Differentiation of Geographical Origins for Cabernet Sauvignon Wines

A Nontarget Metabolomics Study Using UHPLC-Q-TOF/MS

Application Note

Food Testing

Authors

Nana Liang, Ying Liu, Linli Wang, Peiyue Wang, Jinhua Wang, and Shen Han
Beijing Entry-Exit Bureau of Inspection and Quarantine, Beijing, China

Meiling Lu and Shan Zhou
Agilent Technologies, Inc., China

Abstract

Food adulteration and mislabeling may pose potential health risks and trust issues to consumers, particularly for high-valued wine products. The current analytical methods and anticounterfeit labeling techniques are insufficient to determine the identity and point of origins for these products. This application note describes a metabolomic profiling method to trace the origins of wine products based on the work reported by Liang, *et al.* [1]. The reference Cabernet Sauvignon wines collected from five wineries (two from the US and three from China) were initially analyzed using an Agilent UHPLC-Q-TOF/MS platform under accurate TOF/MS scanning mode. The raw data obtained were subjected to data mining using find-by-molecular-feature extraction. The results were imported into Agilent Mass Profiler Professional (MPP) chemometric software to find the characteristic compounds among the groups. Principle components analysis and clustering analysis using the obtained differential compounds demonstrated an ability to separate the two groups of US wines from wines made in China. A partial least square differentiation analysis (PLSDA) model based on them can predict the wine groups with high accuracy. Using the customized polyphenol compounds database and library for wine and other available Agilent PCDLs, twenty-three compounds were tentatively identified, and most of them were endogenous metabolites in grapes, suggesting that the grape metabolites contribute to the major characteristics of wines from different points of origin. The work demonstrates that the metabolomic profiling approach by combining UHPLC-Q-TOF/MS with chemometric analysis is promising in tracing the geographical origins of grape wines.

Agilent Technologies

Introduction

Grape wine has been widely accepted across China. Unfortunately wine adulteration and mislabeling, which may pose potential health risks and trust issues to consumers [1], are often found in the market. To improve and guarantee the quality and safety of wines, it is essential to develop methodologies to monitor wine quality and authenticity, and protect the products with specific geographical origins.

The flavor of wines is mainly due to transportation of specific secondary metabolites in grapes, into the finished wine through the brewing process. Hence, the wine from one particular winery can carry some specific features from the grapes, due to the environmental variations including soil, climate conditions, and different brewing processes. Even for the same genotype of grape, the level of some specific secondary metabolites can vary significantly among the geographic regions, and render a particular taste in the wine products. Therefore, it is possible to differentiate the wines with various origins through the characteristic metabolites patterns. Conventional analytical methods are insufficient to seek such feature patterns and determine the points of origin. A metabolomic profiling approach, based on the feature patterns of the small molecules, is a favored method for wine authenticity and source tracing studies [2-6]. UHPLC hyphenated with high resolution Q-TOF/MS is one of the key metabolomic profiling platforms to obtain the comprehensive picture of the less volatile and heat vulnerable small molecules in samples. Further chemometric analysis facilitates data mining to seek the specific feature patterns potentially used as criteria for classification. Here, we are aiming at discovering the characteristic patterns to determine the geographical origins for Cabernet Sauvignon wines from the US and China using UHPLC-Q-TOF/MS technique together with chemometric analysis.

Experimental

Materials and reagents

Methanol, formic acid, and ammonium acetate were all LC/MS grade, and purchased from Merck, Tedia, and Thermo Fisher, respectively. Deionized water was freshly produced in the lab using a Milli-Q water purification system.

Reference samples collection and preparation

Reference Cabernet Sauvignon (CS) wine samples (total 113) were collected from the wineries directly, including V Sattui (VS) and Robert Mondav (RM) wineries in Napa Valley of US, and three wineries from Zhangjiakou (ZJK), Qinghuangdao (QHD), and Shangdong (SD) provinces of China. Two milliliters of each wine sample was transferred into 2-mL glass sample vials, centrifuged to remove any particulates, and then subjected to analysis by UHPLC-Q-TOF/MS.

Workflow for metabolomic profiling of wine

Raw data were acquired by UHPLC-Q-TOF/MS under the TOF scanning mode with the conditions shown in Table 1. The acquired data were first extracted using the find-by-molecular-feature algorithm (MFE) in Agilent MassHunter Qual. 6.0, and the results were exported as cef files. Alternatively, the total sets of data can be loaded into the Agilent MassHunter Profinder software (V.7.0) for recursive MFE extraction, and exported as cef files. The cef files with the extracted compounds information were then imported into Agilent Mass Profiler Professional (V.13.1.1) for retention time/mass calibration, peak alignment, data filtration, multivariate and univariate statistics, principal component analysis (PCA), and clustering analysis to find the compounds with significant changes among the five sample groups. Modeling, particularly partial least square differentiation analysis (PLSDA), was applied to build a model for predicting the origins of the wine samples. A laboratory-customized polyphenol compounds database and library for wine (wine PCDL) and the METLIN PCDL were applied for identification of the differential compounds.

Detailed LC/MS conditions

Table 1. Instrument Conditions

LC conditions	
Instrument	Agilent 1290 Infinity LC System with built-in degasser
Autosampler	Agilent 1290 Infinity Autosampler with temperature control
Column compartment	Agilent 1290 Infinity Thermostatted Column Compartment
Column	Agilent ZORBAX Eclipse Plus C18, 2.1 × 100 mm, 1.8 µm
Column temperature	40 °C
Mobile phase	A) Aqueous solution containing 5 mmol/L ammonium acetate and 0.1 % formic acid
	B) Methanol/water (95:5) containing 5 mmol/L ammonium acetate and 0.1 % formic acid
Flow rate	0.4 mL/min
Injection volume	2.0 µL
Post time	3 minutes
Gradient elution profile	0–1 minutes: 1 %B, 1–8 minutes: 1–15 %B, 8–15 minutes: 15–45 %B, 15–17 minutes: 45–90 %B, 17–20 minutes: 90 %B
ESI-MS/MS conditions	
Instrument	Agilent 6530 Q-TOF LC/MS system with Agilent dual Jet Stream electrospray ionization source
Ionization mode	Positive
Drying gas temperature	325 °C
Drying gas flow rate	11 L/min
Nebulizer gas pressure	35 psi
Sheath gas temperature	350 °C
Sheath gas flow rate	12 L/min
Capillary voltage	3,500 V
Nozzle voltage	500 V
Scanning mode	TOF scan and target MS/MS scan
Scanning rate	2 spec/sec (TOF scan) 3 spec/sec (target MS/MS)
Scan range	100–1,100 (MS)/50–1,100 (MS/MS)
Reference ions	121.0509/922.0098

Results and Discussion

Optimized UHPLC separation and accurate Q-TOF/MS Detection

Wine is a very complex matrix, containing thousands of compounds. Hence, an appropriate separation is required to elute as many compounds as possible with satisfactory sensitivity, within an acceptable time frame. The total ion chromatogram in Figure 1A shows that by cautiously adjusting the gradient elution using the selected C18 column and the mobile phase, a 20-minute gradient elution is capable of separating the major components in the wine samples. More than 3,000 compounds could be extracted from a single wine sample (Figure 1B) through the MFE algorithm, indicating a high separation efficiency. Thus, such a gradient elution can be applied for further wine sample profiling analysis.

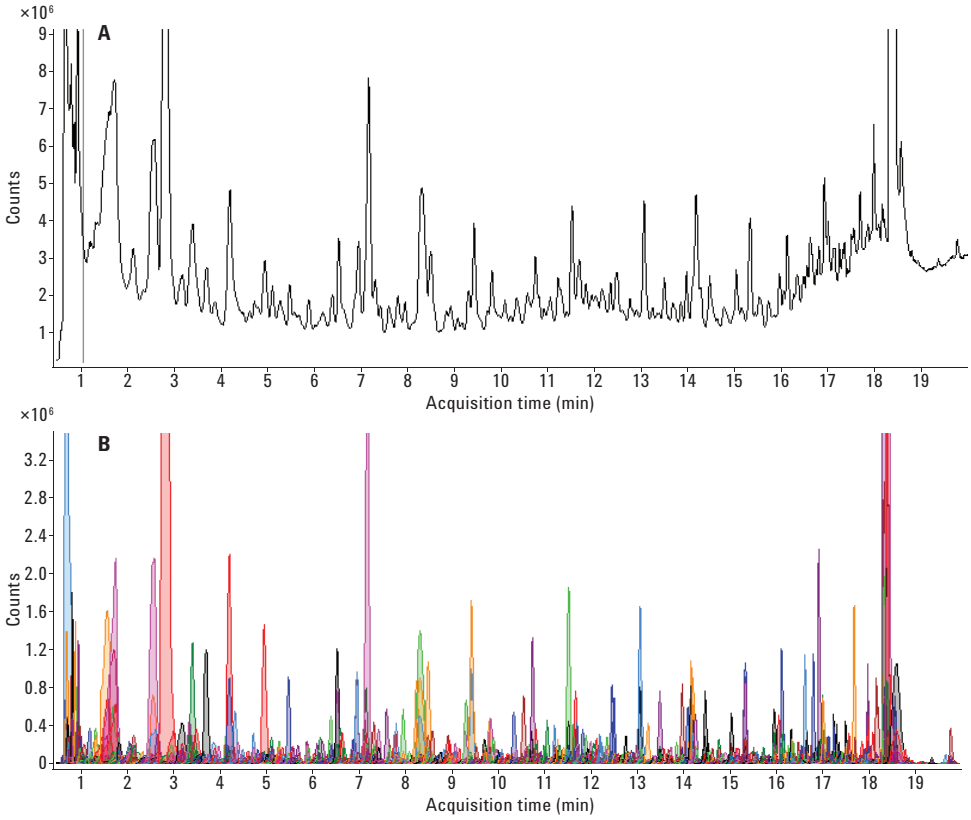

Figure 1. A typical total ion chromatogram (A) and the overlapped extracted compound chromatograms through MFE (B) demonstrating the high separation efficiency of the optimized LC/MS method. The sample is a CS wine from the US.

Chemometric analysis of the chemical profiles varying across wine groups

The data acquired through the Q-TOF scan mode were checked with caution on data reproducibility in both retention time (RT) and mass-to-charge ratio (*m/z*). It was found that both the relative deviations for RT and *m/z* ($|\Delta t_R|/t_{R,avg}$ and $|\Delta(m/z)|/(m/z)_{avg}$) were below 2 % and 5 ppm, respectively, indicating that the data were reliable for metabolomic profiling analysis. Qualified data were then extracted using MFE, and exported as cef files through batch processing software (DA reprocessor). To investigate how the compounds' abundance change among wine samples from different origins, the aforementioned processed data were imported into MPP for chemometric analysis. The RT and mass values for the compounds extracted through MFE from all collected wine samples were aligned in MPP in both the RT and the mass deviation windows. Each aligned compound was then treated as an individual entity in MPP, and annotated using both the RT and accurate mass. Through the above treatment, 29,564 entities in 98 wine samples were obtained. These entities were then sequentially filtered by occurrence frequency (≥80 % in at least one group), sample variability (≤50 % for at least one group), analysis of variance (ANOVA, $P \leq 0.01$), fold change ($FC \geq 3$), and C-C plot with $|P_{corrected}| \geq 0.8$. Eventually, 65 entities with significant differences among the five sample groups were kept. Figure 2 shows the variation of the average abundance for each entity among the groups.

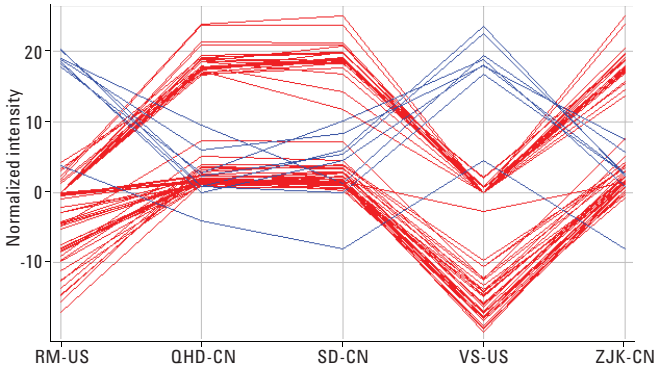

Figure 2. Abundance variation for the 65 differential entities among sample groups after data filtration. Filtration parameters: occurrence frequency, 80 % for ≥1 group; abundance variability, ≤50 % for ≥1 group; ANOVA, P≤ 0.01; fold change, FC≥3 with raw abundance difference cut-off at 1 × 10⁵ for seven out of 10 pairs; C-C plot, $|P_{corrected}| \geq 0.8$.

PCA is a mathematical algorithm used to view the structure of a complex data set; it is commonly used to view similarity among samples. The 65 entities obtained above were subjected to PCA. It was found that the data sets can be well reduced to four dimensions, with a sum of the highest three dimensions or principle components (PCs) explaining 84.76 % of the total covariances (Figure 3A). Even the highest two components can explain 83.04 % of the total covariances (Figure 3B). In addition, wines from two US wineries can clearly be separated from each other and China wines. In comparison, wines from three China wineries overlapped each other (Figure 3). It indicates that these 65 entities contributed to the major variation between China and US wines, and may act as markers for differentiation of the wines from the selected two US wineries and those from three regions of China.

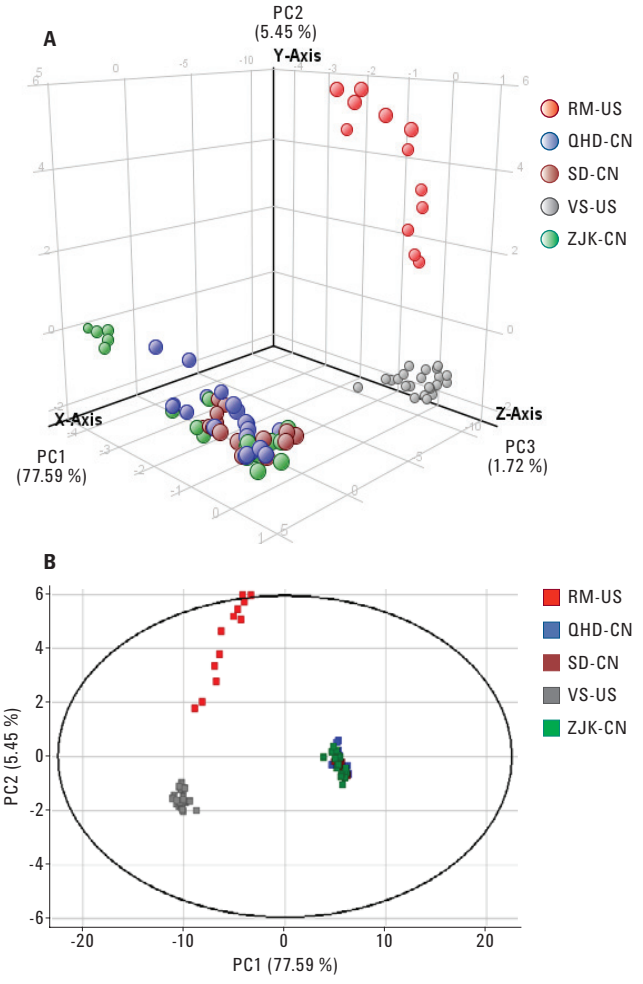

Figure 3. PCA score plot on 65 differential entities. A) 3-D PCA score plot. B) 2-D PCA score plot.

To better illustrate the abundance change of the selected entities across the wine groups, hierarchical cluster analysis (HCA) and K-means analysis were applied to cluster the data sets based on the similarity in entities' abundance. Using the 65 entities obtained through above filtration, a hierarchical condition tree for both wine groups and entities was obtained, as shown in Figure 4A. These wines can be classified mainly into three groups, as seen from the height of the nodes: RM-US, VS-US, and CN, which includes wines from three regions of China. In addition, the 65 entities can be classified mainly into three categories, C1 to C3, based on their similar patterns across wine groups. The entities in group C1 showed relatively lower abundance in US wines, but higher in CN wines. For group C2, the abundance in US wines are significantly higher than that in CN wines. Group C3 clearly shows lower intensity in US wines than in CN wines. The abundance change tendency for these entities are consistent with the K-means analysis shown in Figure 4B. While all three categories of entities can differentiate the three CN groups from the two US groups, some of C1 group can differentiate the two US wines from each other. In addition, fewer than 10 compounds showed visual differences within the three groups of China wines, but the differences were insufficient to differentiate the three groups of wine from each other.

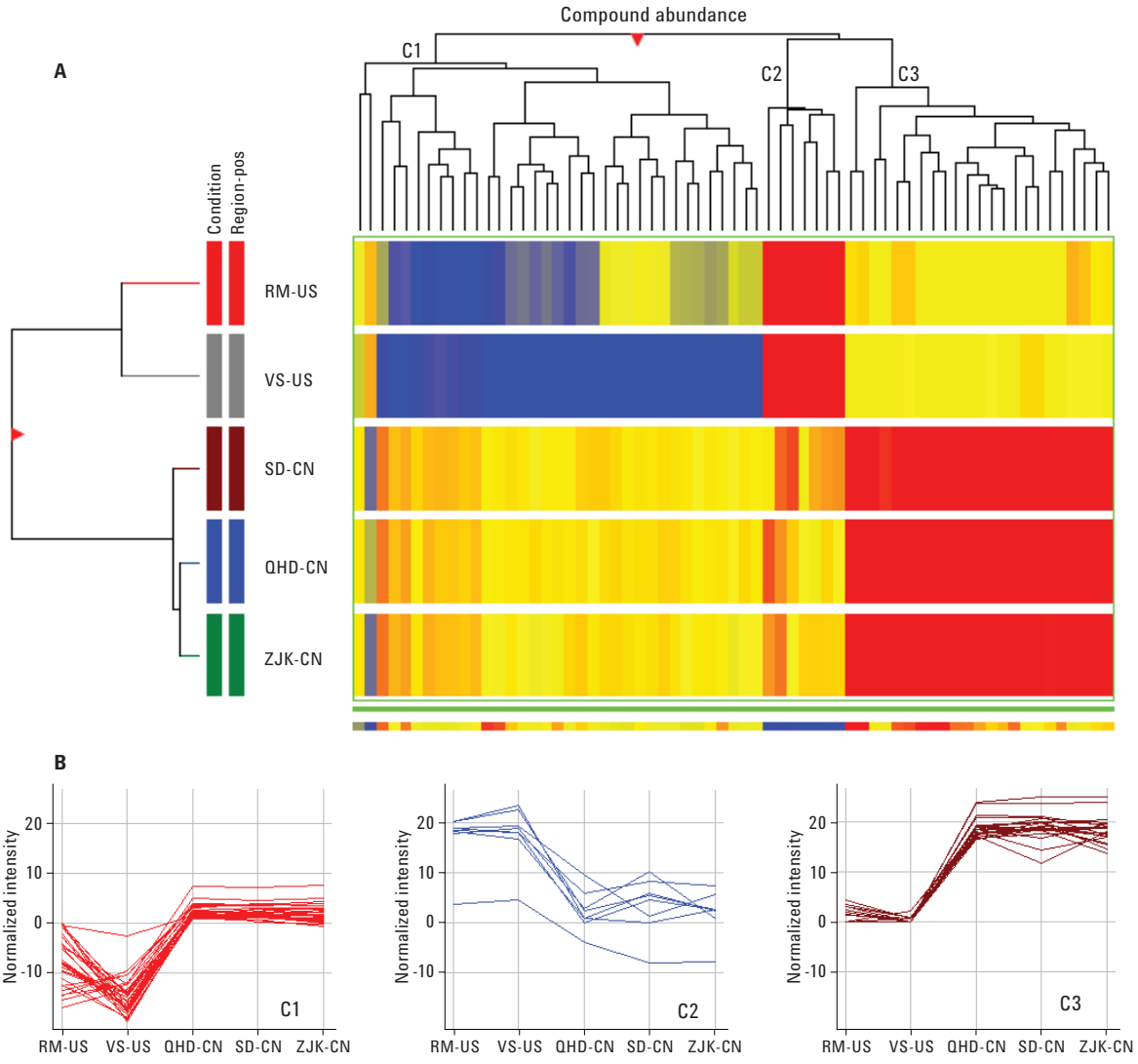

Figure 4. Hierarchical (A) and K-means (B) clustering analysis of the wines based on the 65 entities, showing clear differences between wines from US and China, and also within two groups of US wines.

Validation of the selected differential entities

To avoid false entity annotation during compound alignment, and false positive results during statistical analysis, the selected entities of the potential wine differential markers were validated by direct extraction from the acquired raw data. It was found that the most selected entities were true compounds, and exhibited obvious differences between wines from the selected US and China wineries. Figure 5 shows the representative compounds.

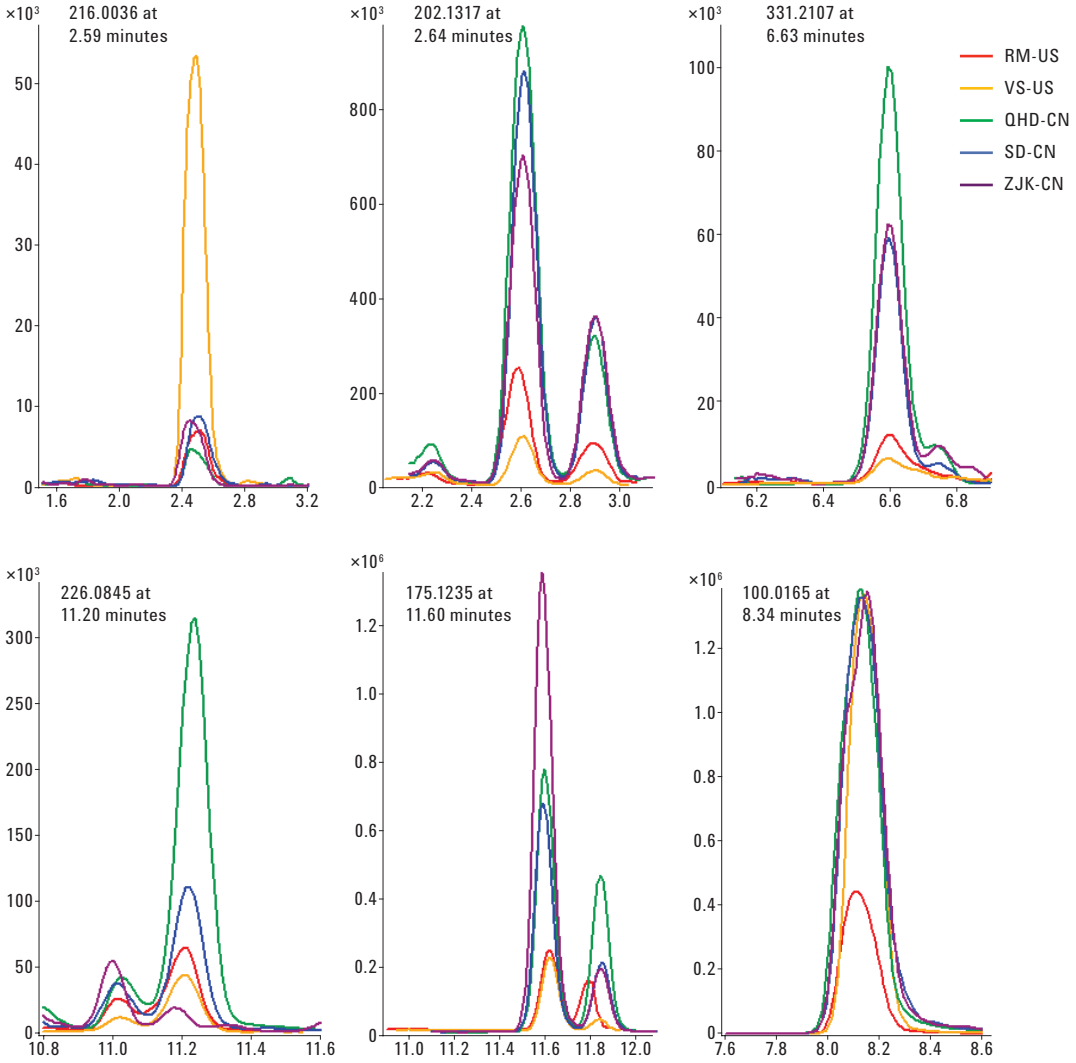

Figure 5. Extracted chromatograms demonstrating the clear variations in abundance for the selected compounds in different wines.

Model building for prediction

Based on the above selected compounds, a PLSDA model was created for predicting the origins of the wines. It was found that the model could predict the two groups of US wines correctly (Figure 6). Though it could not predict the subgroups of wines from China with high accuracy, it could predict them all as one group different from US groups. This shows that the selected markers can be used to differentiate wines from US and China wineries.

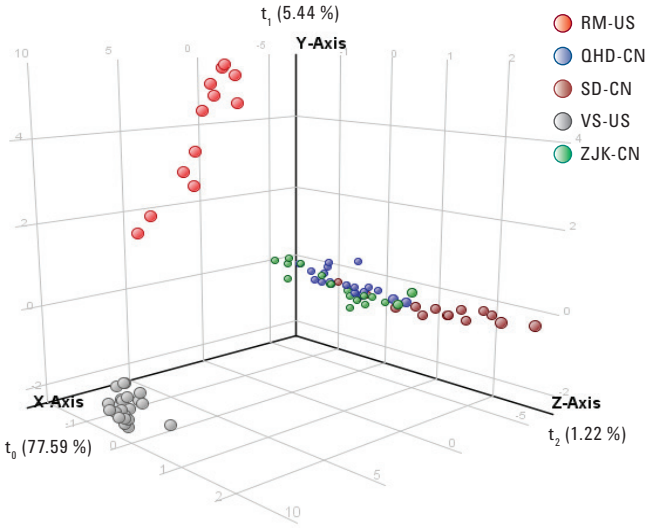

Figure 6. PLSDA score plot based on the 65 entities, showing clear separation of two groups of US wines from the wines of China.

Table 2. Cross-Validation Using the Built Model, Showing Accurate Prediction of Wines from the US and China

	RM-US	VS-US	QHD-CN	SD-CN	ZJK-CN	Accuracy (%)
(True) RM-US	12	0	0	0	0	100.0
(True) VS-US	0	23	0	0	0	100.0
(True) QHD-CN	0	0	18	2	1	85.71*
(True) SD-CN	0	0	3	12	0	80.00*
(True) ZJK-CN	0	0	4	2	13	68.42*
Overall accuracy						86.67*

* When treating the wines from three regions of China as one group, the prediction accuracy for this group of wines was 100 %.

Feature compound identification

To investigate which compounds contribute to the variation among the wines from two US wineries and the wineries in China, the selected markers were identified through an integrated ID Browser in MPP by combining a lab-customized wine PCDL with retention times obtained under the same experimental conditions. Two compounds were very likely to be procyanidin dimer B3 and epi-gallocatechin-epi-catechin dimer (EGCEC) isomers, with mass accuracy and isotopic patterns matching excellently with the theoretical values in the database (Figure 7). Further target MS/MS analysis showed characteristic fragment ions for the monomers of these two compounds. The two compounds were catechin dimers, commonly found in plants, with the former relating to the color of grape wine, and the latter possibly relating to the taste and flavor of the wine. Those compounds not identified using the customized PCDL were tentatively identified by matching the accurate mass and isotope pattern against available PCDLs such as Metlin. Another 21 compounds were tentatively identified (Table 3), and most of them were endogenous compounds from grapes, including polyphenols, esters, organic acids, and small peptides, which may also contribute to the characteristic flavor features of wine.

A

Best	Name	Formula	Mass	Mass (DB)	Diff (mD	Diff (abs. pp	RT	RT Diff	Diff (pp	Score
◉	Dimer (epi)gallocatechin-(epi)catechin-2	C30 H26 O13	594.1377	594.1373	-0.34	0.58	5.148	-0.055	-0.58	91.21

Species	m/z	Height	Score (MS)	Score (mass)	Score (iso. abund)	Score (iso. spacing)	Score (MFG)	Ion Formula
(M+H)+	595.14...	22867		96.68	99.59	99.53	87.42	

m/z	m/z (Calc)	Diff (ppm)	Diff (mDa)	Height	Height (Calc)	Height %	Height % (Calc)	Height Sum %	Height Sum% (Calc)
595.14...	595.1446	0.39	0.2	22840.8	22867	100	100	70	70.1
596.149	596.148	-1.71	-1	7477.8	7603.9	32.7	33.3	22.9	23.3
597.15...	597.1506	-4.98	-3	1924.3	1835	8.4	8	5.9	5.6
598.16...	598.1532	-13.36	-8	393.1	330	1.7	1.4	1.2	1

B

Best	Name	Formula	Mass	Mass (DB)	Diff (mDa)	Diff (abs. ppm)	RT	RT Diff	Diff (ppm)	Score
◉	Procyanidin dimer B3 Cat-(4α→8)-Cat	C30 H26 O12	578.1427	578.1424	-0.3	0.52	9.641	-0.018	-0.52	98.68

Species	m/z	Height	Score (MS)	Score (mass)	Score (iso. abund)	Score (iso. spacing)	Score (MFG)	Ion Formula
(M+Na)+	601.133	6589.5	66.92	75.35	60.09	58.25		
(M+H)+	579.14...	201777.7	99.21	99.82	99.62	97.51		

m/z	m/z (Calc)	Diff (ppm)	Diff (mDa)	Height	Height (Calc)	Height %	Height % (Calc)	Height Sum %	Height Sum% (Calc)
579.14...	579.1497	-0.06	0	201777.7	200522	100	100	70.7	70.2
580.15...	580.1531	-0.28	-0.2	65283.6	66603	32.4	33.2	22.9	23.3
581.15...	581.1557	-3.16	-1.8	15479.7	15653.8	7.7	7.8	5.4	5.5
582.16...	582.1584	-10.03	-5.8	2988.8	2750.8	1.5	1.4	1	1

Figure 7. The compounds match with standards in the customized wine PCDL. The red rectangles indicate how the mass and retention time are close to the theoretical value and the reference value, respectively.

Table 3. Summary of Tentatively Identified Differential Compounds in Wine

t_R/min	Exp. MW	Theor. MW	Mass accuracy/ppm	Formula	Identification
Endogenous compounds related to grapes (12)					
5.170	594.1374	594.1373	0.2	$C_{30}H_{26}O_{13}$	Dimer (epi) gallocatechin-(epi) catechin-2
6.157	620.1523	620.1530	-1.1	$C_{32}H_{28}O_{13}$	Peonidin-3-O-acetylglucoside-4-vinylphenol
9.648	578.1424	578.1424	0	$C_{30}H_{26}O_{13}$	Procyanidin dimer B3 Cat-(4α⟶8)-Cat
12.394	224.1410	224.1413	-1.3	$C_{13}H_{20}O$	Methyl (+)-7-isojasmonate
10.527	196.0738	196.0736	1.0	$C_{10}H_{12}O_4$	Ethyl vanillate
11.749	146.0373	146.0368	3.4	$C_9H_6O_2$	Coumarin
13.747	370.1057	370.1053	1.1	$C_{20}H_{18}O_7$	Sesamolin
4.641	226.0957	226.0954	1.3	$C_{10}H_{14}N_2O_4$	Porphobilinogen
5.322	360.1424	360.1420	1.1	$C_{16}H_{24}O_9$	7-Deoxyloganate
8.363	260.1167	260.1161	2.3	$C_{14}H_{16}N$	Nadoxolol
8.984	176.0687	176.0685	1.1	$C_7H_{12}O_5$	2-Propylmalic acid
6.470	130.0272	130.0266	2.6	$C_5H_6O_4$	Citraconic acid
Exogenous compounds related to grape cultivating and wine processing (1)					
9.125	212.0684	212.0685	0.5	$C_{10}H_{12}O_5$	Asperlin
Small peptides (10)					
4.430	232.1425	232.1423	-0.9	$C_{10}H_{20}N_2O_4$	Ile Thr
4.545	287.1843	287.1845	-0.3	$C_{13}H_{25}N_3O_4$	Leu Gly Val
4.797	317.1952	317.1951	0.3	$C_{14}H_{27}N_3O_5$	Val Leu Ser
5.189	228.1476	228.1474	0.9	$C_{11}H_{20}N_2O_3$	Pro Leu
5.271	260.1375	260.1372	1.2	$C_{11}H_{20}N_2O_5$	Glu Leu
5.681	446.1900	446.1914	-3.1	$C_{20}H_{26}N_6O_6$	Gly Phe Ser His
7.801	301.2001	301.2002	-0.3	$C_{14}H_{27}N_3O_4$	Ile Ala Val
8.736	386.2162	386.2165	-0.8	$C_{17}H_{30}N_4O_6$	Ile Thr Pro Gly
11.573	329.2313	329.2315	-0.6	$C_{16}H_{31}N_3O_4$	Ile Val Val
12.041	548.1867	549.1867	0	$C_{23}H_{28}N_6O_{10}$	His Asp Asp Tyr

Conclusion

UHPLC-Q-TOF/MS is a powerful technique for metabolomic profiling of wines from various sources. Agilent MPP software allows the user to align, filtrate, and cluster data rapidly and efficiently. Using the proposed metabolomic profiling approach, 65 differential wine markers were observed among wines from the US and China. Their relative abundance across the selected wine groups can be used to predict whether a wine sample is from RM or VS wineries of the US, or from the selected three regions of China. Among these compounds, twenty-three were tentatively identified, and some of them are probably related to the color or taste/flavor of the wines, suggesting that geographical variations contribute to the characteristic features of wines. Further confirmation of these markers is undergoing.

References

1. N. N. Liang, Y. Liu, B. L. Wang, *et al.* "Differentiation of grape wines from US and China using modeling building" *China Fermentation* **33(12)**, 23-28 (2014).
2. D. Serrano-Lourido, J. Saurina, S. Hernandez-Cassou, *et al.* "Classification and characterization of Spanish red wines according to their appellation of origin based on chromatographic profiles and chemometric data analysis" *Food Chem.* **135(3)**,1425-31 (2012).
3. S. A. Bellomarino, X. A. Conlan, R. M. Parker, *et al.* "Geographical classification of some Australian wines by discriminant analysis using HPLC with UV and chemiluminescence detection" *Talanta* **80(2)**, 833-838 (2009).
4. S. Kallithraka, I. S. Arvanitoyannis, P. Kefalas, *et al.* "Instrumental and sensory analysis of Greek wines; implementation of principal component analysis (PCA) for classification according to geographical origin" *Food Chem*. **73(4)**,501-504 (2001).
5. L. Vaclavik, O. Lacina, J. Hajslova, J. Zeigenbaum. The use of high performance liquid chromatography–quadrupole time-of-flight mass spectrometry coupled to advanced data mining and chemometric tools for discrimination and classification of red wines according to their variety. *Anal. Chim. Acta*. **45**, 685 (2011).
6. A. Cuadros-Inostroza, P. Giavalisco, J. Hummel, *et al.* Discrimination of wine attributes by metabolome analysis. *Anal. Chem.* **82**, 3573–3580 (2010).

For More Information

These data represent typical results. For more information on our products and services, visit our Web site at www.agilent.com/chem.

www.agilent.com/chem

Printed in the USA
August 22, 2016
5991-7238EN

Agilent Technologies