https://doi.org/10.17221/82/2019-CJFS

Metabolomics-based authentication of wines according to grape variety

*Leos Uttl, Kamila Hurkova, Vladimir Kocourek, Jana Pulkrabova, Monika Tomaniova, Jana Hajslova**

Department of Food Analysis and Nutrition, Faculty of Food and Biochemical Technology, University of Chemistry and Technology Prague, Prague, Czech Republic

**Corresponding author: jana.hajslova@vscht.cz*

Citation: Uttl L., Hurkova K., Kocourek V., Pulkrabova J., Tomaniova M., Hajslova J. (2019): Metabolomics-based authentication of wines according to grape variety. Czech J. Food Sci., 37: 239–245.

Abstract: In 2008, the European Commission highlighted the risk of wine mislabelling regarding the geographical origin and varietal identification. While analytical methods for the identification of wine by geographical origin exist, a reliable strategy for authentication of wine variety is still missing. Here, we investigate the suitability of the metabolomic fingerprinting of ethyl acetate wine extracts, using ultra-high-performance liquid chromatography coupled to high-resolution tandem mass spectrometry. In total, 43 white wine samples (three varieties) were analysed within our study. The generated data were processed by principal component analysis and then by partial least squares discriminant analysis. The resulting statistical models were validated and assessed according to their R^2 (cum) and Q^2 (cum) parameters. The most promising models were based on positive ionisation data, enabling successful classification of 92% of wine samples.

Keywords: authentication; chemometrics; metabolomics; U-HPLC-HRMS/MS; wine variety

Because of its high commercial value and the large volumes produced, wine is a commodity potentially subjected to fraud and mislabelling (Villano *et al.* 2017). Preventing this problem is of high concern in the European Union, which is the world's leading wine producer (European Commission 2019; available at https://ec.europa.eu/agriculture/wine/statistics_en). The price of wine is determined by its quality, which is, in turn, influenced by a number of factors, such as grape growing region, oenological practice, grape variety (or varieties), wine-making techniques, age, and year of vintage (Alanon *et al.* 2015). On this account, it is a legal requirement that wine labelling accurately reflects such information. For characterisation of basic wine quality parameters, such as content of alcohol, sugars, acids, etc., standardised methods have been in routine use. In addition, procedures enabling the detection of various fraudulent practices *e.g.* dilution with cheaper products such as fruit juices, or the addition of undeclared colourants, sweeteners or aroma (Geana *et al.* 2016), have become available in specialised laboratories.

In 2008, the European Commission drew attention to the growing risk of wine mislabelling, particularly regarding geographical origin, and grape variety. As a result, the demand for reliable analytical strategies applicable to the authentication of these wine parameters has become urgent (European Commission 2019; available at https://ec.europa.eu/jrc/en/research-topic/food-authenticity-and-quality).

As regards authentication of wine according to its geographical origin, standardised approaches based on stable isotope ratio analysis (Camin *et al.* 2015; Fan *et al.* 2018) and/or trace element measurement (Fan *et al.* 2018) are currently commonly employed. On the contrary, discrimination of wine according to

Supported by the Ministry of Agriculture of Czech Republic, Project No. QJ1530272, by the Operative Program, Grant No. CZ.2.16/3.1.00/21537 and CZ.2.16/3.1.00/24503, and by the Ministry of Education, Youth and Sport of the Czech Republic, Project No. 43760/2015.

https://doi.org/10.17221/82/2019-CJFS

the grape variety used for production still remains a challenging task, although a number of studies aimed at varietal authentication have been performed in the recent decade. In some of them, proton nuclear magnetic resonance was employed (Godelmann *et al.* 2013). However, chromatography coupled with mass spectrometry was the most common method used. While discrimination based on volatiles fingerprinting was reported in only a few studies (Ziolkowska *et al.* 2016), liquid chromatography coupled with high-resolution mass spectrometry (HPLC-HRMS) represented the dominating technique used for this purpose. The overview of wine sets tested by HPLC-HRMS and the outcome of chemometric analysis are presented in Table 1. As shown here, many different discriminatory markers have been identified and used to build statistical models for classification of wine categories differing in grape variety. It is worth noting that most of the studies were focused on red wines only. As mostly white wines are produced in the Czech Republic, we focused on this category. Riesling, Pinot Gris, and Roter Traminer representing the group of 10 most popular white wine varieties in the Czech Republic were involved in the authentication study.

MATERIAL AND METHODS

Chemicals and samples. HPLC grade methanol was purchased from Merck and HPLC grade ethyl acetate from Fluka Analytical. Deionised water was obtained from a Milli-Q purification system supplied by Merck. Mobile phase additives formic acid and ammonium formate (purity ≥ 98%) were purchased from Sigma-Aldrich.

Monovarietal commercial bottled white wines (in total 43) were purchased directly from winemakers. The samples represented wines of Czech geographic origin, produced by different winemakers, from different grape varieties: Riesling (n = 14), Pinot Gris (n = 17) and Roter Traminer (n = 12); and vintages (2013 to 2015, individual years were as much as equally represented in each variety group); thus, a very variable sample set was obtained.

Sample preparation. Four mililiters aliquot of the sample was transferred into a 15-ml plastic cuvette. In the next step, 4 ml of water acidified with formic acid (pH 2) and 4 ml of ethyl acetate were added, the cuvette was intensively shaken for 3 min and then centrifuged at 10 000 rpm for 5 min at 5°C. A 3-ml aliquot of the upper phase was recovered and evaporated to dryness under a gentle stream of nitrogen. The residue was reconstituted to a final volume of 500 µl with a mixture of methanol and water (50:50, v/v). In this way, the samples were six-time pre-concentrated. At the same time, a quality control sample (*QC*) was obtained by the mixing of equal volumes of all the tested samples. Until U-HPLC-HRMS/MS analysis, all extracts were stored in the freezer at –18°C.

Instrumental conditions. For non-target analysis, the Dionex UltiMate 3000 RS U-HPLC system (Thermo Fisher Scientific, USA) coupled to the quadrupole time-of-flight (QTOF) SCIEX TripleTOF® 6600 mass spectrometer (AB SCIEX, Canada) was used.

Chromatographic separation was performed using HSS T3 column (1.8-µm, 2.1 mm × 100 mm; Waters). The mobile phases consisted of (A) 5 mM ammonium formate and 0.1% formic acid in Milli-Q water and (B) 5 mM ammonium formate and 0.1% formic acid in methanol. For both polarities, the elution multistep gradient was used as follows: 0.0 min (95% A, flow 0.40 ml/min), from 0 to 1 min (95% A, flow 0.40 ml/min), from 1 to 11 min (0% A, flow 0.55 ml/min); from 11 to 12 min (0% A, flow 0.60 ml/min); from 12 to 12.1 min (95% A, flow 0.40 ml/min); from 12.1 to 14 min (95% A, flow 0.40 ml/min). The column was kept at a temperature of 45°C. The sample injection volume was set at 4 µl.

The TripleTOF instrument was equipped with a DuoSpray™ ion source. The instrument was operated either in positive (ESI+) or negative mode (ESI–); parameter settings used for the measurement were: capillary voltage, + 5.0 kV (ESI+) and –4.5 kV (ESI–); nebuliser pressure: 50 psi, drying gas pressure: 50 psi, curtain gas pressure: 35 psi, source temperature: 480°C; and declustering potential 80 V.

To collect MS and MS/MS spectra, full mass scan (TOF MS) and information-dependent acquisition (IDA) methods were simultaneously used. The TOF MS spectra were acquired in the *m/z* range 100–1200 at an acquisition rate of 1.5 spectra s^{-1} (periodic cycle time 0.65 s), and the product ion (*PI*) spectra were acquired in the *m/z* range 50–1200 at an acquisition rate of 1.5 spectra s^{-1} (periodic cycle time 0.65 s). The collision energy was of 35 V with the spread of ± 15 V. In order to achieve the highest mass accuracy throughout the measurement, an automatic calibration was regularly performed (every 10 samples) by the calibration delivery system (CDS, APCI calibration solution). In order to avoid

https://doi.org/10.17221/82/2019-CJFS

systematic bias due to analytical variation, the in-batch order of all samples analysed was random and the *QC* sample was analysed at regular intervals through the analysis (every 10 wine samples).

The instrument was controlled by Analyst 1.7.1 TF (AB SCIEX, Canada) and the qualitative analysis was performed using PeakView 2.2 software (AB SCIEX, Canada).

Repeatability assessment. Based on the repeated measurement (n = 10) of the *QC* sample, the repeatability was determined as a relative standard deviation (RSD) of the response for 10 ions (compounds) randomly selected throughout the whole range of retention times. The RSD values ranged between 2–10%.

Data processing and statistical analysis. For data processing, MarkerView 1.3.1 software (AB SCIEX) was used. As a starting point, two automated algorithms, the first one for peak finding and the second one for retention time (*RT*) and *m/z* alignment, were applied. Peak detection parameters were set to a minimum peak width of 0.02 Da, a noise threshold of 10 and subtraction multiple factor of 1.5. The parameters used for *RT* and *m/z* alignment were as follows: *RT* range 0–12; *m/z* 100–1200, *RT* tolerance 0.2 min; mass tolerance 0.03 Da. In this way, two separate data matrices (one for ESI+ and one for ESI–) consisting of lists of features were obtained. Subsequently, total area sum normalisation was performed for each sample and the data were pre-processed using Pareto scaling. Finally, to obtain the first overview of the data structure, principal component analysis (PCA) was performed.

Simultaneously, the data were exported into SIMCA 13.0 software (Umetrics), where PCA and a partial least squares discriminant analysis (PLS-DA) were performed. Prior to the actual PCA and PLS-DA, the data were log transformed (to lower the data skewness) and Pareto-scaled. The most prominent variables for the statistical model building were selected, based on the variable importance in projection (VIP) plot.

The chemometric models obtained were accompanied with the R^2 (cumulative) and Q^2 (cumulative) parameters, which were used to determine the validity of the models. R^2 (cum) indicates the variation described by all components in the model, and Q^2 (cum) is a measure of how accurately the model can predict class membership. Both parameters were calculated by a seven-round internal cross-validation of the data, using a default option of SIMCA. R^2 (cum) and Q^2 (cum) values higher than 0.5 indicate good quality of the model; values close to 1 then indicate an excellent PLS-DA model (Triba *et al.* 2015). For the validation of models, a permutation test with 100 permutations was used.

RESULTS AND DISCUSSION

In this study, the experience obtained in our previous research focused on the critical assessment of HRMS for food authentication was employed. Nevertheless, careful tuning of sample analysis had to be performed in the first phase. Much attention was also paid to the processing of the generated data, because it is a critical step in any fingerprinting strategy (Han *et al.* 2017). In the paragraphs below, the strategy employed in this pilot study is introduced.

U-HPLC-HRMS/MS metabolomics fingerprinting method. Like several other similar studies concerned with varietal origin-based wine authentication (Table 1), the reversed phase HPLC-QTOF-MS technique was used for non-target analysis of wine metabolome components. As shown in Table 1, most of the differentiating markers were secondary metabolites, such as polyphenols, which is not surprising – red wines are richer in these compounds because they are made using the entire grape (skin and seeds included), while white wines, on which this study is focused, are produced by using only the free-running grape juice. To pre-concentrate these important secondary metabolites, ethyl acetate extraction of acidified wine samples was performed. When comparing direct injection of wine with the analysis of organic extract, up to 20-times higher signal intensities of some phenolic compounds (e.g. catechin or fertaric acid) were obtained, probably not only due to the injection of higher matrix equivalent, but also to a less intensive matrix suppression, as most polar metabolites are transferred into the organic extract in a limited amount.

Figure 1 shows an example of the total ion chromatogram of a wine sample. As can be seen here, a number of compounds were detectable in both ionisation modes. Nevertheless, to assess the classification potential of these fingerprints, multivariate statistics were employed for data processing.

Chemometric analysis. Chemometric analysis of generated data is always a critical step in authentication based on non-target screening. Raw data files obtained by the analysis of 43 samples were

https://doi.org/10.17221/82/2019-CJFS

Table 1. Analytical approaches used for authentication of wine according to the origin

Sample preparation	Analytical technique	Origin-based discrimination	No. of varieties	Varieties	MS detection	Chemometric	Var. in classification	Metabolites	Parameters of classification models and main results	Source
Direct injection	RP-HPLC–(ESI+) – QTOF-MS	varietal	3	Cabernet Sauvignon, Merlot, Pinot Noir	Non-targeted	PLS-DA	3	25	ESI+, PLS-DA model, recognition ability = 100%, prediction ability = 95.6%	(Vaclavik *et al.* 2011)
Filtration	RP-U–HPLC–(ESI-) –QTOF-MS	varietal	3	Syrah, Merlot, Pinot Noir	Non-targeted	–	–	2 (phenols)	–	(Delcambre & Saucier 2013)
Direct injection/ethyl acetate extraction	RP-U-HPLC–(ESI+/-) –QTOF-MS, DART – QTOF-MS	varietal	9	Merlot, Pinot Noir, Tempranillo, Shiraz, Chardonnay Blanc, Riesling, Sauvignon Blanc, Silvaner	Non-targeted	OPLS-DA	3, 2	–	Direct injection – ESI+, Sauvignon Blanc, Riesling and Silvaner, $R^2 = 0.9$, $Q^2 = 0.84$, 100% correctly classified samples (U-HPLC –QTOF–MS, best model)	(Rubert *et al.* 2014)
Filtration	RP-HPLC–(ESI+) – QTOF–MS	varietal and geographical	8	Aspiran, Bonarda, Cabernet Sauvignon, Malbec, Merlot, Sangiovese, Syrah, Tempranillo	Non-targeted	D-UPLS	4	4 (anthocyanins)	100% correctly classified samples	(Pisano *et al.* 2015)
Filtration, centrifugation, dilution in water, derivatisation	RP-UPLC–(ESI+) – QqQ–MS	varietal	2	Merlot, Cabernet Sauvignon	Targeted	–	–	5 (amines and phenols)	–	(Malec *et al.* 2017)
Filtration	RP-HPLC-(ESI+/–) – QqQ-MS	varietal, geographical and year of vintage	8	Cabernet Sauvignon and 7 not specified	Targeted	OPLS-DA	2	50 (45 nonanthocyanins and 5 anthocyanins)	$R^2 = 0.69$, $Q^2 = 0.64$, correctly classified – Cabernet Sauvignon 93%, other varieties 85%	(Li *et al.* 2018)

Figure 1. Pinot Gris ESI+ (blue) and ESI– (pink) total ion chromatograms

processed by MarkerView software. After automated peak finding and retention time (*RT*) and *m/z* alignment, two separate data matrices were obtained. ESI+ matrix contained 2 096 features and ESI– contained 1 045 features. To remove signal redundancy, all isotopic peaks and the background peaks originated from blank samples were excluded. After this step, reduced data matrices contained 1729 features (ESI+) and 703 features (ESI–). Subsequently, to obtain a first overview of the data structure, PCA was performed (Figure 2A and B). Data points (samples) are coloured according to the variety of the respective wine samples. In both ionisation polarities, samples were significantly mixed (regarding wine variety). It was therefore necessary to proceed deeper within the data processing. For this purpose, data matrices were exported into the SIMCA program. The robustness of the analytical procedure was confirmed by the tight clustering of the QC samples (Figure 2A and B).

In the SIMCA program, both data matrices were log-transformed and Pareto-scaled. Pre-processed data were filtered using VIP plot. In the VIP plot, the variables are ordered according to their VIP scores. The higher the VIP score of a variable is, the higher the effect on class separation the variable has (Xi *et al.* 2014). In general, if the VIP score is higher than 1, the variable might be considered as important. In all the presented models, a high amount (hundreds) of variables reached a VIP score of 1 or higher. Nevertheless, statistical models with a high ratio of 'number of variables/number of samples' are more prone to be over-fitted (Xi *et al.* 2014). Therefore, the numbers of the variables in all our models were reduced (only variables with the highest VIP score were used). Therefore, the ratios of 'number

https://doi.org/10.17221/82/2019-CJFS

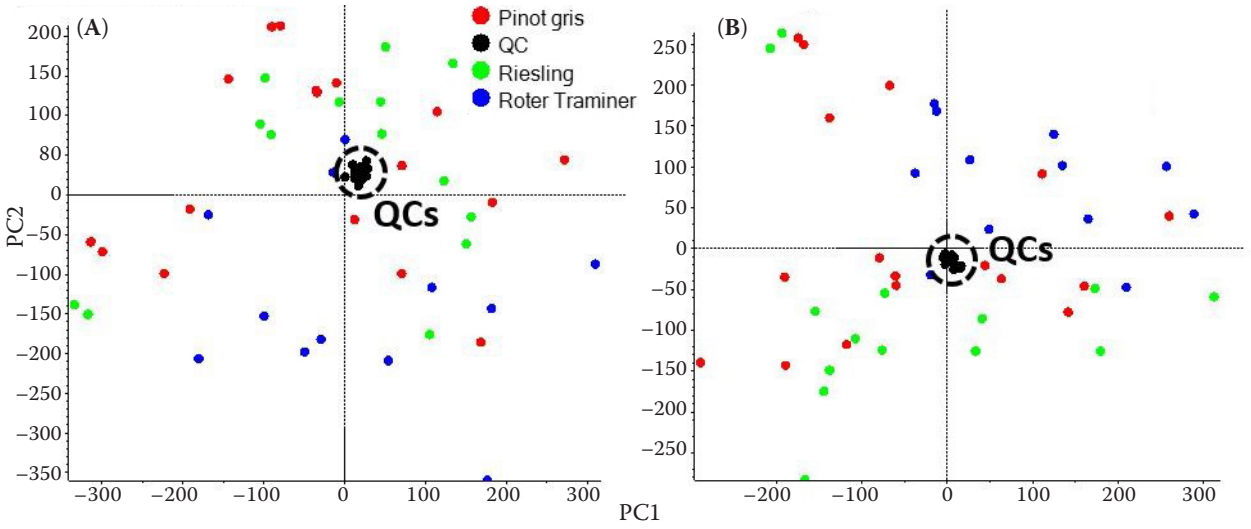

Figure 2. PCA score plots distribution of wine samples and tight clustering of QC sample repeated measurements (black) in ESI+ (**A**) and ESI– (**B**)
red – Pinot Gris; green – Riesling; blue – Roter Traminer

of variables/number of samples' were lower than 1. Selected variables were subsequently used to build PLS-DA models.

In the case of ESI+ data, the VIP threshold was set to 3 (considering the abovementioned criteria). Therefore, only variables with VIP scores higher than 3 were further processed. In this way, the number of features (variables) was reduced from 1 544 to 14. In the next step, the remaining variables were manually inspected, in order to select the most significant ones and to create an adequate statistical model. Out of the remaining 14 variables, 11 variables (with the highest VIP scores) were selected for the final statistical model. PCA and PLS-DA score plots for the final model are illustrated in Figure 3A and B. The performance of the final model was characterised by the following parameters: R^2 (cum) = 92%; Q^2 (cum) = 90%, which corresponds to a very good model. The model was validated by a permutation test with 100 permutations. The plot for the permutation test is presented in Figure 4A–C. The intercepts were as follows: Pinot Gris, R^2 = (0.0, 0.0483), Q^2 = (0.0, –0.286); Riesling, R^2 = (0.0, 0.0231), Q^2 = (0.0, –0.28); Roter Traminer, R^2 = (0.0, 0.0236), Q^2 = (0.0, –0.253).

In order to obtain models accompanied with even higher R^2 (cum) and Q^2 (cum) parameters (capable of better sample classification), statistical models comparing only two wine varieties (binary models) were also constructed. Their PLS-DA cross-validation results are summarised in Table 2. The same data handling procedure was applied for ESI– data. The results

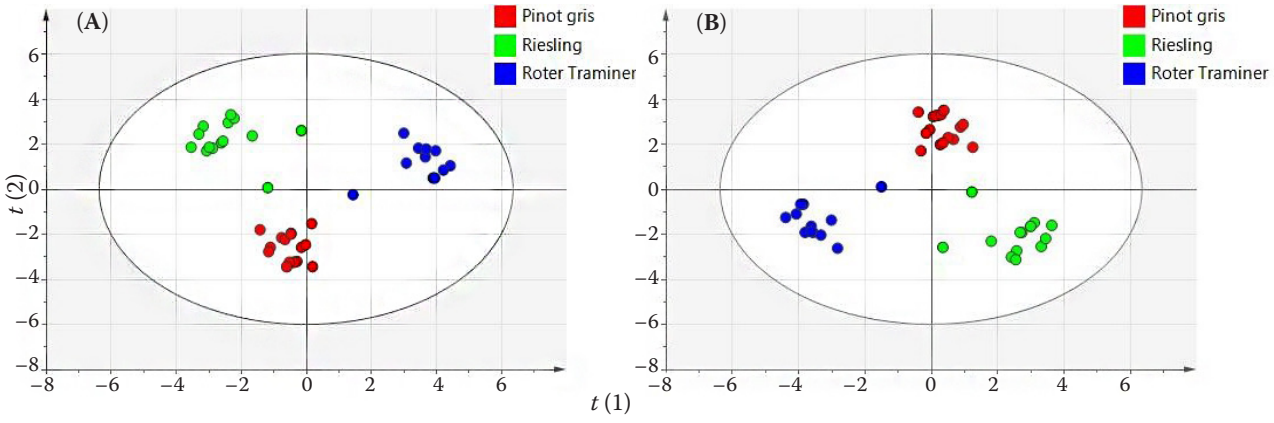

Figure 3. PCA (**A**) and PLS-DA (**B**) score plots showing classification of wine samples in ESI+
red – Pinot Gris; green – Riesling; blue – Roter Traminer

https://doi.org/10.17221/82/2019-CJFS

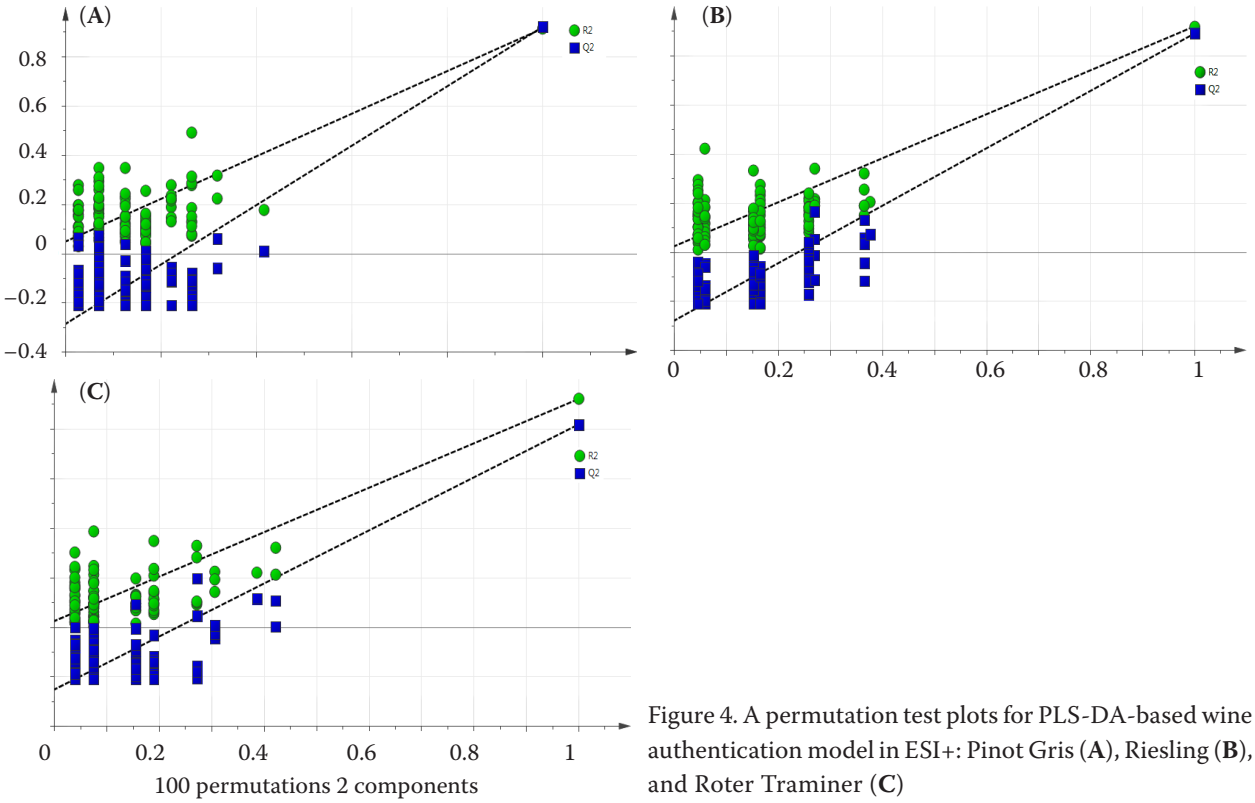

Figure 4. A permutation test plots for PLS-DA-based wine authentication model in ESI+: Pinot Gris (**A**), Riesling (**B**), and Roter Traminer (**C**)

of cross-validation of all negative ionisation-based models are also summarised in Table 2.

The most important markers for varietal classification were: Riesling – *m/z* = 327.0694, *RT* = 4.46 min, ESI+; *m/z* = 598.2102, *RT* = 6.06, ESI+, Roter Traminer – *m/z* = 485.1973, *RT* = 7.80 min, ESI+; *m/z* = 353.1554, *RT* = 8.02 min, ESI+, Pinot Gris – *m/z* = 449.1770, *RT* = 6.18, ESI+.

CONCLUSIONS

The presented pilot study performed on authentic white wines produced in the Czech Republic has demonstrated that the metabolomic fingerprinting procedure based on ethyl acetate sample extraction, followed with U-HPLC–HRMS/MS analysis and multivariate data processing (PCA and PLS-DA),

Table 2. Summarised cross-validation results of all PLS-DA statistical models

Number of groups	Ionisation mode	Number of variables	R^2 (cum)	Q^2 (%)	Varieties	Permutation test intercepts	
						R^2	Q^2
3	ESI+	11	92	90	Pinot Gris	0.0, 0.0483	0.0, –0.286
					Riesling	0.0, 0.0231	0.0, –0.28
					Roter Traminer	0.0, 0.0236	0.0, –0.253
	ESI–	15	87	84	Pinot Gris	0.0, 0.11	0.0, –0.265
					Riesling	0.0, 0.0773	0.0, –0.282
					Roter Traminer	0.0, 0.0917	0.0, –0.266
2	ESI+	12	96	92	Roter Traminer × Pinot Gris	0.0, 0.2	0.0, –0.193
	ESI+	10	95	91	Roter Traminer × Riesling	0.0, 0.204	0.0, –0.192
	ESI+	12	95	90	Riesling × Pinot Gris	0.0, 0.189	0.0, –0.211
	ESI–	14	93	87	Roter Traminer × Pinot Gris	0.0, 0.222	0.0, –0.278
	ESI–	13	95	87	Roter Traminer × Riesling	0.0, 0.218	0.0, –0.2
	ESI–	15	94	86	Riesling × Pinot Gris	0.0, 0.299	0.0, –0.206

https://doi.org/10.17221/82/2019-CJFS

is a promising tool for recognition of grape varieties used for wine production. Statistical models constructed using the data generated in ESI positive ionisation mode provided very good classification power; superior classification power being obtained for binary models.

To accommodate all variabilities and to obtain even more robust classification models, incorporation of wines from consecutive harvest years and different regions would be needed.

References

Alanon M.E., Perez-Coello M.S., Marina M.L. (2015): Wine science in the metabolomics era. Trac Trends in Analytical Chemistry, 74: 1–20.

Camin F., Dordevic N., Wehrens R., Neteler M., Delucchi L., Postma G., Buydens L. (2015): Climatic and geographical dependence of the H, C and O stable isotope ratios of Italian wine. Analytica Chimica Acta, 853: 384–390.

Delcambre A., Saucier C. (2013): High-Throughput OEnomics: Shotgun Polyphenomics of Wines. Analytical Chemistry, 85: 9736–9741.

Fan S., Zhong Q., Gao H., Wang D., Li G., Huang Z. (2018): Elemental profile and oxygen isotope ratio (δ18O) for verifying the geographical origin of Chinese wines. Journal of Food and Drug Analysis, 26: 1033–1044.

Geana E.I., Popescu R., Costinel D., Dinca O.R., Stefanescu I., Ionete R.E., Bala C. (2016): Verifying the red wines adulteration through isotopic and chromatographic investigations coupled with multivariate statistic interpretation of the data. Food Control, 62: 1–9.

Godelmann R., Fang F., Humpfer E., Schutz B., Bansbach M., Schafer H., Spraul M. (2013): Targeted and nontargeted wine analysis by 1H NMR spectroscopy combined with multivariate statistical analysis. Differentiation of important parameters: grape Variety, geographical grigin, year of vintage. Journal of Agricultural and Food Chemistry, 61: 5610–5619.

Han T.L., Yang Y., Zhang H., Law K.P. (2017): Analytical challenges of untargeted GC-MS-based metabolomics and the critical issues in selecting the data processing strategy. F1000Research, 6: 967–967.

Li S.Y., Zhu B.Q., Reeves M.J., Duan C.Q. (2018): Phenolic analysis and theoretic design for chinese commercial wines' authentication. Journal of Food Science, 83: 30–38.

Malec P.A., Oteri M., Inferrera V., Cacciola F., Mondello L., Kennedy R.T. (2017): Determination of amines and phenolic acids in wine with benzoyl chloride derivatization and liquid chromatography-mass spectrometry. Journal of Chromatography A, 1523: 248–256.

Pisano P.L., Silva M.F., Olivieri A.C. (2015): Anthocyanins as markers for the classification of Argentinean wines according to botanical and geographical origin. Chemometric modeling of liquid chromatography-mass spectrometry data. Food Chemistry, 175: 174–180.

Rubert J., Lacina O., Fauhl-Hassek C., Hajslova J. (2014): Metabolic fingerprinting based on high-resolution tandem mass spectrometry: a reliable tool for wine authentication? Analytical and Bioanalytical Chemistry, 406: 6791–6803.

Triba M.N., Le Moyec L., Amathieu R., Goossens C., Bouchemal N., Nahon P., Savarin P. (2015): PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters. Molecular Biosystems, 11: 13–19.

Vaclavik L., Lacina O., Hajslova J., Zweigenbaum J. (2011): 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. Analytica Chimica Acta, 685: 45–51.

Villano C., Lisanti M.T., Gambuti A., Vecchio R., Moio L., Frusciante L., Carputo D. (2017): Wine varietal authentication based on phenolics, volatiles and DNA markers: State of the art, perspectives and drawbacks. Food Control, 80: 1–10.

Xi B., Gu H., Baniasadi H., Raftery D. (2014): Statistical analysis and modeling of mass spectrometry-based metabolomics data. Methods in molecular biology (Clifton, N.J.), 1198: 333–353.

Ziolkowska A., Wasowicz E., Jelen H.H. (2016): Differentiation of wines according to grape variety and geographical origin based on volatiles profiling using SPME-MS and SPME-GC/MS methods. Food Chemistry, 213: 714–720.

Received: 2019–04–11
Accepted after corrections: 2019–05–22
Published online: 2019–07–24