enhanced fruit quality; however, the complexity of the dataset led to several approximations that hindered a comprehensive biological assessment (12). Therefore, the present study aimed to examine metabolic signatures associated with viticulture management strategies on field-grown grapevine by applying analytical chemistry and quantitative metabolomics to promote the advantages of multivariate statistics, by providing a real example (ELR in grapevine) where a comprehensive metabolomics assessment can help directing agronomical decisions due to the seasonal variations. In addition to MANOVA, we used the recently developed statistical algorithm sPLS-DA, a sparse version of the PLS for discrimination analysis (13). To the best of our knowledge has scarcely been used for fruit quality assessment (14). The generated knowledge will encourage food and agriculture scientists to adopt novel multivariate approaches for omics data integration for more accurate decision-making purposes.

2. Materials and methods

2.1. Vineyard and Plant material

The experiment was conducted in a commercial vineyard (Lemon Creek Winery, 41° 96' N, 86° 44' W) situated in Michigan (USA), during the seasons 2016 and 2017. Eight year-old grapevines (cv. Merlot, clone 14) grafted on 101–14 rootstock, were planted in a Spinks loamy fine soil with a spacing of 2.3 m between vines and 3.0 m between rows. Vines were trained to a bilateral vertical-shoot-position (VSP) Guyot training system and cane-pruned during the two winters, leaving 40 buds per vine. The crop level was adjusted 4 weeks after flowering to about 44 clusters per vine in both years through the removal of tertiary clusters and shoots containing a single cluster, leaving fruit equally distributed between apical and basal clusters on each shoot. Recommended crop-protection practices were followed, and the pest-management program was based on scouting experience and weather conditions. Weather data, including daily temperature and daily precipitation, were recorded during the experiment by an automated weather station in Berrien Springs, MI, from the Michigan Automated Weather Network (MAWN), located 6.0 km from the vineyard. Growing degree days (GDD) were calculated with the Baskerville–Emin method using a base temperature of 10°C. No irrigation was used, and standard summer vineyard practices were applied, with shoots mechanically hedged on July 15 and 14 in 2016 and 2017, respectively.

2.2. Sample collection

Each year, the experiment was conducted in five rows of vines where treatments were applied to 10-vine plots in a randomized block design with three blocks. Within treatment plots, three 'target' vines were chosen randomly, and three representative shoots per vine were tagged. Four different leaf-removal strategies were used: pre-bloom hand leaf removal of six basal leaves (PB-Man), pre-bloom mechanical leaf removal (PB-Mech), after-bloom hand leaf removal of six basal leaves (AB-Man), after-bloom mechanical leaf removal (AB-Mech), and one control with no cluster-zone leaf

removal at veraison (Ct). 60 berries per replicate were collected weekly from the basal cluster on the three tagged shoots (20 berries per vine, ~7 berries per cluster). Each berry was removed from clusters by cutting the pedicel at approximately 3–4 mm from the receptacle using scissors. 30 of these berries were placed in Ziploc bags and returned to campus in a cooler, where processing took place the following day for basic fruit chemistry analysis. The other 30 berries were immediately frozen in liquid nitrogen and used later for metabolomic analysis. Experimental measurements are reported in detail elsewhere (12). Data collection on vine growth and development, cluster morphology and fruit chemical parameters, were performed over the two seasons. Slightly different weather conditions, in 2016, seasonal growing degree days (GDD) were 1980 with a mean day temperature during vine growth and development (June–September) of 22.5 °C. 2017 seasonal growing degree days (GDD) were 1682 with a mean day temperature (June–September) of 20.3 °C.

2.3 Fruit Set, Yield, and Cluster Morphology.

Clusters from 20 random shoots gathered at bloom and fruit-set were photographed in the field against a dark background, and the floret counts in each image were estimated using the software ImageJ (Version 1.51). The same 20 random clusters were harvested, and destructively counted. Linear relationships were evaluated between the actual number of florets (Y) and the florets calculated from photos. Regressions for 2016 and 2017 were obtained (2016 $Y = 1.421X$, $r^2 = 0.89$ and 2017 $Y = 1.537X$, $r^2 = 0.88$) respectively. Moreover, regressions between the actual number of berries (Y) and the berries calculated from photos (X) were also developed. The regressions were: in 2016 ($Y = 1.685X$, $r^2 = 0.92$) and 2017 ($Y = 1.367X$, $r^2 = 0.84$). These regressions were applied to estimate floret counts from field photos of target clusters to estimate the true numbers of florets (EL-17) and berries (EL-27) nondestructively and the percentage of fruit set at harvest (FS-38). At harvest, the yield per vine and the total number of clusters per vine were recorded and weight measured using a portable Pelouze 4010 scale. Clusters were harvested, placed in quart-size resealable bags, and immediately transported to campus for storage in –20 °C until further analysis. Clusters were thawed and initially weighed before counting the number of berries. The weight and length of the rachis, as well as the wing length, were recorded, and then the cluster-compactness index (CI) was calculated as reported in (15). A subset of ten frozen berries per treatment plot were used for separation of skin, peel, and seed tissues and analysis of their average weight. Before removal of skins, berries were weighed (g) and individually measured for diameter (mm). After skins were removed, both skins and seeds were blotted dry with a KimWipes™ and weighed, and the pulp weight was calculated by their subtraction from the individual-berry weight.

2.4 Berry Composition.

Following cluster-morphology analysis, berries from individual clusters were crushed within Ziploc® bags at room temperature, and the free-run juice was collected into 100 mL beakers. °Brix was measured using a digital refractometer (ATA-3810 PAL-1, Pulse, Inc.), and pH was measured with a 370 Thermo Orion pH meter (Thermo Fisher Scientific, Inc.). Total acidity (TA) was analyzed using a Multi-T 2.2 digital titrator (Laboratory Synergy, Inc.) with each sample consisting of 10 mL of clear juice diluted to 100 mL with distilled water and titrated with 0.1 M sodium hydroxide to pH 8.2 using an equation to yield the TA (g/L).

2.5 LC-MS-MS Method

Whole berries from each treatment plot (n=30) were weighted and grounded into a fine paste in a 50 mL conical tube using a Tekmar 1810 (110 V) homogenizer at 4000 rpm for 10 min in an ice bath to avoid oxidation. Then, 15 mL of 80% methanol and 1% formic acid (v/v) were added to the homogenate, vortexed for 10s, and sonicated in an cold Ultrasonic Bath (1.9 L, 117 V, 50 Hz, VWR) for 30 min to fully degrade cell-wall structure. Then, the sample was centrifuged (Eppendorf 5810R) at 4000 rpm for 10 min at 4 °C; the supernatant was collected and filtered through a 0.45 µm syringe filter. 100 µL were used for total phenolic quantification by Folin-Ciocalteu method (16).

A targeted metabolomic profiling of color specialized metabolites (both anthocyanins and flavonols) was performed with LC-MS/MS, using telmisartan (Sigma-Aldrich) as the internal standard. Samples were analyzed using a Waters Acquity UHPLC interfaced to a Waters Xevo G2-XS Q-ToF mass spectrometer. 10 μ L of filtered supernatant were injected onto a Waters HSS-T3 UHPLC column (2.1 μ m, 100 mm, 1.7 μ m particle size) held at 40 °C using a binary gradient of water with 0.1% formic acid (solvent A) and acetonitrile (solvent B). The solvent flow rate was 0.3 mL/min and it was started at time 0 with 100% (solvent A), held for 0.5 min at 100% (solvent A), ramped to 50% (solvent B) at 6 min, ramped to 99% (solvent B) at 6.5 min, held at 99% (solvent B) until 8.5 min, reverted back to the starting condition of 100% (solvent A) at 8.51 min, and held until 10 min. Compounds were ionized by electrospray ionization in positive-ion mode with a capillary voltage of 3.0 kV, a cone voltage at 35 V, a source temperature of 100 °C, and a desolvation temperature of 350 °C. Data were acquired using a data-independent MSE method that consisted of two separate acquisition functions, one with no collision energy and the other with a collision-energy ramp of 20–80 V. No standards were used, instead accurate masses were obtained using lock mass correction with leucine enkephalin and the reference compound (**Table S1**, Supplemental information). Fifteen anthocyanins, five flavonols, and phenylalanine, and were identified based on their accurate masses, relative retention times, and fragmentation patterns at 3 MSI reporting level according to Sumner, et al. (17). Peak areas were obtained using Quanlynx (part of the Waters Masslynx software) and normalized to the internal-standard abundance and to the dry weight of the extracted material. Three independent biological replicates were performed to assure reproducibility of our analysis.

2.6. Sugar analysis.

Berry sugars were resolved and quantified using a Shimadzu Prominence HPLC system, equipped with a RID-10A detector according to Mikulic-Petkovsek and coworkers (18). Briefly, grape juice was 200-fold diluted with nanopure water and filtered through a 0.45 μ m nylon membrane (Phenomenex, Torrance, CA), then 5 μ L of sample was injected into a 250 mm $\times$ 4 mm Kinetex® RCM monosaccharide Ca^{+2} column (Phenomenex, Torrance, CA). Column temperature was set at 80 °C. The mobile phase was HPLC-grade water at a 0.6 mL/min flow rate, run time was 30 min, and RI was used to monitor the eluted analytes. Carbohydrate LC performance check Mix (Restek®) was used to build standard curves from 1.0 mg/mL to 0.02 mg/mL ($p < 0.05$, R^2 0.97), and retention times were obtained. All analyses were performed in triplicates.

2.7 Description of the dataset for multivariate analysis.

We used a three-steps workflow for quantitative metabolomics of grape metabolome: pre-analytical work, analytical work, and data processing (**Figure 1**). Pre-analytical work included vineyard sample collection, storage, and sample preparation in the laboratory. The analytical step included sample processing and metabolomics. Finally, after data acquisition, data processing and multivariate analysis were applied to extract information on identity, absolute and/or relative abundance of individual metabolites.

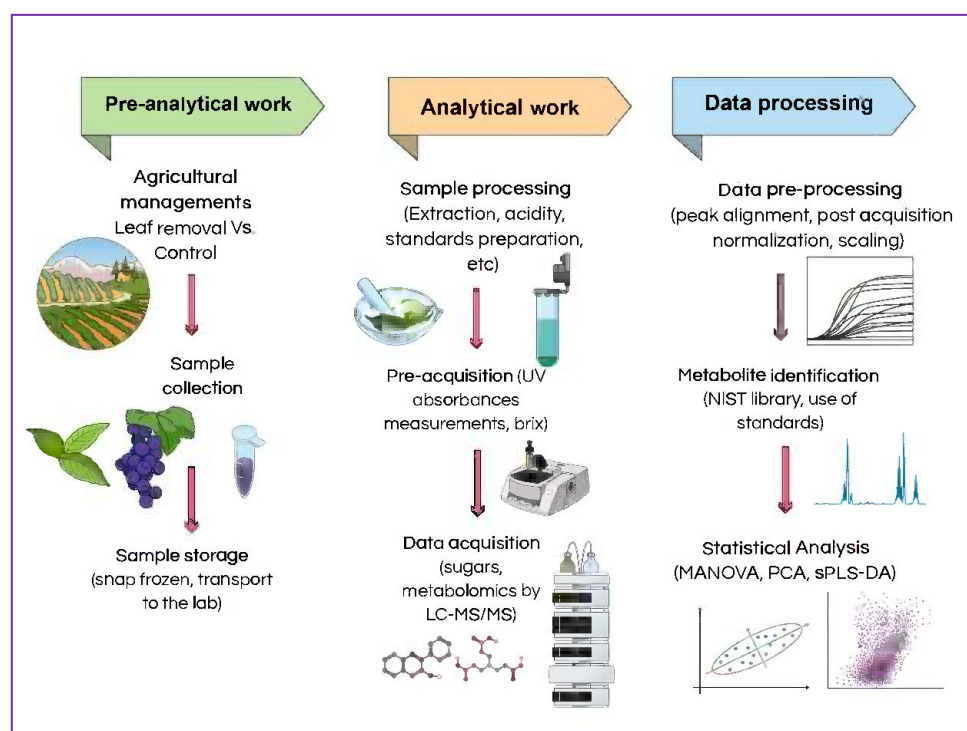


Figure 1. Workflow for general quantitative metabolomics.

The final dataset contained 26 metabolic features and 23 morphological and chemical variables measured on vines or grape clusters. The complete morphological and chemical variables used for multivariate analysis are listed in **Table 1**.

Table 1. Description of the parameters utilized for the statistical analysis.

Parameter	Description
Yield (kg/vine)	Weight of the total clusters per vine measured at harvest
Cluster weight (g)	Average cluster weight measured at harvest
Cluster number	Cluster number per vine at harvest
Rachis length (cm)	Length of the main rachis of the cluster measured at harvest
Wing length (cm)	Length of the cluster wing measured at harvest
Rachis weight (g)	Weight (g) of the cluster rachis measured at harvest
Cluster compactness	The ratio between total berry number per cluster and the rachis length, expressed as the cluster compactness index (CCI).
Final berry number	Number of berries per cluster measured at harvest
FS-38	Fruit-set calculated as the ratio between the number of flowers per cluster at bloom and the number of berries at harvest.
FS-27	Fruit-set calculated as the ratio between the number of flowers per cluster at bloom and the number of berries at berry set
Flower number	Number of the flowers at bloom
Berry diameter (cm)	Berry diameter measured at harvest
Pea-size berry number	Number of berries after fruit set
Berry weight (g)	Berry weight measured at harvest
Harvest berry number	Number of berries per cluster at harvest
Skin weight (g)	Weight of the berry skin measured at harvest
Pulp weight (g)	Weight of berry pulp measured at harvest
Skin/pulp ratio	Ratio between the weight of the berry skin and the weight of the berry pulp measured at harvest
Seed weight (g)	Weight of the seeds in the berry measured at harvest
Seed number	Number of berry seeds measured at harvest
Individual seed weight (g)	Average seed weight measured at harvest
Total soluble solids (°Brix)	Brix concentration measured by refractometer of the berries at harvest
pH	pH of the berry juice measured at harvest
Titrateable acidity (g/L)	Titrateable acidity of the berry juice measured at harvest

2.8. Classification methods data analysis.

Data referring to sugar accumulation (sucrose, glucose and fructose) in grape berry was statistically evaluated by multivariate analysis of variance (MANOVA), considering three dependent variables (year, time and treatment). Results were expressed as the mean value $\pm$ standard deviation (SD) of the independent measurements ($n = 3$), performed for both years (Tables S2-S9 Supplemental information). Wilks's lambda (Λ) and Pillai's tests were applied to examine the individual main effect and combined effects of year, time, and treatment on grape sugars accumulation. Secondary metabolite variables and production variables were also assessed by MANOVA examining the individual and combined effect of year and treatment, using the Λ -test. MANOVA was performed in SPSS v.22 (IBM, Armonk, NY). Principal Component Analysis (PCA), and sparse-Partial Least Square Determinant analysis (sPLS-DA) were performed using MetaboAnalyst in its webserver version (19). sPLS-DA is based on the use of PLS regression (20) for discriminant purposes; a Lasso penalization is added to select only a certain number of parameters. The sPLS algorithm used in this work is based on the PLS-SVD approach, and it has been shown to be robust for both binary and multiclass problems (21). Before performing the multivariate algorithms, data were mean centered. For PCA, the main principal components were chosen by visualization of the scree plot and considering the cumulative variance (8). For sPLS-DA, ten variables were kept per each principal component. The performance of the sPLS-DA model was assessed by cross validation, and the number of components to be discussed was chosen by looking at the error rate (14) and the cumulative variance. All graphs were made in OriginPro v. 2019 (OriginLab, Northampton, MA, USA).

3. Results and discussion

3.1. Quantification of grape metabolome

Intervention with leaf removal significantly reduced the total area per vine in both years (2016-2017). PB-Man has the highest percentage of leaf area removal followed by PB-Mech, on the contrary AB-Mech has the lowest leaf area removal percentage followed by AB-Man (Figure 2). Thus, PB-Mech was more efficient in decreasing cluster compactness, similar to the PB-Man, which was also observed in previous research (22, 23).

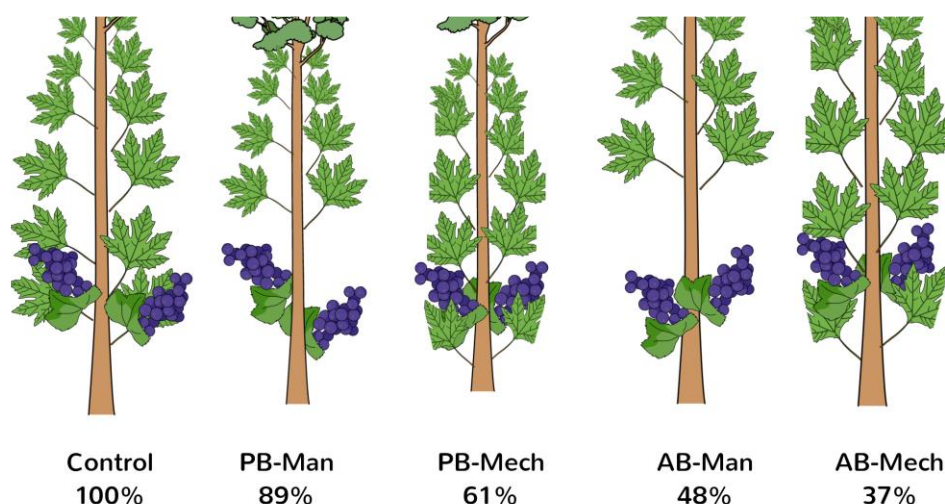


Figure 2. Depiction of fruit-zones from representative shoots of treatments and percentage of leaf area removed average of the 2 experimental years. C, control; PB-Man, pre-bloom manual leaf removal; PB-Mech, pre-bloom mechanical leaf removal; AB-Man, after-bloom manual leaf removal; AB-Mech, after-bloom mechanical leaf removal.

With the purpose to validate the robustness of viticultural practices across seasonal changes, vine growth and development assessments included yield components (yield per vine, cluster

weight, and berry weight and diameter), cluster morphology and basic fruit parameters (°Brix, pH and titratable acidity). In contrast to our previous study (12), metabolomics data were not averaged between seasons (2016 and 2017). We completed the grape metabolome portfolio by quantifying sugars as critical primary metabolites, aiming to obtain a deep understanding of the effects of ELR on grapevine physiology and fruit quality (Figure 3).

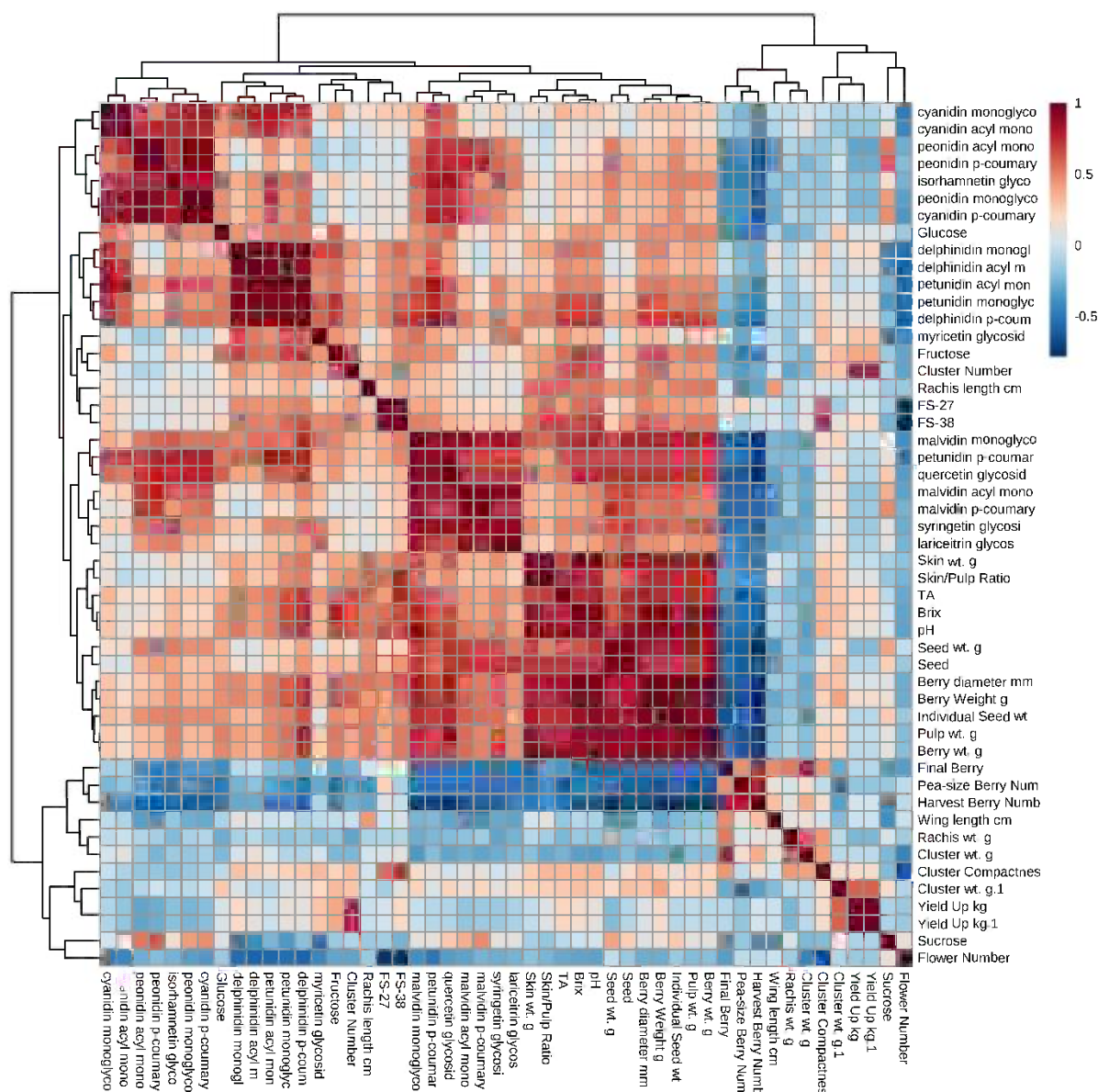


Figure 3. Pearson Correlation map of the metabolomic and agronomical responses.

3.2. MANOVA reveals overall robustness of leaf-removal strategies.

Results of MANOVA analysis are presented in **Table 2**. Agricultural practices (indicate as ‘treatment’) and season (‘year’), as well as their interaction in both fruit metabolomic and vine and cluster variables were tested. Despite similarities in climate between growing seasons, Wilks’ lambda (Λ) was lower for the “treatment” variable (2.18×10^{-9} , $p = 0.001$), than for the “year” variable (4.48×10^{-5} , $p = 0.023$), but both significant at $p < 0.05$. This is different from many studies showing the greater effect of year than treatment in agriculture (24, 25). Specifically, season variability affected, beside glucose accumulation as described in the previous section, seed number and weight ($p < 0.001$), berry number and weight ($p < 0.05$), without significantly affect the accumulation of secondary metabolites,

except peonidin- and delphinidin-glucosides ($p < 0.001$). *Vice versa*, agricultural practices heavily impacted the profile of secondary metabolites, except for malvidins-based anthocyanins ($p > 0.05$).

Table 2. MANOVA results for the metabolomic and vegetative data considering individual and combined effects.

	Effect	Value	F	Hypothesis df	Error df	Sig.
Treatment	Pillai's Trace	3.690	2.507	76.000	16.000	0.021
	Wilks' Lambda	2.18×10^{-9}	13.079	76.000	6.330	0.001
Year	Pillai's Trace	1.000	1175.423 ^a	19.000	1.000	0.023
	Wilks' Lambda	4.48×10^{-5}	1175.423 ^a	19.000	1.000	0.023
Treatment * Year	Pillai's Trace	3.461	1.351	76.000	16.000	0.255
	Wilks' Lambda	1.65×10^{-8}	7.791	76.000	6.330	0.006

^a Exact statistic.

Overall, MANOVA confirms the robustness of our viticultural strategy against relatively small seasonal variability, suggesting a deeper statistical surveillance, which will be discussed in the next section.

3.2. Sugar accumulation during ripening.

Multivariate analysis of variance (MANOVA) was performed to evaluate the role of seasonal impact and/or viticultural management on sugar accumulations. Three variables (treatment, time, year) and their interactions were tested. Viticultural interventions had a significant effect on sugar composition, disregarding the seasonal variations. This is probably due to overall similar seasonal heat accumulation of the two seasons (12, 26, 27), which conditioned vines to respond similarly in both years to the ERL treatments (28). Wilks' lambda (Λ) was lower for the "treatment" variable (0.318, $p < 0.001$), whereas $\Lambda = 1$ for the "year" variable, indicating no significant seasonal differences. Variation during ripening was also significant ($\Lambda = 0.310$, $p < 0.001$); however, the combination "time $\times$ treatment" (i.e. comparing the accumulation of sugars during ripening when different treatments were applied), was not statistically different ($\Lambda = 0.431$, $p > 0.05$). Between years, only glucose accumulation was significantly higher in 2016 ($p < 0.05$). A closer look at the pairwise comparison between treatments indicates that mechanical treatments (both AB and PB) were not statistically different among them ($p > 0.05$); however, performing a manual leaf removal decreased the accumulation of both glucose and fructose ($p < 0.05$) at both pre-bloom and after-bloom stages.

4.2. sPLS-DA discloses potential of PB-Removal.

There are several discriminant and classification techniques suitable to be applied to medium-to-large *omics* dataset. In the past, we successfully used principal component analysis (PCA) for exploratory analysis of metabolites profiles (29, 30). The goal of PCA is to replace all correlated variables by a smaller set of uncorrelated variables, known as principal components (8), that retain most of the information of the original dataset. In the present work, PCA was unsuccessful in discriminating among treatments (**Figure 4a**) and treatments scattered across the first three principal dimensions. Limitations of PCA are particularly evident when integrating *omics*: PCA decomposition relies on the maximization of variance, whereas biological behaviors are often not related to the highest variance, being most features unchanged during physiological transitions (31).

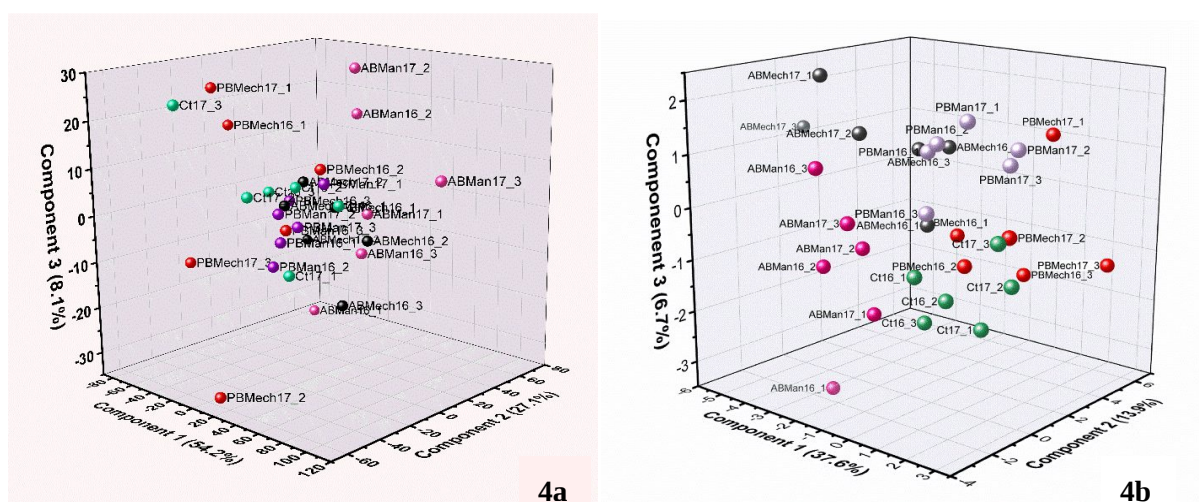


Figure 4. Principal Component Analysis (PCA) of the data set (4a) compare against 3D score plot between the selected three principal components (4b). The explained variances are shown in brackets.

We opted to use the sPLS-DA, which has been shown to be a powerful statistical method to classify large dataset by grouping critical variables (14). Although PLS algorithms are widely used for predictive purposes, they also have great classification outcomes when combined with supervised classification techniques, like DA. Particularly, sPLS-DA has additional advantages in producing robust and easy-to-interpret models (14). When applied to our dataset, sPLS-DA provided a clear landscape of the impact of the viticultural management technique (ELR) on grape quality at harvest. As for PCA, variables were combined in principal components, the first four explaining 68.3% of variance (**Figure 4b**). Compared to PCA, the four treatments and the control are well clustered, despite of season (2016 vs. 2017), which is a good indication that the leaf removal practices are consistent even in conditions of year-by-year weather differences. After-bloom leaf removal samples are more scattered than pre-bloom ones, independently of the method of application (manual vs. mechanical). The loading for variables selected by the sPLS-DA model for each component are presented in **Figure 5**, as well as the impact (low-to-high, as absolute value) of each loading in the considered treatment. The first component (PC1, 37.6% variance) contrasts variables characterizing decreased berry maturity (i.e. °Brix) and berry attributes (weight and diameter), with the harvest berry number per cluster. Looking at the overall impact of the variables, PC1 places AB leaf removal strategies (manual and mechanical) downside in respect to PB ones, being AB-Mech<AB-Man in terms of several berry attributes. The second component (PC2, 13.8% variance) describes almost exclusively positive accumulation of anthocyanins monoconjugates. Interestingly, this component can classify mechanical vs. manual, with the former more suitable to boost secondary metabolite's portfolio. Third (PC3, 6.7% variance) and forth (PC4, 10.1% variance) components are less clearly defined. PC3 contrasts a lower accumulation of flavonol glycosides with positive attributes of cluster morphology, such as rachis and wing length, which are predominant in PB manual removal, due to the severe defoliation early in the development of the cluster (12, 15) impacting photosynthate portioning in the shoot.

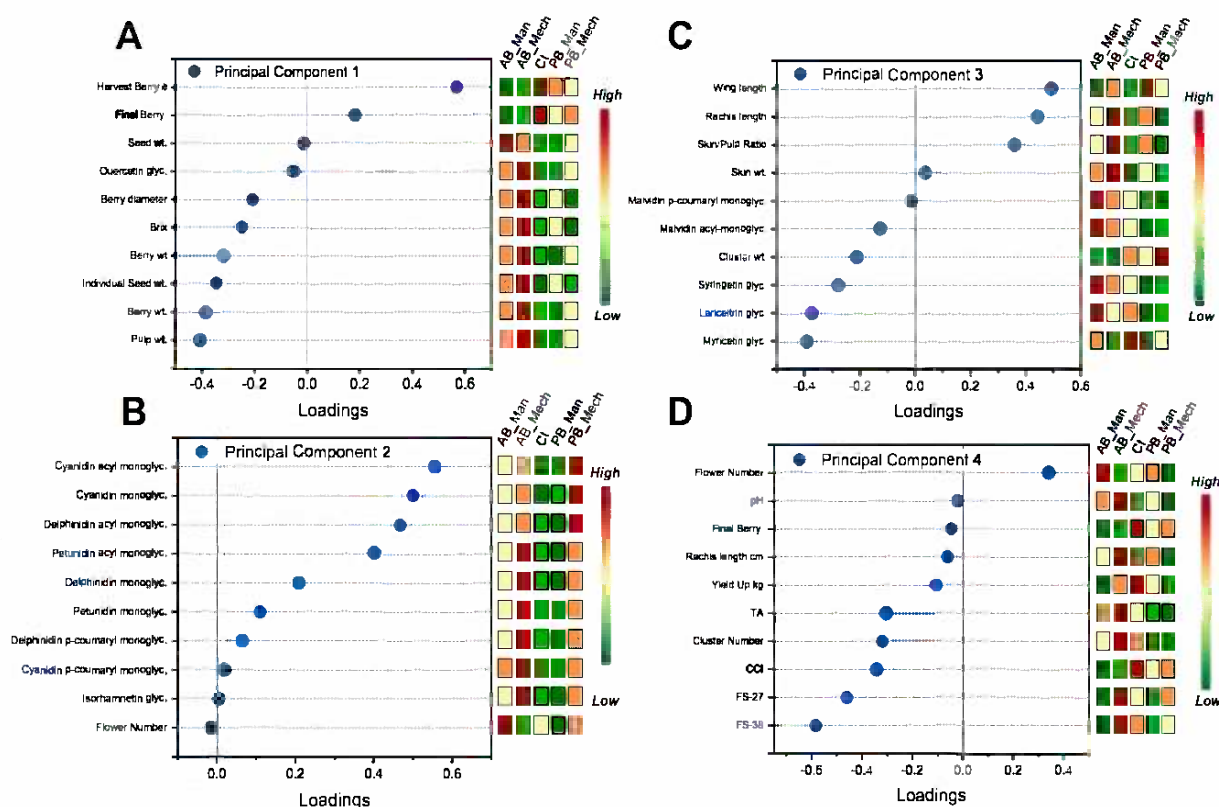


Figure 5. Plots A, B, C and D shown component 1, 2, 3, and 4 respectively, and the variable selected by the sPLS-DA model for a given component. The variables are ranked by the value of their loadings, and the impact of individual variables in each treatment is color-coded in the right panels from low-to-high in absolute values (mean by 2 years).

Flavonols (syringetin, laricitrin, and myricetin), as well as malvidin acyl- and *p*-coumaroyl derivatives, represent the downstream products of anthocyanin pathway (12, 32, 33) and confer the characteristic pigmentation to the berries of red grape cultivars. Finally, PC4 contrasts negative development of cluster's architecture (cluster number per vine, cluster compactness or CCI) and fruit set (FS-27, FS-38), calculated as number of berries per cluster at fruit set (FS-27) and at harvest (FS-38) over the number of flowers at bloom. The 3D score plot and loadings visualization allow an easier and global data interpretation in comparison to the univariate statistics approach. Treatments having higher score in PC1 and PC3 resulted in both reduced berry attributes (i.e. weight, diameter) and reduced flavonols accumulation, probably due to the treatment at the after-bloom (AB) stage: this is the case of both manual and mechanical AB leaf removal treatments, as well as control. When comparing mechanical vs. manual techniques, mechanical leaf removal had a greater PC2 score at both physiological stages, indicative of higher anthocyanins accumulation from veraison to harvest. This is consistent with other reports in Sangiovese (23) and Cabernet Sauvignon (34). These findings confirm that mechanical leaf removal enhance anthocyanins accumulation, opening the door to "precision agriculture enhancement" of specific anthocyanins with the intervention either at pre-bloom (i.e. delphinidin, cyanidin, petunidin) or after-bloom (i.e. delphinidin, petunidin).

5. Conclusions

Initial challenges encountered in interpreting fruit metabolic data and vine response parameters as affected by different leaf removal techniques were resolved by applying MANOVA and sPLS-DA. From MANOVA, we deduced that pre-bloom leaf removal had a more consistent impact on phenolics than sugars between years, suggesting that sugar accumulation variation is more related to the

seasonal variability of GGD accumulation and daily mean temperature. sPLS-DA enlightened the ability of pre-bloom leaf removal to boost fruit technological maturity and phenolic ripening in red (*Vitis vinifera* L.) Merlot grown in a cool climate. Thus, this technique is recommended for cool climates viticulture as an effective strategy to consistently enhance fruit quality.

Acknowledgments: This work was financially supported by AgBioResearch at Michigan State University (Project GREEN RE100290 and GR17041), the USDA Hatch project MICL02526, and the Michigan Grape and Wine Industry Council. Cristobal Torres-Palacios was supported by CONACYT Fellowship 711082. We appreciate the assistance of Jeff Lemon and Tim Lemon (Lemon Creek Winery, Berrien Springs, MI) for hosting the trial and for their help in vineyard maintenance and their assistance in performing the mechanized experimental treatments. Authors thanks Dr. Carlo Barnaba (MSU) for his critical review of the manuscript and Tony Schillmiller from the Mass Spectrometry and Metabolomics Core Center at MSU for his technical guidance with the MS identification. The authors declare no conflicts of interest.

References

1. Sumner LW, Mendes P, Dixon RA. Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry*. 2003;62(6):817-36.
2. Serrano A, Espinoza C, Armijo G, Inostroza-Blancheteau C, Poblete E, Meyer-Regueiro C, et al. Omics Approaches for Understanding Grapevine Berry Development: Regulatory Networks Associated with Endogenous Processes and Environmental Responses. *Frontiers in Plant Science*. 2017;8.
3. Giacometti J, Tomljanović AB, Josić D. Application of proteomics and metabolomics for investigation of food toxins. *Food Research International*. 2013;54(1):1042-51.
4. Ren S, Hinzman AA, Kang EL, Szczesniak RD, Lu LJ. Computational and statistical analysis of metabolomics data. *Metabolomics*. 2015;11(6):1492-513.
5. Craig A, Cloareo O, Holmes E, Nicholson JK, Lindon JC. Scaling and normalization effects in NMR spectroscopic metabonomic data sets. *Analytical Chemistry*. 2006;78(7):2262-7.
6. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 2003;19(2):185-93.
7. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in H-1 NMR metabonomics. *Analytical Chemistry*. 2006;78(13):4281-90.
8. Cadima J, Jolliffe IT. Loading and correlations in the interpretation of principle components. *Journal of Applied Statistics*. 1995;22(2):203-14.
9. Cerretani L, Maggio RM, Barnaba C, Toschi TG, Chiavaro E. Application of partial least square regression to differential scanning calorimetry data for fatty acid quantitation in olive oil. *Food Chemistry*. 2011;127(4):1899-904.
10. Granato D, Putnik P, Kovačević DB, Santos JS, Calado V, Rocha RS, et al. Trends in chemometrics: Food authentication, microbiology, and effects of processing. *Comprehensive Reviews in Food Science and Food Safety*. 2018;17(3):663-77.
11. Cuadros-Rodríguez L, Ruiz-Samblás C, Valverde-Som L, Pérez-Castaño E, González-Casado A. Chromatographic fingerprinting: An innovative approach for food identification and food authentication—A tutorial. *Analytica chimica acta*. 2016;909:9-23.

12. VanderWeide J, Medina-Meza IG, Frioni T, Sivilotti P, Falchi R, Sabbatini P. Enhancement of Fruit Technological Maturity and Alteration of the Flavonoid Metabolomic Profile in Merlot (*Vitis vinifera* L.) by Early Mechanical Leaf Removal. *Journal of agricultural and food chemistry*. 2018;66(37):9839-49.
13. Fabres PJ, Collins C, Cavagnaro TR, Rodríguez López CM. A concise review on multi-omics data integration for terroir analysis in *Vitis vinifera*. *Frontiers in plant science*. 2017;8:1065.
14. Lê Cao K-A, Boitard S, Besse P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC bioinformatics*. 2011;12(1):253.
15. Acimovic D, Tozzini L, Green A, Sivilotti P, Sabbatini P. Identification of a defoliation severity threshold for changing fruitset, bunch morphology and fruit composition in Pinot Noir. *Australian journal of grape and wine research*. 2016;22(3):399-408.
16. Ainsworth EA, Gillespie KM. Estimation of total phenolic content and other oxidation substrates in plant tissues using Folin–Ciocalteu reagent. *Nature protocols*. 2007;2(4):875-7.
17. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis. *Metabolomics*. 2007;3(3):211-21.
18. Mikulic - Petkovsek M, Schmitzer V, Slatnar A, Stampar F, Veberic R. Composition of sugars, organic acids, and total phenolics in 25 wild or cultivated berry species. *Journal of food science*. 2012;77(10):C1064-C70.
19. Chong J, Xia J. MetaboAnalystR: an R package for flexible and reproducible analysis of metabolomics data. *Bioinformatics*. 2018;34(24):4313-4.
20. Lorber A, Wangen L, Kowalski B. A theoretical foundation for the PLS algorithm. 1987.
21. de la Luz Cádiz-Gurrea M, Micol V, Joven J, Segura-Carretero A, Fernández-Arroyo S. Different behavior of polyphenols in energy metabolism of lipopolysaccharide-stimulated cells. *Food Research International*. 2019;118:96-100.
22. Poni S, Casalini L, Bernizzoni F, Civardi S, Intrieri C. Effects of early defoliation on shoot photosynthesis, yield components, and grape composition. *American Journal of Enology and Viticulture*. 2006;57(4):397-407.
23. Intrieri C, Filippetti I, Allegro G, Centinari M, Poni S. Early defoliation (hand vs mechanical) for improved crop control and grape composition in Sangiovese (*Vitis vinifera* L.). *Australian Journal of Grape and Wine Research*. 2008;14(1):25-32.
24. Zenoni S, Dal Santo S, Tornielli GB, D'Inca E, Filippetti I, Pastore C, et al. Transcriptional Responses to Pre-flowering Leaf Defoliation in Grapevine Berry from Different Growing Sites, Years, and Genotypes. *Frontiers in Plant Science*. 2017;8.
25. Schultze SR, Sabbatini P, Luo LF. Interannual Effects of Early Season Growing Degree Day Accumulation and Frost in the Cool Climate Viticulture of Michigan. *Annals of the American Association of Geographers*. 2016;106(5):975-89.
26. Cohen SD, Tarara JM, Gambetta GA, Matthews MA, Kennedy JA. Impact of diurnal temperature variation on grape berry development, proanthocyanidin accumulation, and the expression of flavonoid pathway genes. *Journal of Experimental Botany*. 2012;63(7):2655-65.
27. Schultze SR, Sabbatini P, Andresen JA. Spatial and Temporal Study of Climatic Variability on Grape Production in Southwestern Michigan. *American Journal of Enology and Viticulture*. 2014;65(2):179-88.
28. Zapata D, Salazar-Gutierrez M, Chaves B, Keller M, Hoogenboom G. Predicting key phenological stages for 17 grapevine cultivars (*Vitis vinifera* L.). *American Journal of Enology and Viticulture*. 2017;68(1):60-72.
29. Medina-Meza IG, Aluwi NA, Saunders SR, Ganjyal GM. GC–MS profiling of triterpenoid saponins from 28 quinoa varieties (*Chenopodium quinoa* Willd.) grown in Washington State. *Journal of agricultural and food chemistry*. 2016;64(45):8583-91.

30. Chen Y-S, Aluwi NA, Saunders SR, Ganjyal GM, Medina-Meza IG. Metabolic fingerprinting unveils quinoa oil as a source of bioactive phytochemicals. *Food Chemistry*. 2019.
31. Yao F, Coquery J, Lê Cao K-A. Independent principal component analysis for biologically meaningful dimension reduction of large biological data sets. *BMC bioinformatics*. 2012;13(1):24.
32. de Orduna RM. Climate change associated effects on grape and wine quality and production. *Food Research International*. 2010;43(7):1844-55.
33. De Lorenzis G, Rustioni L, Parisi SG, Zoli F, Brancadoro L. Anthocyanin biosynthesis during berry development in corvina grape. *Scientia Horticulturae*. 2016;212:74-80.
34. VanderWeide J, Tombesi S, Castellarin SD, Sabbatini P. Canopy architecture and fruit microclimate, not ripening-related phytohormones, control phenylpropanoid accumulation in response to early leaf removal in 'Merlot' (*Vitis vinifera* L.) grapevines. *Plant Physiology and Biochemistry*. 2020;157:291-302.

Supplemental information

Table S1. Mass and retention time of identified grape metabolites

Compound	ESI-mode	Monoisotopic Mass (g/mol)	Adduct Type	Retention Time (minutes)	Mass Error (ppm)
phenylalanine	positive	166.0865	M+H	3.04	-1.8
<i>Anthocyanins</i>					
cyanidin 3-glycoside	positive	449.1081	M+	3.56	-0.7
peonidin 3-glycoside	positive	463.1235	M+	3.83	-1.1
delphinidin 3-glycoside	positive	465.1032	M+	3.36	-0.2
petunidin 3-glycoside	positive	479.1185	M+	3.62	-1.0
malvidin 3-glycoside	positive	493.1341	M+	3.85	-1.0
cyanidin acetyl-3-glycoside	positive	491.1184	M+	4.17	-1.2
peonidin acetyl-3-glycoside	positive	505.1340	M+	4.43	-1.2
delphinidin acetyl-3-glycoside	positive	507.1135	M+	3.97	-0.8
petunidin acetyl-3-glycoside	positive	521.1289	M+	4.20	-1.2
malvidin acetyl-3-glycoside	positive	535.1445	M+	4.43	-1.3
cyanidin p-coumaroyl-3-glycosid	positive	595.1452	M+	4.68	0.0
peonidin p-coumaroyl-3-glycosid	positive	609.1603	M+	4.91	-0.8
delphinidin p-coumaroyl-3-glycosid	positive	611.1403	M+	4.50	0.3
petunidin p-coumaroyl-3-glycosid	positive	625.1553	M+	4.70	-0.6
malvidin p-coumaroyl-3-glycosid	positive	639.1708	M+	4.91	-0.9
<i>Flavonols</i>					
quercetin 3-glycoside	positive	465.1030	M+H	4.62	-0.6
isorhamnetin 3-glycoside	positive	479.1186	M+H	4.97	-0.8
myricetin 3-glycoside	positive	481.0975	M+H	4.30	-1.5
laricitrin 3-glycoside	positive	495.1125	M+H	4.63	-2.8
syringetin 3-glycoside	positive	509.1291	M+H	4.95	-0.8

433 **Table 2.** Merlot 3-O-glucoside Anthocyanin Quantitative (mg/kg DW) 2016

Treatment	cyanidin 3-glucoside	peonidin 3-glucoside	delphinidin 3-glucoside	petunidin 3-glucoside	malvidin 3-glucoside	Total 3-glucosides	Total* anthocyanins (mg/kg DW)
C	6.8 b	23.1 c	7.1 bc	8.9 bc	55.3	101.4 b	194.3 b
PB-MA	8.0 b	26.8 bc	5.9 c	7.5 c	46.7	95.0 b	180.2 b
PB-ME	15.7 a	40.7 a	11.8 a	13.1 ab	54.3	135.6 a	239.7 a
AB-MA	8.7 b	29.1 bc	6.8 bc	9.1 abc	58.5	112.2 ab	216.6 b
AB-ME	14.1 a	38.8 ab	11.0 ab	13.5 a	59.3	136.7 a	240.8 a

434

435 **Table S3.** Merlot 3-O-glucoside Anthocyanin Quantitative (mg/kg DW) 2017

Treatment	cyanidin 3-glucoside	peonidin 3-glucoside	delphinidin 3-glucoside	petunidin 3-glucoside	malvidin 3-glucoside	Total 3-glucosides	Total anthocyanins (mg/kg DW)
C	6.7 b	16.3 b	10.2	10.9	48.8	92.93	154.9
PB-MA	9.7 ab	22.0 ab	10.2	11.2	47.4	100.5	158.1
PB-ME	16.1 a	28.3 a	18.5	17.0	51.8	131.7	205.7
AB-MA	7.2 b	18.5 ab	10.4	11.7	57.0	104.8	176.6
AB-ME	8.0 b	16.8 b	11.8	12.0	47.6	96.2	154.8

436

437

438 **Table S4.** Merlot Acylated Anthocyanin Quantitative (mg/kg DW) 2016

Treatment	delphinidin 3-acyl-glucoside	cyanidin 3-acyl-glucoside	petunidin 3-acyl-glucoside	peonidin 3-acyl-glucoside	malvidin 3-acyl-glucoside	Total 3-acyl-glucosides
C	1.5 b	1.5 b	2.8 b	8.9 b	31.1	41.6
PB-MA	1.3 b	1.8 b	2.4 b	10.5 ab	25.6	45.8
PB-ME	3.0 a	3.8 a	4.9 a	14.9 a	29.8	51.4
AB-MA	1.5 b	1.9 b	3.0 b	11.3 ab	33.8	54.9
AB-ME	2.7 a	3.2 a	4.8 a	14.0 a	30.2	56.4

439 **Table S5.** Merlot Acylated Anthocyanin Quantitative (mg/kg DW) 2017

Treatment	cyanidin 3-acyl- glucoside	peonidin 3- acyl-glucoside	delphinidin 3-acyl- glucoside	petunidin 3-acyl- glucoside	malvidin 3-acyl- glucoside	Total 3- acyl- glucosides
C	1.2 b	5.2 b	2.1 b	3.2 b	22.5	34.3
PB-MA	1.7 b	6.3 ab	1.9 b	3.0 b	18.9	31.8
PB-ME	3.4 a	8.8 a	4.4 a	5.8 a	22.0	44.4
AB-MA	1.3 b	6.2 ab	2.0 b	3.3 b	25.8	38.6
AB-ME	1.4 b	5.0 b	2.5 ab	3.6 ab	19.0	31.6

440

441 **Table S6.** Merlot Coumarylated Anthocyanin Quantitative (mg/kg DW) 2016

Treatment	delphinidin 3- coumaryl- glucoside	cyanidin 3- coumaryl- glucoside	petunidin 3- coumaryl- glucoside	peonidin 3- coumaryl- glucoside	malvidin 3- coumaryl- glucoside	Total 3- coumaryl- glucosides
C	1.3 ab	2.2 c	2.8 ab	9.9 b	31.0	47.1
PB-MA	1.1 b	2.4 bc	2.4 b	10.6 ab	27.2	43.6
PB-ME	1.9 a	3.7 a	3.7 ab	14.3 ab	24.0	47.6
AB-MA	1.4 ab	2.7 abc	3.1 ab	12.0 ab	33.9	53.0
AB-ME	1.8 a	3.6 ab	4.0 a	14.5 a	25.8	49.7

442

443 **Table S7.** Merlot Coumarylated Anthocyanin Quantitative (mg/kg DW) 2017

Treatment	delphinidin 3-coumaryl- glucoside	cyanidin 3- coumaryl- glucoside	petunidin 3- coumaryl- glucoside	peonidin 3- coumaryl- glucoside	malvidin 3- coumaryl- glucoside	Total 3- coumaryl- glucosides
C	1.4	1.3 b	2.5	4.7	17.8	27.7
PB-MA	1.3	1.7 ab	2.3	5.4	15.2	25.7
PB-ME	2.1	2.5 a	3.2	6.5	15.2	29.6
AB-MA	1.5	1.5 ab	2.7	5.8	21.7	33.2

AB-ME	1.7	1.5 b	2.6	4.9	16.4	27.1
--------------	-----	-------	-----	-----	------	------

Table S8. Merlot Flavanol Quantitative (mg/kg) 2016

Treatment	quercetin 3-glucoside	syringetin 3-glucoside	laricitrin 3-glucoside	myricetin 3-glucoside	isorhamnetin 3-glucoside	Total 3-glucosides
C	6.04 ab	0.151	0.111	0.605	0.620 b	7.53 bc
PB-MA	5.58 b	0.116	0.082	0.382	0.473 b	6.64 c
PB-ME	8.23 a	0.157	0.098	0.581	1.12 a	10.2 a
AB-MA	7.60 ab	0.159	0.099	0.429	0.713 b	9.95 ab
AB-ME	7.95 ab	0.155	0.091	0.399	1.13 a	9.73 ab

Table S9. Merlot Flavanol Quantitative (mg/kg) 2017

Treatment	quercetin 3-glucoside	syringetin 3-glucoside	laricitrin 3-glucoside	myricetin 3-glucoside	isorhamnetin 3-glucoside	Total 3-glucosides
C	5.18	0.142 ab	0.101 a	0.801	0.600	6.82
PB-MA	5.43	0.118 ab	0.083 ab	0.559	0.5367	6.72
PB-ME	6.23	0.137 ab	0.084 ab	0.742	0.710	7.90
AB-MA	6.53	0.162 a	0.108 a	0.707	0.670	9.06
AB-ME	5.64	0.108 b	0.063 b	0.492	0.533	6.84