# Current trends in ŒNO-NMR based metabolomics

José Enrique Herbert-Pucheta[1], Paz Austin-Quiñones[2], Francisco Rodríguez-González[3], Cristina Pino-Villar[4], Guadalupe Flores-Pérez [5], Santiago José Arguello-Campos[5], and VictorVillalobos Arámbula[5]

[1]Departamento de Química Orgánica, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional. Prolongación de Carpio y Plan de Ayala s/n, Colonia Santo Tomás, 11340 Ciudad de México
[2]Consejo Mexicano Vitivinícola A.C. Montecito No 38 Piso 15 Despacho 22 - WTC México, C.P. 03810, Ciudad de México
[3]Casa Madero, Carretera Paila 102- Parras Km. 18.2; Hacienda San Lorenzo. Parras de la Fuente, C.P. 27980, Coahuila, Mexico
[4]Bodegas de Santo Tomás, Km 49 Carretera Federal No. 1 Ensenada – La Paz, Ensenada Baja California, Mexico
[5]Secretaria de Agricultura y Desarrollo Rural, Municipio Libre No. 377, Colonia Santa Cruz Atoyac, CP 03310, Alcaldía Benito Juárez, Ciudad de México, Mexico

**Abstract.** Present work discusses strengths and limitations of two Nuclear Magnetic Resonance outliers obtained with a water-to-ethanol solvent multi pre saturation acquisition method, recently included in the Compendium of International Methods of Analysis of Wines and Musts, published as OIV-MA-AS316-01, and their accuracy for metabolomics analysis. Furthermore, it is also presented an alternative to produce more discriminant and sensitive NMR data matrices for metabolomics studies, comprising the use of a novel NMR acquisition strategy in wines, the double pulsed-field gradient echo (DPFGE) NMR scheme, with a refocusing band-selective uniform-response pure-phase selective pulse, for a selective excitation of the 5-10 ppm chemical shift range of wine samples, that reveals novel broad aromatic $^1$H resonances, directly associated to complex polyphenols. Both aromatics and full binnedOIV-MA-AS316-01,as well as the selective 5-10 ppm DPFGE NMR outliers were statistically analyzed with diverse non-supervised Principal Component Analysis (PCA) and supervised Partial Least Squares - Discriminant Analysis (PLS-DA), sparse (sPLS-DA) least squares- discriminant analysis, and orthogonal projections to latent structures discriminant analysis (OPLS-DA). Supervised multivariate statistical analysis of DPFGE and aromatics' binned OIV-MA-AS316-01NMR data have shown their robustness to broadly discriminate geographical origins and narrowly differentiate between different fermentation schemes of wines from identical variety and region.

## 1 Introduction

The oenological industry has benefited from the use of Proton Nuclear Magnetic Resonance ($^1$H-NMR) spectroscopy in combination with Multivariate Statistical Analysis (MSA) as a foodomics tool for retrieving discriminant features related to geographical origins, grape varieties, and further quality controls. Said *omics* methods have gained such attention that Intergovernmental Organizations and Control Agencies are currently recommending their massive use amongst Member States as quality compliances for tracking oenological standard and degradation parameters, fermentation products, polyphenols, amino acids, geographical origins, *appellations d'originecontrôlée* and type of monovarietal strains in wines [1].

Food integrity control for guaranteeing its *safety*, *quality*, *authenticity,* and *traceability* is a central matter for consumers' *protection*, *inspection* and *custom* agencies, nowadays. For that, worldwide control agencies constantly promote the use of analytical methods for detecting fraudulent practices such as counterfeited commodities, included oenological products. The alcoholic beverages are especially prone to simply being adulterated since they can be easily mixed with an important plethora of cheaper liquids [2]. Since the last years, standard analytical reference methods are transitioning towards *omics* techniques, combining either separation and/or spectroscopic schemes. However, most of them are in turn targeted screening strategies. More recently, due to the implementation of robust Artificial Intelligence (AI) algorithms in metabolomics [3], counterfeiting techniques are focused in recording unique sample fingerprints, with non-targeted metabolomics approaches. Such holistic methods related to food integrity control can produce *broad* or *narrow* analytical signatures, depending on the capacity of the technique to respectively produce low- or high-resolution fingerprints. Typically, broad analytical fingerprints are obtained with portable instrumentation such as UV-VIS, Raman or NIR approaches, whilst narrow fingerprints for elucidating complex chemical diversities are obtained by means of coupled chromatography with mass spectrometry, and more recently with nuclear magnetic resonance spectroscopy [4, 5]. In alcoholic beverages, *broad* and *narrow* analytical fingerprints have been used to respectively trace *brand* and *generic* authentication, like for instance, to differentiate rums manufactured from sugar cane juice from those made from sugar cane molasses [6], and for the assessment of both quality and authenticity of Scotch Whisky, based on non-targeted fingerprinting approach using gas chromatography coupled to tandem high-resolution mass spectrometry followed by multidimensional chemometric processing [7].

Despite the known restricted Limits of Quantification (LOQ) and Detection (LOD), proton Nuclear Magnetic Resonance Spectroscopy ($^1$H-NMR), in combination with diverse multivariate statistical analysis, has been an emergent *foodomics* approach to construct multiple food

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (http://creativecommons.org/licenses/by/4.0/).

matrices, with the known advantages of being nondestructive, easy sample manipulation, (ultra)-fast, reproducible and reliable, compared to chromatography coupled with MS techniques [8, 9].Said foodstuff metabolomics strategy has emerged over the last decades for the implementation of models to trace the food quality, origin, manufacture, and authenticity [5, 10].

Several reports have emerged over the last years in terms of both NMR acquisitions and MSA combined methodologies for developing different foodomics approaches. The first report applied water-to-ethanol NMR multi-presaturation schemes during mixing times and recovery delays within 1D-NOESY experiments in a set of approximately 600 German wine samples. This data matrix allowed classifying grape varieties, geographical origins, and aging of five wine-growing areas of Southern Germany (Rheinpfalz, Rheinhessen, Mosel, Baden, and Württemberg), with a principal component analysis (PCA), linear discrimination analysis (LDA), and multivariate analysis of variance (MANOVA) [11]. An Independent Component Analysis (ICA) combined with LDA achieved noticeable improvements to generate discriminative features within the NMR data matrix of German wine samples [12]. To discriminate between Italian "*Fiano di Avellino*" wines produced with the same grape variety but fermented with commercial or autochthonous yeast starters, the authors used a $T_1$-relaxation filter as a strategy for ethanol suppression, instead of water-to-ethanol multi-suppression, proton NMR profiling in combination with PCA, LDA, and a hierarchical cluster analysis (HCA) [13]. Recently, $^1$H-NMR targeted metabolomics has discriminated between Chinese wine regions [14] and grape varieties such as Cabernet Sauvignon, Merlot, and Cabernet Gernischt dry red wines [15], as well as different Chardonnay dry white wines treated with different inactive yeasts before aging [16]. Discriminative features to differentiate between grape varieties and fermentation processes were reported to be ethyl acetate, lactic acid, alanine, succinic acid, proline, malic acid, and gallic acid for red wines and 2,3-butanediol, ethyl acetate, malic acid, valine, succinic acid, lactic acid, tartaric acid, glycerol, gallic acid, choline, proline, and alanine for white wines. Furthermore, specific oenological improvements, such as the use of *Hanseniaspora vineae* yeast strains with respect standard fermentations to enhance aromatic profiles in Spanish Albillo white wines, were evaluated with both $^1$H-NMR and GC-FID-targeted metabolomics [17].

Present work stresses a set of NMR-MSA novelties for developing robust NMR *foodomics*. In one hand, it is evaluated the discriminant capacity of the NMR data matrix that is produced with the OIV Type 4 method reported in the Compendium of International Analysis of Methods No. OIV-MA-AS316-01 [1], that uses a multi-presaturation noesy pulse for suppressing water and ethanol signals during mixing and recovery delay times [1, 4, 5, 11, 18, 19]. After data processing, two signal bucketing strategies were tested for preparing the NMR inputs for MSA workflow: full $^1$H-NMRspectra and exclusively $^1$H- aromatic resonances from a chemical shift range between 5.5-10 ppm. These NMR data matrixes were submitted to diverse MSA strategies such as unsupervised Principal Component Analysis (PCA), as well as diverse supervised approaches such as standard partial (PLS-DA), sparse (sPLS-DA) least squares-discriminant analysis, and orthogonal projections to latent structures discriminant analysis (OPLS-DA).Accuracy of the OIV-MA-AS316-01 NMR data matrices to discriminate Mexican wines from three different geographical origins, each produced from diverse grape varieties and years of vintage [5],were evaluated by means of the quality of produced MSA holistic score plots. In parallel, a novel NMR outlier for "*wine-omics"*, produced with a Double Pulsed Field Gradient Echo (DPFGE) acquisition method[4], was also evaluated with PCA, PLS-DA, sPLS-DA, and OPLS-DA multivariate statistical analysis for their accuracy to differentiate between wines from the same geographical origin and variety, produced with different fermentation schemes [4].Pairwise comparisons between OPLS-DA score and loading plots produced with OIV-MA-AS316-01 and DPFGE NMR outliers, were carried out to demonstrate strengths and limitations of each data matrix herein analyzed, describe relevant discriminant factors such as geographical and fermentation processes' identities. Finally, a brief discussion stressing perspectives and challenges for NMR Big-Data acquisition, pre- and post-processing, multivariate statistical analysis, interpretation, management, and sharing, in relation to the digital transition of the vitiviniculture sector, stressed within the 2020-2024 OIV Strategic Digital Transformation Plan [20] is presented.

## 2 Materials and methods

### 2.1 Wine sampling

A set of 76 Mexican Merlot and Cabernet Sauvignonmonovarietal winesfrom 3 different regions:

- Queretaro (Finca Sala Vivé Freixenet, Ezequiel Montes, México, [20º39'55" N, 99º53'54" W], highlighted in blue within Figs. 1, 2, and 4);
- Baja California(Monte Xanic, Valle de Guadalupe, México, [32º17'34" N, 115º5'28" W], highlighted in red within Figs. 1, 2, and 4;
- Coahuila (Casa Madero, Parras, México, [25º27'2" N, 102º10'37" W], highlighted in black within Figs. 1, 2, and 4);

and from 2017-2020 year of vintages, were obtained for the present study. All samples from Querétaro (Merlot) and Baja California (Cabernet sauvignon) were fermented with standard inoculation schemes with the use of *Saccharomyces cerevisiae* yeast strains, respectively from Zymaflore®, Laffort, France, and D254™, Lallemand, Montreal, QC, Canada. Cabernet sauvignon grape varieties (*Vitisvinifera* L.) from Casa Madero, Parras, Coahuila were fermented with three different fermentation schemes [4]:

- Standard inoculation with *Saccharomyces cerevisiae* yeast strain, from D254™, Lallemand, Montreal, QC, Canada (green in Fig. 4).
- A first-step co-inoculation with non-*Saccharomyces Candida zemplinina* yeast strain (Enartis Ferm, San Martino Trecate, Italy), followed by a later inoculation with a D254™, *Saccharomyces cerevisiae* yeast strain (Lallemand, Montreal, QC, Canada; magenta in Fig. 4).
- Inoculation with a mixed strain containing *Saccharomyces uvarum* and *Saccharomyces cerevisiae*ex. ph. r. bayanus (hereinafter mentioned as *Saccharomyces bayanus ex uvarum*) yeast strain (Enartis Ferm, San Martino Trecate, Italy; yellow in Fig. 4).

### *2.1.1 NMR spectroscopy. Acquisition details*

For all batches, 540 µL of wines were dissolved in 60 µL of deuterium oxide solution, with 99.9% deuteration mixed with 0.05 wt% of 3-(trimethylsilyl) propionic-2,2,3,3-d4 acid and sodium salt as the internal reference (CAS No. 7789-20-0), and 0.1% of phosphonate $KH_2PO_4$ (CAS No. 7778-70-0) buffer was prepared and pH-adjusted to a value of 3.1.

All wine NMR spectra were recorded at 14.1 Teslas of static magnetic field on a Bruker 600 AVANCE III HD equipped with a 5-mm 1H/D BBO probe head with z-gradient. The following set of NMR experiments were conducted:

(a) Standard quantitative $^1$H-one dimensional direct polarization NMR experiments (q-$^1$H-NMR) were carried out for measuring the %alcohol content [4] by previously calibrating the 90º hard pulse (10.55 µs @ 23.69 kHz).By using 16 transients of 65,536 complex points, having recycling delays of 15 s and with acquisition times of 1723 milliseconds, have produced experimental times of 16 minutes and 43 seconds per spectrum. No apodization function was applied during Fourier-Transform.

(b) {$^1H_{water\_presat}$ NMR}(Figs. 3A, B, and C, bottom spectra): 1D single-pulse NOESY spectra with an off-resonance shaped-pulse water-to-ethanol multi-presaturation pulse during the relaxation delay (3 s), and mixing times (100 milliseconds), comprises the method OIV-MA-AS316-01, recently published at the Compendium of International Methods of Analysis of Wines and Musts of the International Organization of Vine and Wine[1, 4, 5, 11, 18, 19], herein after mentioned with this code. Power level irradiation amplitudes of $1.04 \times 10^{-3}$ W were used for solvent multi-suppression. The acquisitions were as follows: 64 transients were collected into 65,536 complex data points, with acquisition times of 1500 ms, produce experimental times of 4 min and 30 s per experiment.

(c) Selective Double Pulsed-Field-Gradient Echo (DPFGE) $^1$H NMR spectra (Figs. 3A, B, and C, top spectra) for the selective excitation of aromatic $^1$H spin systems (5.5-11 ppm, 3360 Hz). The acquisitions were as follows: from an optimized q-$^1$H NMR with a spectral width of 13 ppm and a transmitter frequency offset of 4.5 ppm, the pulse sequence schemed in Figure 2, top was implemented. The REBURP selective π refocusing band-selective uniform response pure-phase pulse that is flanked by two gradient pulses during an echo period that allows to exclusively refocus the selected aromatic chemical shift range whilst it is simultaneously defocusing the intense water-to-ethanol hydroalcoholic chemical shift range was calibrated with the aid of the programs Shape tool and NMRSIM (Bruker Biospin) in order to selectively excite a frequency range of 3360 Hz from a frequency offset of the REBURP pulse, defined at +2300 Hz with respect to the carrier frequency at 4.5 ppm that allowed exciting a chemical shift range between 5.5 and 11 ppm. The pulse length of the REBURP π pulse was defined at 1900 ms with a power level of 0.223 watts. The acquisitions were as follows: 64 transients were collected into 262,144 complex data points, with acquisition times of 3076 ms and recovery delays of 2 s, and produced experimental times per wine batch of 5 min and 24 s.

### *2.1.2 NMR spectroscopy. Processing details*

NMR postprocessing for producing the MSA input variables was carried out as follows: ppm calibration and manual phase corrections were conducted using Bruker Top Spin 4.0.8 software. Global and intermediate baseline corrections, least-squares or parametric time warping NMR alignments, variable size bucketing for untargeted profiling, and data matrix normalization were carried out with NMR Proc Flow software [21].

## 2.2 Multivariate statistical analysis

Data pre-processing comprising normalization by sum (for adjust differences amongst samples), transformation (Log) and autoscaling (mean centering divided by standard deviation of each variable), applied to remove any possible variation during experimental phase, in order to make features as comparable between them as possible, as well as statistical analysis workflow for obtaining the Principal Component (PCA); standard (PLS-DA), sparse partial least-square discriminant analysis (sPLS-DA), and Orthogonal Projections to Latent Structures Discriminant Analysis (OPLS-DA), from the constant sum normalized DPFGE and OIV-MA-AS316-01data matrixes, were developed with Metabo Analyst 5.0 software [22]. In all cases, T2 Hotelling's regions depicted by ellipses in score plots of each model define a 95% confidence interval [23]. Supervised OPLS-DA was carried out with Monte-Carlo cross-validations with 10 test partitions per 100 permutations for testing [24]. Reliability of each classification per supervised model (except for the reduced sPLS-DA model), was

evaluated in terms of goodness of the fit ($R^2$) and goodness of prediction ($Q^2$) [25-26].

## 3 Results

The robustness of the OIV-MA-AS316-01 data matrix for non-targeted NMR based metabolomics, is schemed in Figures 1 and 2. Recently included in the Compendium of International Methods of Analysis of Wines and Musts of the International Organization of Vine and Wine, and conceived as a Type IV methodfor targeted analysis of glucose, malic acid, acetic acid, fumaric acid, shikimic acid and sorbic acid in wine [1], present work evaluates two different data matrixes produced from OIV-MA-AS316-01 acquisition method, for non-targeted NMR based metabolomics analysis, in order to obtain models to discriminate between geographical origins. Figure 1 presents the score plots of unsupervised principal component analysis (PCA) and its supervised counterpart -Partial Least Square Discriminant Analysis (PLS-DA)- of two OIV-MA-AS316-01 data matrixes: full $^1$H chemical shift binning (right in Fig. 1) and a reduced binning of all aromatic $^1$H resonances (left in Fig. 1). As clearly observed, full binning produced a reduced discriminant capacity, with an unsupervised PCA, with poor separations amongst groups using two dimensional projections, with a total variance of 70.0% (PC1=58.1%, PC2= 11.9%). In striking contrast, with an equivalent PCA variance (69.7%, PC1= 49.9%, PC2= 19.8%), the OIV-MA-AS316-01 aromatic $^1$H binning NMR data matrix, produces a model to discriminate between wine samples from Baja California, Coahuila, and Queretaro, even since the unsupervised machine learning algorithm.

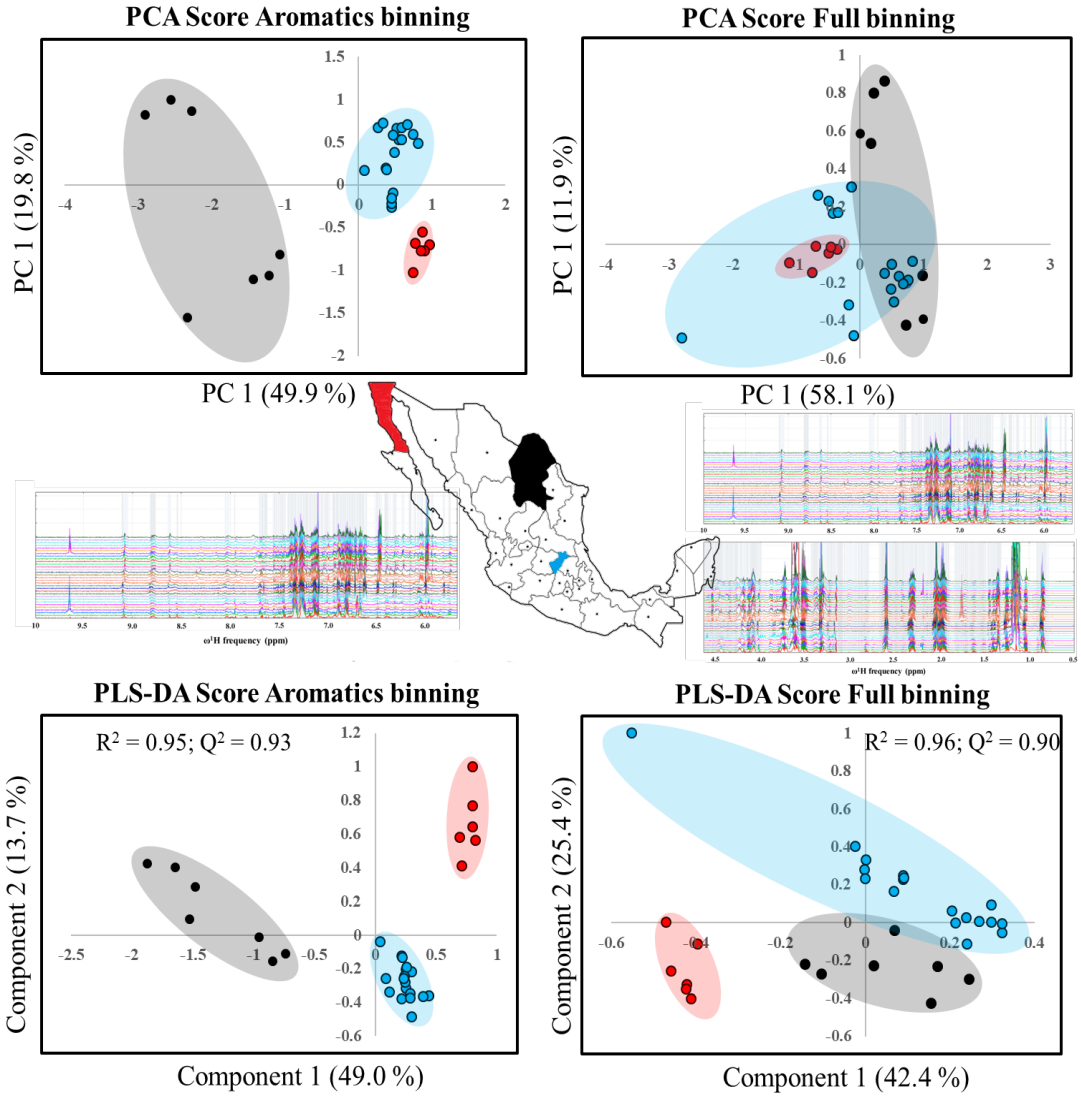



**Figure 1.** Unsupervised Principal Component Analysis (PCA, top) and supervised partial least square discriminant analysis (PLS-DA, bottom) score plots of two data matrixes obtained with the OIV-MA-AS316-01 acquisition method: full $^1$H-chemical shift binning (right) and signal bucketing of exclusively aromatic signals from 5.5-10 ppm of proton chemical shift range (left). Both signal bucketing strategies are highlighted in the middle of the figure. MSA score plots model discriminant fingerprints related to wines' geographical origin, assigning red T2 Hotelling's ellipse and scoresto Baja California, blackT2 Hotelling's ellipse and scores to Coahuila and blue T2 Hotelling's ellipse and scores to Queretaro samples, in agreement with country's colour map.

The last strongly suggest that wines' aromatic $^1$H NMR outliers are sufficiently discriminant towards key factors such as the geographical origins. Unsupervised principal component analysis is generally used for organizing the NMR data matrix and for determining correlations between discriminant factors (geographical origin, fermentation schemes) and outliers (OIV-MA-AS316-01 aromatic $^1$H binning/OIV-MA-AS316-01 full$^1$H binning), whereas an important number of metabolomics works claim the need to test the discriminant capacity of PCA models as a prerequisite to evaluate the quality of the MSA data inputs. Present work has evaluated diverse NMR outliers that can be obtained from the OIV-MA-AS316-01 (data not shown), concluding that the OIV-MA-AS316-01 aromatic $^1$H binning outlier (Fig. 1, top left)it the sole that can accurately discriminate geographical origins with an unexpensive PCA approach.

The last is confirmed with the supervised PLS-discriminant analysis, whereas the fully binned OIV-MA-AS316-01 matrix produces a discriminative holistic geographical origin fingerprint ($R^2$ = 0.96, $Q^2$ = 0.90), demonstrating the advantages of extending the MSA treatment of a non-discriminant data matrix (OIV-MA-AS316-01 full $^1$H chemical shift binning outlier), to a supervised approach. Evidently, the PLS-DA treatment of the OIV-MA-AS316-01 aromatic $^1$H binning data matrix, produces a highly discriminant holistic fingerprint for Mexican wines' geographical origins ($R^2$ = 0.96, $Q^2$ = 0.90, Fig. 1 bottom left). Interestingly, both PCA and PLS-DA treatments of the OIV-MA-AS316-01 aromatic $^1$H chemical shift binning NMR outlier produce comparable fingerprints, with a positive PC1 coordinate for Querétaro and Baja California scores.

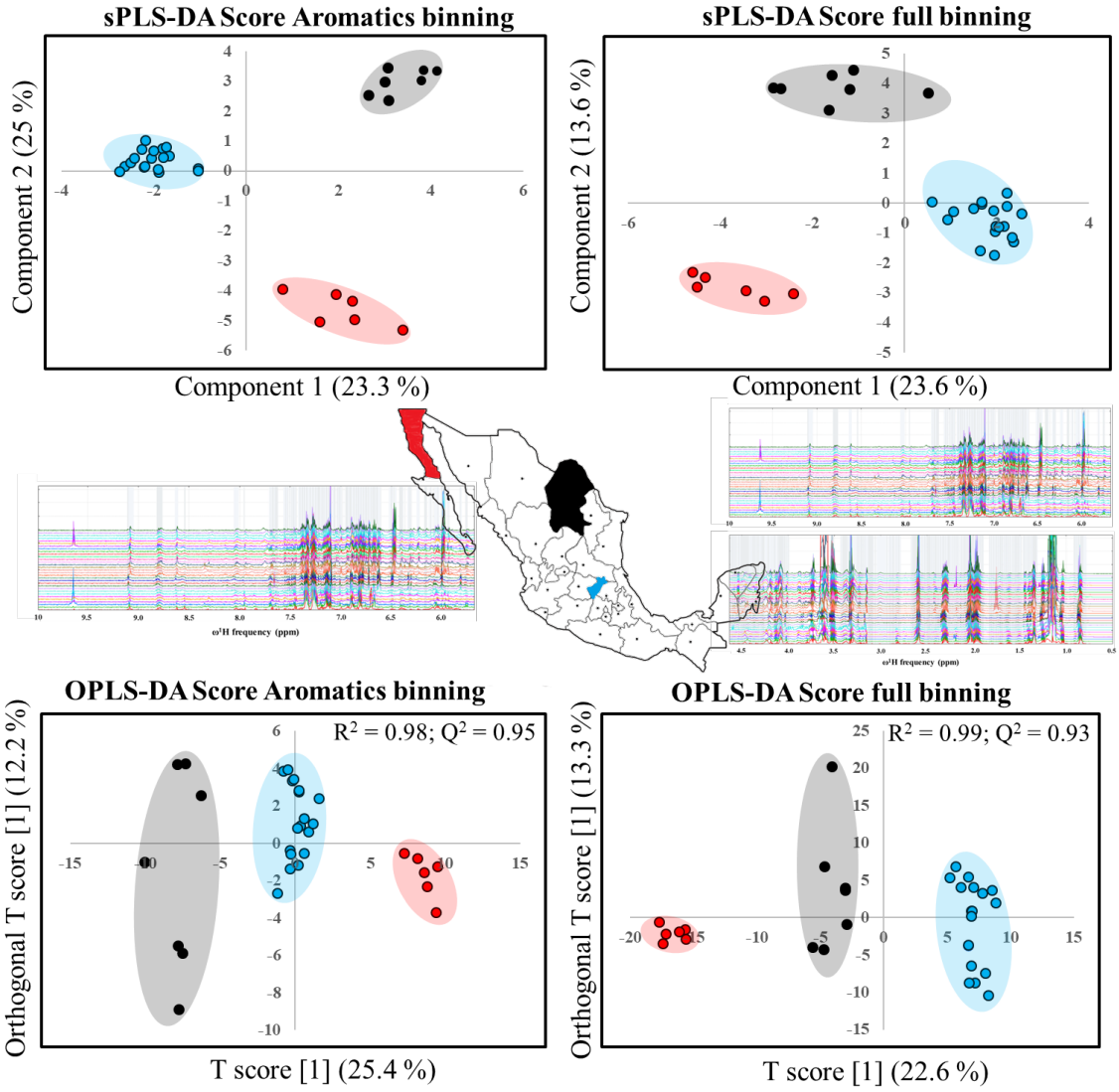
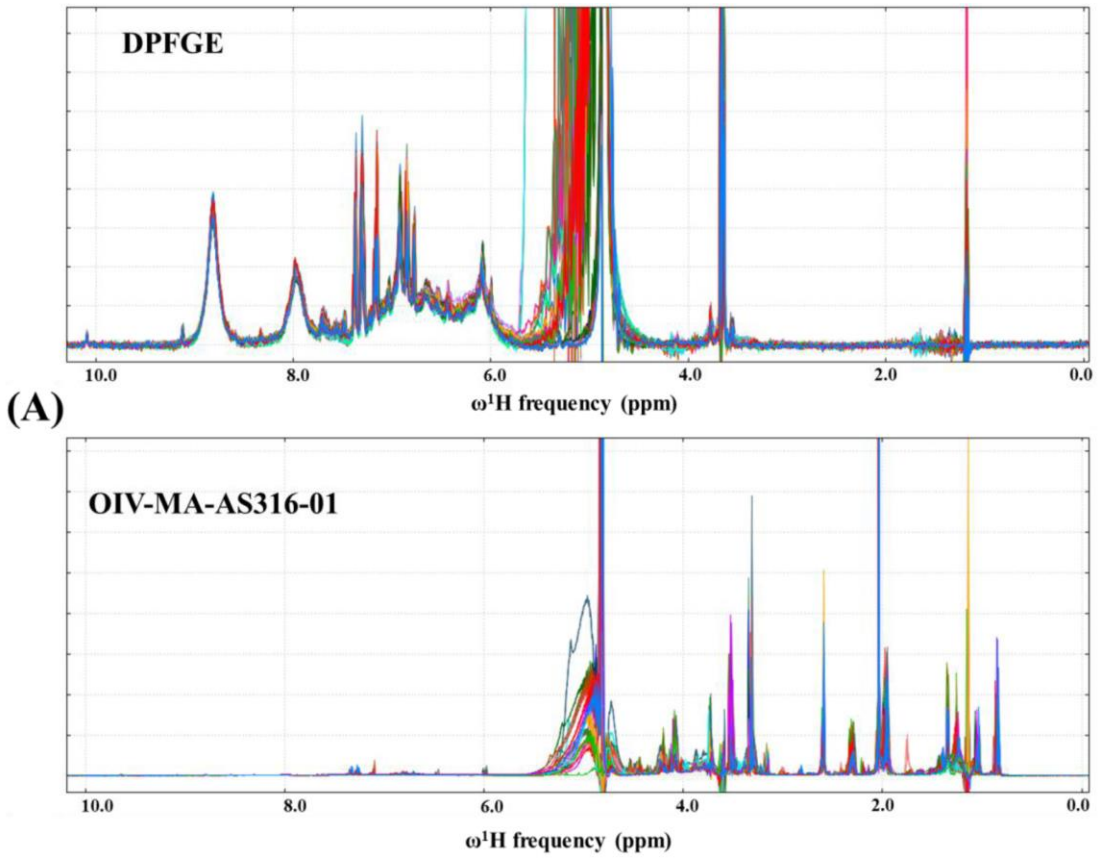
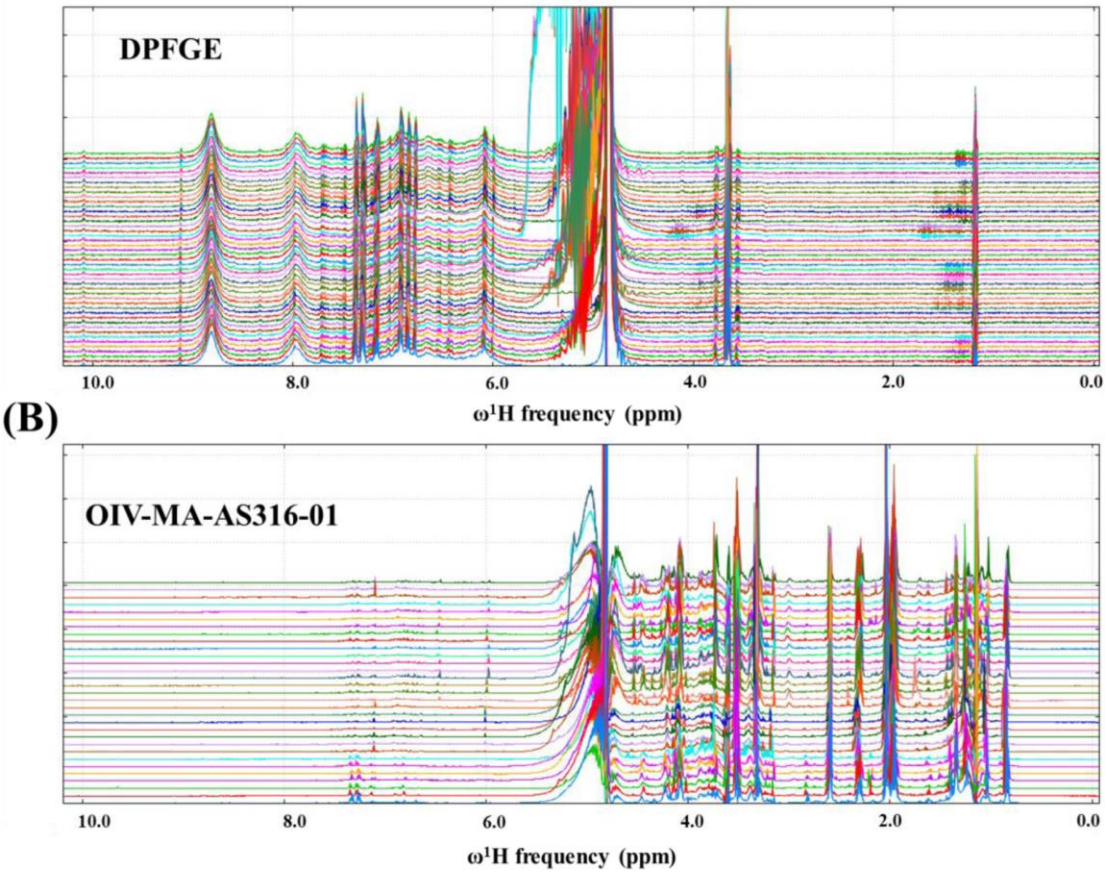



**Figure 2.** Supervised sparse partial least square discriminant analysis (sPLS-DA, top) and orthogonal projections to latent structures discriminant analysis (OPLS-DA, bottom) score plots of two data matrixes obtained with the OIV-MA-AS316-01 acquisition method: full $^1$H-chemical shift binning (right) and signal bucketing of exclusively aromatic signals from 5.5-10 ppm of proton chemical shift range (left). Both signal bucketing strategies are highlighted in the middle of the figure. MSA score plots model discriminant fingerprints related to wines' geographical origin, assigning red T2 Hotelling's ellipse and scores to Baja California, black T2 Hotelling's ellipse and scores to Coahuila and blue T2 Hotelling's ellipse and scores to Queretaro samples, in agreement with country's colour map.


DPFGE
10.0
8.0
6.0
4.0
2.0
0.0
ω¹H frequency (ppm)
(A)
OIV-MA-AS316-01
10.0
8.0
6.0
4.0
2.0
0.0
ω¹H frequency (ppm)


DPFGE
10.0
8.0
6.0
4.0
2.0
0.0
ω¹H frequency (ppm)
(B)
OIV-MA-AS316-01
10.0
8.0
6.0
4.0
2.0
0.0
ω¹H frequency (ppm)

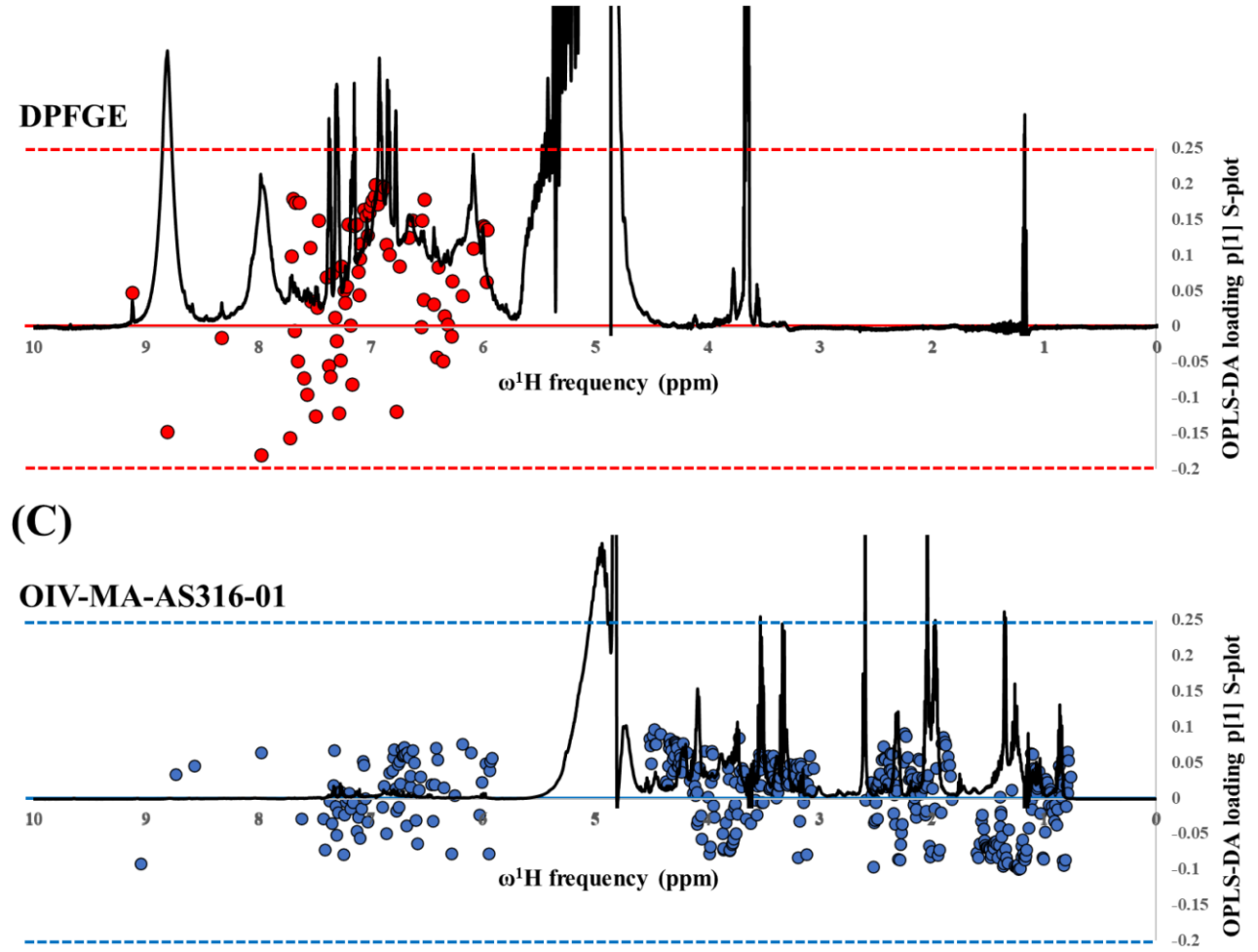



**Figure 3.** (A) Overlaid NMR spectra that defines DPFGE (top) and OIV-MA-AS316-01 (bottom) metabolomics outliers; (B): Stacked NMR spectra that defines DPFGE (top) and OIV-MA-AS316-01 (bottom) metabolomics outliers and (C): OPLS-DA loading p[1] S-plots of DPFGE (top) and OIV-MA-AS316-01 (bottom) data matrixes, as a function of $^{1}$H –NMR chemical shifts, indicating the importance of relevant metabolites to respectively discriminate fermentation schemes of wines from same variety -geographical origin and geographical origins of wines from diverse grape varieties and year of vintages.

In order to maximize separations amongst samples, supervised sparse PLS-DA and OPLS-DA deep learning algorithms [22] were applied to each NMR data matrix (Fig. 2). Whilst sPLS-DA performs variable selection and classification of key features between cohorts in a one-step procedure, presenting excellent predictive performances at competitive computational costs, OPLS-DA permits us to obtain optimal information from the dataset by the identification of a more refined multivariate subspace for the maximum group separations by applying Monte-Carlo Cross Validations with a set of partitions per number of permutations, often used as alternatives when PCA and/or partial least square algorithms fails to expose group separations. As observed in Figure 2, both OIV-MA-AS316-01 NMR outliers produce robust geographical origin score plots for Mexican wines.

The discriminant capacity of aromatic moieties present in wines such as flavonoids (tannins, proanthocyanidins, etc.) or flavonols (quercetin-3-O-glucoside or cleaved quercetin) observed since unsupervised PCA machine learning algorithms were applied to OIV-MA-AS316-01 aromatics $^{1}$H chemical shift binning outliers, opens the venue to propose novel NMR acquisition schemes -in the basis of maximizing aromatic protons' signal resolution and sensitivity, in a fast and easy way - that can produce unprecedented data matrixes for novel generation *ŒNO-NMR* based metabolomics approaches. Last requirements are fully covered with the novel Double Pulsed Field Gradient Echo (DPFGE) NMR experiment, whereas the acquisition details have been elsewhere published [4].

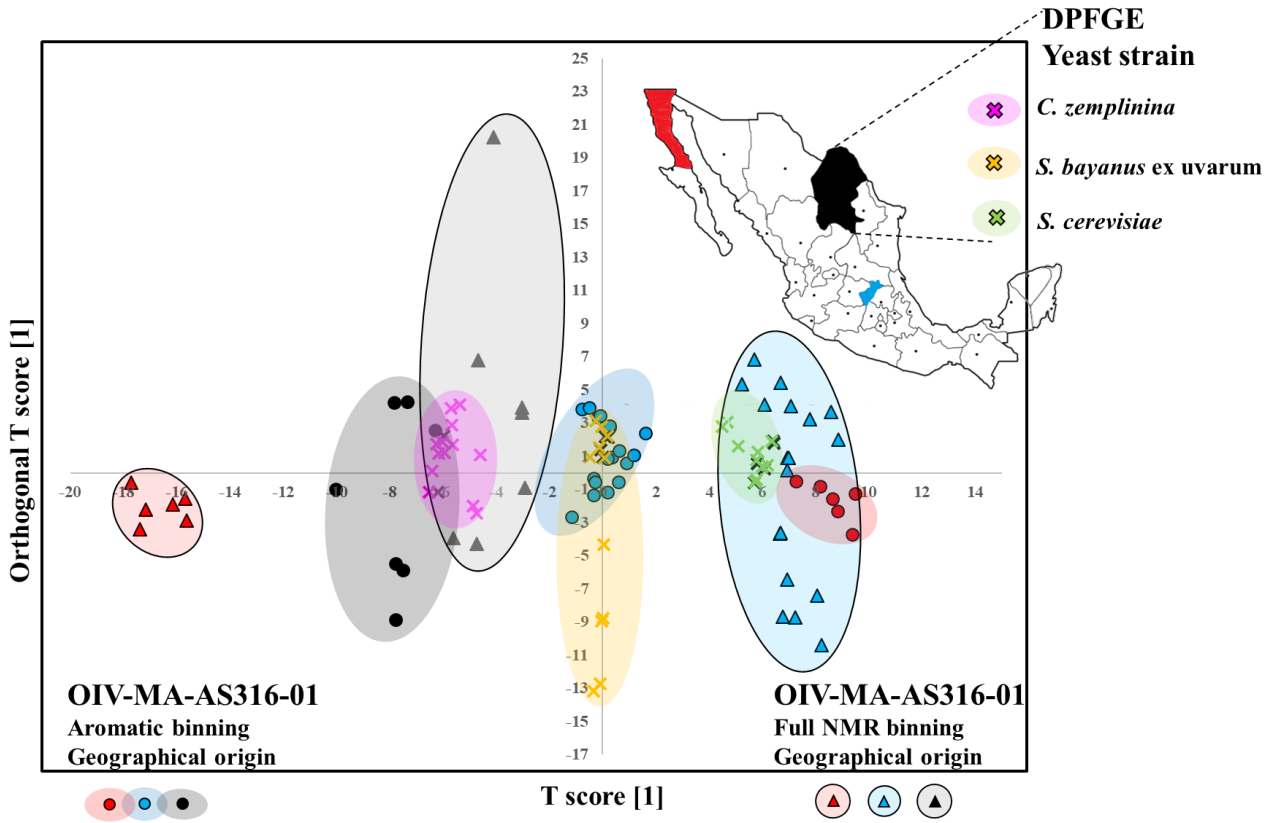



**Figure 4.** Comparative orthogonal projections to latent structures discriminant analysis (OPLS-DA) score plots of the full data set: OIV-MA-AS316-01 aromatic binning NMR data matrix (red, blue, and black circled score plots to discriminate wines' geographical origin, as shown in Mexico's map); OIV-MA-AS316-01 full binning NMR data matrix (red, blue, and black triangled score plots to discriminate wines' geographical origin, as shown in Mexico's map) and aromatics' DPFGE data matrix to discriminate different fermentation schemes of wines from identical geographical origin (Coahuila, as shown in Mexico's map)and Cabernet sauvignon variety (green: Standard inoculation with Saccharomyces *cerevisiae* D254™ yeast strain; magenta: a first-step co-inoculation with non-*Saccharomyces Candida zemplinina* yeast strain, followed by a later inoculation with a D254™, *Saccharomyces cerevisiae* yeast strain and yellow: inoculation with *Saccharomyces bayanus ex uvarum*combined yeast strain).

Comparative DPFGE and OIV-MA-AS316-01 NMR outliers, serving as inputs for machine and/or deep learning routines for non-targeted and targeted *ŒNO-NMR* metabolomics, are shown in Figure 3. Overlaid (A) and stacked (B) DPFGE and OIV-MA-AS316-01 NMR spectra in Figure 3 respectively illustrate the reproducibility and the complexity of each data matrix. By plotting both overlaid and stacked DPFGE and OIV-MA-AS316-01 NMR spectra at comparable conditions, it is unambiguously detected the important incrementin signal to noise ratio (S/N),within the $^{1}$H aromatic chemical shift range, with the use of the DPFGE scheme (S/N= 68.72 with the use of an NMR receiver gain of 203), with respect its OIV-MA-AS316-01multipresat counterpart (S/N= 10.3 with the use of an NMR receiver gain of 40)[4].Finally, the discriminant capacity of each binned chemical shift in both DPFGE (Fig. 3C top) and OIV-MA-AS316-01 NMR (Fig. 3C bottom) outliers, can be elucidated by means of their reported OPLS-DA loading p[1] S-plotnet values. In this sense, the DPFGE OPLS-DA loadings are at least 4- 5 times higher in magnitude and thus, more discriminant that p[1] loadings obtained from OIV-MA-AS316-01 NMR spectra, as shown in Figure 3C. Besides, novel broad resonances within the 5.5-10 ppm aromatic chemical shift range are observed in the DPFGE NMR spectra, whereas by their line widths at half height (LWHH) of 78-86 Hz, respectively for signals at 8.813 and 7.98 ppm (Fig. 3, DPFGE NMR outliers), said resonances must correspond to polyphenols of considerable hydrodynamic radii size [27, 28].

Figure 4 depicts an overall OPLS-DA score plot, including the full set of Mexican Cabernet sauvignon and Merlot monovarietal wines from three different geographical regions, diverse years of vintage and three different fermentation schemes (non-*Saccharomyces Candida zemplinina*, standard *Saccharomyces cerevisiae,* and *Saccharomyces bayanus ex uvarum*). In other words, it is presented the full data set of Mexican wines analyzed for the present study, with two NMR outliers: DPFGE and OIV-MA-AS316-01, whereas for the later, it is included another set of outliers obtained with two chemical shift binning strategies (*red, blue, and black circled scores produced from* ***OIV-MA-AS316-01 $^{1}$H-aromatic binning*** *NMR data matrix; red, blue, and black triangled score plots produced from* ***OIV-MA-AS316-01 $^{1}$H-full binning*** *NMR data matrix*). Interestingly it is observed that scores obtained from different methods, either by a different NMR outlier produced from the same NMR experiment, or by different NMR outliers

obtained from different acquisition schemes, produce correlations amongst groups, as depicted by overlays between scores with the common group "Coahuila" (black triangled and circled scores and magenta crosses). However, maximum group separations amongst the multivariate subspace provided by OPLS-DA deep learning algorithms will better reflect accurate discriminations between discriminant factors (geographical origin, fermentation schemes) and outliers (OIV-MA-AS316-01 aromatic $^{1}$H binning/OIV-MA-AS316-01 full $^{1}$H binning / DPFGE) by using the same acquisition, processing and multivariate statistical analysis.

## 4 Conclusions

The OIV-MA-AS316-01 NMR acquisition scheme, as part of the OIV methods recently published at the Compendium of International Methods of Analysis of Wines and Musts, can produce different data matrixes with specific discriminant capacities to differentiate between geographical origins, as a function of the machine – deep learning algorithm used in MSA metabolomics workflows. First, the OIV-MA-AS316-01 aromatics 1H chemical shift binning outlier is the most accurate data matrix to discriminate several factors in wines, such as geographical origins, even with unsupervised PCA approaches. This observation strongly suggests that simple and complex polyphenols such as flavonoids or flavonols, are the most discriminative metabolites in wines, towards diverse factors such as geographical origins, varieties, year of vintages, manufacturing processes, amongst several more. In complement, a huge plethora of data matrixes that can be derived from the OIV-MA-AS316-01 NMR method (full or partial binning pre-processing) might produce discriminant holistic fingerprints of diverse factors such as geographical origins, with the use of supervised MSA deep learning algorithms such as PLS-DA, its sparse version and the robust OPLS-DA.

The Double Pulsed Field Gradient Echo NMR experiment, with a selective irradiation over the aromatic chemical shift range in $^{1}$H NMR wine spectra, has proven its robustness to even classify wines from the same geographical origin, grape variety, and year of vintage, but produced with different fermentation schemes. The DPFGE NMR data matrix easily produces novel $^{1}$H broad resonances, as new generation fingerprints for future wine-omics studies, like for instance optimization of novel fermentation schemes or climate change impacts.

The OIV Digital transformation Plan includes four main axes: i) Digital platforms network, ii) Digital Garage, iii) Advanced Analytics Labs and iv) Digital Transformation Observatory Hub. The Hub currently reports the importance of digitalization mainly in vineyard, winery and distribution within vine and wine value chain. Specific impacts relating vine and wine digitalization and Advanced Analytics Labs mostly considering the use of Artificial Intelligence and thus the consolidation of omics sciences, including ŒNO-NMR based metabolomics comprise: a) improved data collection, b) increased traceability, c) improved product quality, d) improved environmental and social sustainability, e) creation of new products, f) inclusion and accessibility for smaller produces, and g) design of strategies towards mitigating climate change. Currently in 2022, ŒNO-NMR based metabolomics is under massive expansion for crop monitoring and management, as well as for wine production and quality process monitoring.

## Acknowledgments

Different funding sources are acknowledged: J.E.H.-.P thanks financial support from Instituto Politécnico Nacional (IPN) program "Programa Institucional de Contratación de Personal Académico de Excelencia (PICPAE)" and "Estímulos al Desempeño de los Investigadores (EDI)" EDI-IPN financial support grant No. F-00318. All authors thanks CONACyT program No. LN295321 "Laboratorios Nacionales" and CONACyT-México infrastructure grant: INFRA-2016 (269012) for funding Nuclear Magnetic Resonance experimental time.

J.E.H.-P., F.R.-G., and C.P.-V. acknowledge the Consejo Mexicano Vitivinícola and the Mexican Ministry of Agriculture (SADER) for all financial support for attending OIV duties (expert meetings, congresses, etc.).

J.E.H.-P. specially acknowledge Eng. Isela Mejía-Fonseca for unconditional assistance.

## References

1. https://www.oiv.int/en/technical-standards-and-documents/methods-of-analysis/compendium-of-international-methods-of-analysis-of-wines-and-musts
2. M. Esteki, J. Simal-Gandara, Z. Shahsavari, S. Zandbaaf, E. Dashtaki, Y.A. Vander Heyden, Food Control **93**, 165 (2018) doi.org/10.1016/j.foodcont.2018.06.015
3. Y. Pomyen, K. Wanichthanarak, P. Poungsombat, J. Fahrmann, D. Grapov, S. Khoomrung, Comp. Struct. Biotechnol. J. **18**, 2818 (2020) doi.org/10.1016/Jj.csbj.2020.09.033
4. J. E. Herbert-Pucheta, J.D. Lozada-Ramírez, A.E. Ortega-Regules, L.R. Hernández, C. Anaya de Parrodi, Molecules **26**, 4146 (2021) doi.org/10.3390/molecules26144146
5. R. López-Aguilar, H. Zuleta-Prada, A. Hernández-Montes, J. E. Herbert-Pucheta, Foods **10**, 157 (2021) doi.org/10.3390/foods10010157
6. L. Franitza, L. Nicolotti, M. Granvogl, P. Schieberle, J. Agric. Food Chem. **66**, 3038 (2018) doi.org/10.1021/acs.jafc.8b00180
7. M. Stupak, I. Goodall, M. Tomaniova, J. Pulkrabova, H. Jana, Anal. Chimica Acta **1042**, 60 (2018) doi.org/10.1016/j.aca.2018.09.017
8. M. Amargianitaki, A. Spyros, Chem. Biol. Technol. Agric. **4**, 1 (2017) doi.org/10.1186/s40538-017-0092-x
9. P.A. Solovyev, C. Fauhl-Hassek, J.Riedl, S. Esslinger, L. Bontempo, F. Camin, Compr. Rev. Sci. Food Saf. **20**, 2040 (2021) doi:10.1111/1541-4337.12700

10. M. Viskić, L. M. Bandić, A.M.J. Korenika, Jeromel, Foods **10**, 120 (2021) doi:10.3390/foods10010120
11. R. Godelmann, F. Fang, E. Humpfer, B. Schütz, M. Bansbach, H. Schäfer, M. Spraul, J. Agric. Food Chem. **61**, 5610 (2013) doi:10.1021/jf400800d
12. Y. Monakhova, R. Godelmann, T. Kuballa, S. Mushtakova, D. Rutledge, Talanta **141**, 60 (2015) doi:10.1016/j.talanta.2015.03.037
13. P. Mazzei, R. Spaccini, N. Francesca, G. Moschetti, A. Piccolo, J. Agric. Food Chem. **61**, 10816 (2013) doi:10.1021/jf403567x
14. L. Gougeon, G. Da Costa, I. Le Mao, W. Ma, P. Teissedre, F. Guyon, T. Richard, Food Anal. Methods **11**, 3425 (2018) doi:10.1007/s12161-018-1310-2
15. B. Hu, J. Gao, S. Xu, J. Zhu, X. Fan, X. Zhou, Appl. Biol. Chem. **63**, 24 (2020) doi:10.1186/s13765-020-00509-x
16. L. Kubicova, F. Hadacek, G. Bachmann, W. Weckwerth, V. Chobot, Antioxidants **8**, 476 (2019) doi:10.3390/antiox8100476
17. J.M. Del Fresno, C. Escott, I. Loira, J.E. Herbert-Pucheta, R. Schneider, F. Carrau, R. Cuerda, A. Morata, Fermentation **6**, 66 (2020) doi:0.3390/fermentation6030066
18. J.E. Herbert-Pucheta, I. Mejía-Fonseca, L.G. Zepeda-Vallejo, D. Milmo-Brittingham, G. Padilla-Maya, "The "Wine-T1" NMR experiment for novel wine-metabolome fingerprinting with nuclear-spin relaxation" EDP-Sciences BIO Web of Conferences (2019), 02029, 1-5. DOI: 10.1051/bioconf/20191202029
19. J.E. Herbert-Pucheta, C. Pino-Villar, F. Rodríguez-González, G. Padilla-Maya, D. Milmo-Brittingham, L.G. Zepeda-Vallejo, EDP-Sciences BIO Web of Conferences; (2019), **02016**, 1-5. DOI: 10.1051/bioconf/20191502016
20. D. Jacob, C. Deborde, M. Lefevbre, M. Maucourt, A. Moing, Metabolomics **13**, 36 (2017) doi:10.1007/s11306-017-1178-y
21. J. Chong, D.S. Wishart, J. Xia, Curr. Protoc. Bioinformatics **68**, e68 (2019) doi:10.1002/cpbi.86
22. E. Holmes, R.L. Loo, J. Stamler, M. Bictash, I.K.S. Yap, Q. Chan, T. Ebbels, M. De Iorio, I.J. Brown, K.A. Veselkov, et al. Nature **453**, 396 (2008) doi:10.1038/nature06882
23. X. Qing-Song, L. Yi-Zeng, Chemom. Intell. Lab. Syst. **56**, 1 (2001) doi:10.1016/S0169-7439(00)00122-2
24. E. Szymańska, E. Saccenti, A.K. Smilde, J.A. Westerhuis, Metabolomics **8**, 3 (2012) doi.org/10.1007/s11306-011-0330-3
25. B. Worley, R. Powers, Curr. Metabolomics **4**, 97 (2016) doi.org/10.2174/2213235X04666160613122429
26. J.E. Herbert-Pucheta, C.A. López-Morales, E. Medina-Rivero, S. Estrada Parra, S.M. Pérez-Tapia, L.G. Zepeda-Vallejo, J. Pharm. Biomed. Anal. **196**, 113940 (2021) 10.1016/j.jpba.2021.113940
27. S. Vázquez-Leyva, L. Vallejo-Castillo, C.A. López-Morales, J.E. Herbert-Pucheta, L.G. Zepeda-Vallejo, M. Velasco-Velázquez, L. Pavón, S.M. Pérez-Tapia, E. Medina-Rivero, Anal. Chem. **91**, 14392 (2019) 10.1021/acs.analchem.9b02873